Master Formulae Book

Comprehensive Equations Concepts

Milav Dabgar

June 3, 2026

Contents

Part I

Stanford

CHAPTER 1

Aa228V Validation Safety Critical Systems

Discrete Reachability

$$\begin{aligned} v(4) &= 1 \quad \text{(Target state)} \\ v(1) &= 0 \quad \text{(State 1 is absorbing and not the target, so } T \text{ is unreachable)} \end{aligned}$$

$$\begin{aligned} v(2) &= P_{21}v(1) + P_{23}v(3) \\ v(3) &= P_{32}v(2) + P_{34}v(4) \end{aligned}$$

$$\begin{aligned} v(2) &= 0.2(0) + 0.8v(3) \implies v(2) = 0.8v(3) \\ v(3) &= 0.5v(2) + 0.5(1) \implies v(3) = 0.5v(2) + 0.5 \end{aligned}$$

$$\begin{aligned} v(3) &= 0.5(0.8v(3)) + 0.5 \\ v(3) &= 0.4v(3) + 0.5 \\ 0.6v(3) &= 0.5 \\ v(3) &= \frac{5}{6} \approx 0.833 \end{aligned}$$

$$v(2) = 0.8\left(\frac{5}{6}\right) = \frac{4}{6} = \frac{2}{3} \approx 0.667$$

$$\begin{aligned} z_i^{\min} &= \lfloor y_i^{\min}/\Delta x_i \rfloor \Delta x_i \\ z_i^{\max} &= \lceil y_i^{\max}/\Delta x_i \rceil \Delta x_i \end{aligned}$$

$$\begin{aligned} z_1^{\min} &= \lfloor 1.3/1.0 \rfloor \times 1.0 = 1 \times 1.0 = 1.0 \\ z_1^{\max} &= \lceil 3.8/1.0 \rceil \times 1.0 = 4 \times 1.0 = 4.0 \end{aligned}$$

$$\begin{aligned} z_2^{\min} &= \lfloor 3.9/0.5 \rfloor \times 0.5 = \lfloor 7.8 \rfloor \times 0.5 = 7 \times 0.5 = 3.5 \\ z_2^{\max} &= \lceil 4.2/0.5 \rceil \times 0.5 = \lceil 8.4 \rceil \times 0.5 = 9 \times 0.5 = 4.5 \end{aligned}$$

$$\begin{aligned} \text{Widths of } Y: w_{Y,1} &= 3.8 - 1.3 = 2.5, \quad w_{Y,2} = 4.2 - 3.9 = 0.3 \\ \text{Vol}(Y) &= 2.5 \times 0.3 = 0.75 \end{aligned}$$

$$\begin{aligned} \text{Widths of } Z: w_{Z,1} &= 4.0 - 1.0 = 3.0, \quad w_{Z,2} = 4.5 - 3.5 = 1.0 \\ \text{Vol}(Z) &= 3.0 \times 1.0 = 3.0 \end{aligned}$$

$$\begin{aligned} \text{Expansion in dim 1} &= 3.0 - 2.5 = 0.5 \leq 2\Delta x_1 = 2.0 \\ \text{Expansion in dim 2} &= 1.0 - 0.3 = 0.7 \leq 2\Delta x_2 = 1.0 \end{aligned}$$

Explainability

$$
\begin{aligned}
w_1(1) + w_2(1) + b &= 0.8 \\
w_1(1.5) + w_2(1) + b &= 0.9 \\
w_1(1) + w_2(1.5) + b &= 0.85
\end{aligned}
$$

Guest Lecture: Mining Failures in Vision-Based Controllers

$$
\begin{aligned}
S(x_1) &= (1)^2 + 2(1)^2 = 3 \\
S(x_2) &= (-1)^2 + 2(2)^2 = 9 \\
S(x_3) &= (0)^2 + 2(-1)^2 = 2 \\
S(x_4) &= (2)^2 + 2(0)^2 = 4
\end{aligned}
$$

$$
\begin{aligned}
(x_2 - \mu^{(1)})(x_2 - \mu^{(1)})^T &= \begin{bmatrix} -1.5 \\ 1 \end{bmatrix} \begin{bmatrix} -1.5 & 1 \end{bmatrix} = \begin{bmatrix} 2.25 & -1.5 \\ -1.5 & 1 \end{bmatrix} \\
(x_4 - \mu^{(1)})(x_4 - \mu^{(1)})^T &= \begin{bmatrix} 1.5 \\ -1 \end{bmatrix} \begin{bmatrix} 1.5 & -1 \end{bmatrix} = \begin{bmatrix} 2.25 & -1.5 \\ -1.5 & 1 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
x = G(1) &= \begin{bmatrix} 2(1) \\ (1)^2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\
u = C(x) &= 0.5(2) + 1 = 2 \\
L(1) = -(2)^2 &= -4
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial u} &= -2u = -2(2) = -4 \\
\frac{\partial C}{\partial x} &= \begin{bmatrix} 0.5 & 1 \end{bmatrix} \\
\frac{\partial G}{\partial z} &= \begin{bmatrix} 2 \\ 2z \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\hat{P}_F &= \frac{1}{3} \sum_{i=1}^{3} I_F(x_i) w(x_i) \\
&= \frac{1}{3}((0)(0.2) + (1)(0.2) + (1)(0.2)) \\
&= \frac{0.4}{3} \approx 0.133
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(\hat{P}_F) \approx \frac{1}{3(2)} \sum_{i=1}^{3} (I_F(x_i) w(x_i) - 0.133)^2 \\
&= \frac{1}{6} \left[(0 - 0.133)^2 + (0.2 - 0.133)^2 + (0.2 - 0.133)^2 \right] \\
&= \frac{1}{6} [0.0177 + 0.0045 + 0.0045] \\
&= \frac{1}{6} [0.0267] \approx 0.00445
\end{aligned}
$$

Runtime Monitoring

$$\mathcal{D} = \{(1,1),(2,1),(2,2),(3,3)\}$$

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 30 \\ 0 \\ 5 \\ -2 \\ 32 \end{bmatrix}$$

$$A\mathbf{x}_1 = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 25 \\ 4 \end{bmatrix} = \begin{bmatrix} 25 \\ -25 \\ 4 \\ -4 \\ 29 \end{bmatrix}$$

$$A\mathbf{x}_2 = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 28 \\ 6 \end{bmatrix} = \begin{bmatrix} 28 \\ -28 \\ 6 \\ -6 \\ 34 \end{bmatrix}$$

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$$|\Sigma| = (2)(2) - (0)(0) = 4$$

$$\Sigma^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

$$(\mathbf{x}_{test}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}_{test}-\boldsymbol{\mu}) = \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (2)(1) + (2)(1) = 4$$

$$p(\mathbf{x}_{test}) = \frac{1}{\sqrt{(2\pi)^2(4)}} \exp\left(-\frac{1}{2}(4)\right)$$

$$= \frac{1}{\sqrt{16\pi^2}} e^{-2} = \frac{1}{4\pi} e^{-2}$$

$$p(\mathbf{x}_{test}) = \frac{1}{12.5664}(0.1353) \approx (0.0796)(0.1353) \approx 0.0108$$

$$\hat{p}_i = \frac{\exp(z_i/T)}{\sum_{j=1}^{K} \exp(z_j/T)}$$

$$\exp(2.0) \approx 7.389$$

$$\exp(1.0) \approx 2.718$$

$$\exp(0.1) \approx 1.105$$

$$\Sigma = 7.389 + 2.718 + 1.105 = 11.212$$

$$\hat{p}_1 = \frac{7.389}{11.212} \approx 0.659$$

$$\hat{p}_2 = \frac{2.718}{11.212} \approx 0.242$$

$$\hat{p}_3 = \frac{1.105}{11.212} \approx 0.099$$

$$\mathbf{z}/T = [1.0, 0.5, 0.05]$$

$$\exp(0.5) \approx 1.649$$

$$\exp(0.05) \approx 1.051$$

$$\Sigma = 2.718 + 1.649 + 1.051 = 5.418$$

$$\hat{p}_1 = \frac{2.718}{5.418} \approx 0.502$$

$$\hat{p}_2 = \frac{1.649}{5.418} \approx 0.304$$

$$\hat{p}_3 = \frac{1.051}{5.418} \approx 0.194$$

$$k = \lceil (n+1)(1-\alpha) \rceil = \lceil (9+1)(1-0.20) \rceil = \lceil 10 \times 0.80 \rceil = \lceil 8 \rceil = 8$$

$$\hat{q} = 0.50$$

$$\hat{\mu}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^{M} f_m(\mathbf{x})$$

$$\sigma^2_{epistemic}(\mathbf{x}) = \frac{1}{M-1} \sum_{m=1}^{M} (f_m(\mathbf{x}) - \hat{\mu}(\mathbf{x}))^2$$

$$\hat{\mu}_A = \frac{20.1 + 20.3 + 19.8 + 20.0 + 20.2}{5} = \frac{100.4}{5} = 20.08$$

$$\sigma^2_A = \frac{1}{5-1} \left[(20.1 - 20.08)^2 + (20.3 - 20.08)^2 + (19.8 - 20.08)^2 + (20.0 - 20.08)^2 + (20.2 - 20.08)^2 \right]$$

$$\sigma^2_A = \frac{1}{4} \left[(0.02)^2 + (0.22)^2 + (-0.28)^2 + (-0.08)^2 + (0.12)^2 \right]$$

$$\sigma^2_A = \frac{1}{4} [0.0004 + 0.0484 + 0.0784 + 0.0064 + 0.0144]$$

$$\sigma^2_A = \frac{0.148}{4} = 0.037$$

$$\hat{\mu}_B = \frac{15.0 + 25.5 + 10.2 + 30.1 + 18.4}{5} = \frac{99.2}{5} = 19.84$$

$$\sigma^2_B = \frac{1}{4} \left[(15.0 - 19.84)^2 + (25.5 - 19.84)^2 + (10.2 - 19.84)^2 + (30.1 - 19.84)^2 + (18.4 - 19.84)^2 \right]$$

$$\sigma^2_B = \frac{1}{4} \left[(-4.84)^2 + (5.66)^2 + (-9.64)^2 + (10.26)^2 + (-1.44)^2 \right]$$

$$\sigma^2_B = \frac{1}{4} [23.4256 + 32.0356 + 92.9296 + 105.2676 + 2.0736]$$

$$\sigma^2_B = \frac{255.732}{4} = 63.933$$

$$f(x'_1) = 2.1^2 + \sin(0.1) \approx 4.41 + 0.0998 = 4.5098$$
$$f(x'_2) = 1.9^2 + \sin(-0.1) \approx 3.61 - 0.0998 = 3.5102$$
$$f(x'_3) = 2.0^2 + \sin(0.2) \approx 4.0 + 0.1987 = 4.1987$$

$$\alpha_1^c = \frac{1}{4}(0.5 + 0.1 + 0.2 + 0.4) = \frac{1.2}{4} = 0.3$$
$$\alpha_2^c = \frac{1}{4}(-0.2 - 0.1 - 0.1 - 0.4) = \frac{-0.8}{4} = -0.2$$

$$\begin{aligned}\sum_k \alpha_k^c A^k &= 0.3 \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} - 0.2 \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 0.6 & 0 \\ 0.3 & 0.9 \end{bmatrix} - \begin{bmatrix} 0.2 & 0.2 \\ 0 & 0.4 \end{bmatrix} \\ &= \begin{bmatrix} 0.4 & -0.2 \\ 0.3 & 0.5 \end{bmatrix}\end{aligned}$$

CHAPTER 2

CS109 Probability

Counting I

$$|A_1 \cup A_2 \cup \cdots \cup A_k| = |A_1| + |A_2| + \cdots + |A_k|$$

$$|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \times |A_2| \times \cdots \times |A_k|$$

$$|A \cup B| = |A| + |B| - |A \cap B|$$

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

What is Probability?

$$P(E) = \frac{|E|}{|S|}$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$$

Conditional Probability and Bayes

$$\begin{aligned} P(D) &= P(D|M_1)P(M_1) + P(D|M_2)P(M_2) \\ &= (0.02)(0.60) + (0.05)(0.40) \\ &= 0.012 + 0.020 = 0.032 \end{aligned}$$

$$\begin{aligned} P(T^+) &= P(T^+|D)P(D) + P(T^+|D^c)P(D^c) \\ &= (0.99)(0.01) + (0.05)(0.99) \\ &= 0.0099 + 0.0495 = 0.0594 \end{aligned}$$

Independence

$$P(E) = \frac{1}{6}$$

$$P(F) = \frac{6}{36} = \frac{1}{6}$$

$$P(E \cap F) = \frac{1}{36}$$

$$P(E)P(F) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

$$P(\text{at least one success}) = 1 - P(\text{no successes in all } n \text{ trials})$$

$$P(\text{System works}) = P(\text{C}_1 \text{ works}) \times P(\text{C}_2 \text{ works}) \ldots$$

$$P(\text{System works}) = 1 - P(\text{All components fail})$$

$$P(\text{Bucket is empty}) = \left(\frac{m-1}{m}\right)^n$$

$$P(\text{All fail}) = 0.05 \times 0.05 \times 0.05 \times 0.05 = (0.05)^4$$

$$P(H_1) = P(H_1|A)P(A) + P(H_1|B)P(B) = (0.5)(0.5) + (0.9)(0.5) = 0.25 + 0.45 = 0.70$$

$$P(H_1 \cap H_2) = P(H_1 \cap H_2|A)P(A) + P(H_1 \cap H_2|B)P(B)$$

$$P(H_1 \cap H_2|A) = P(H_1|A)P(H_2|A) = (0.5)(0.5) = 0.25$$

$$P(H_1 \cap H_2|B) = P(H_1|B)P(H_2|B) = (0.9)(0.9) = 0.81$$

$$P(H_1 \cap H_2) = (0.25)(0.5) + (0.81)(0.5) = 0.125 + 0.405 = 0.53$$

$$P(X_1, X_2, \ldots, X_n|Y) = \prod_{i=1}^{n} P(X_i|Y)$$

$$P(Y|X_1, \ldots, X_n) = \frac{P(X_1, \ldots, X_n|Y)P(Y)}{P(X_1, \ldots, X_n)} = \frac{P(Y)\prod_{i=1}^{n} P(X_i|Y)}{P(X_1, \ldots, X_n)}$$

$$P(X_1 \cap X_2|S) = P(X_1|S)P(X_2|S) = 0.5 \times 0.3 = 0.15$$

$$P(X_1 \cap X_2|S^c) = P(X_1|S^c)P(X_2|S^c) = 0.01 \times 0.05 = 0.0005$$

$$P(X_1 \cap X_2) = P(X_1 \cap X_2|S)P(S) + P(X_1 \cap X_2|S^c)P(S^c) = 0.06 + 0.0003 = 0.0603$$

$$P(S|X_1 \cap X_2) = \frac{0.06}{0.0603} \approx 0.995$$

Inference II

$$\begin{aligned}
P(X_1 = \text{Rain}) &= P(X_1 = \text{Rain} \mid X_0 = \text{Rain})P(X_0 = \text{Rain}) \\
&\quad + P(X_1 = \text{Rain} \mid X_0 = \text{Sun})P(X_0 = \text{Sun}) \\
&= (0.7)(0.5) + (0.3)(0.5) = 0.35 + 0.15 = 0.5 \\
P(X_1 = \text{Sun}) &= P(X_1 = \text{Sun} \mid X_0 = \text{Rain})P(X_0 = \text{Rain}) \\
&\quad + P(X_1 = \text{Sun} \mid X_0 = \text{Sun})P(X_0 = \text{Sun}) \\
&= (0.3)(0.5) + (0.7)(0.5) = 0.15 + 0.35 = 0.5
\end{aligned}$$

$$\begin{aligned}
U(X_1 = \text{Rain}) &= P(E_1 = \text{Umbrella} \mid X_1 = \text{Rain})P(X_1 = \text{Rain}) \\
&= (0.9)(0.5) = 0.45 \\
U(X_1 = \text{Sun}) &= P(E_1 = \text{Umbrella} \mid X_1 = \text{Sun})P(X_1 = \text{Sun}) \\
&= (0.2)(0.5) = 0.10
\end{aligned}$$

$$
\begin{aligned}
P(X_1 = \text{Rain} \mid E_1 = \text{Umbrella}) &= \frac{0.45}{0.55} \approx 0.818 \\
P(X_1 = \text{Sun} \mid E_1 = \text{Umbrella}) &= \frac{0.10}{0.55} \approx 0.182
\end{aligned}
$$

$$
\begin{aligned}
P(X_2 = \text{Rain} \mid E_1) &= P(X_2 = \text{Rain} \mid X_1 = \text{Rain})P(X_1 = \text{Rain} \mid E_1) \\
&\quad + P(X_2 = \text{Rain} \mid X_1 = \text{Sun})P(X_1 = \text{Sun} \mid E_1) \\
&= (0.7)(0.818) + (0.3)(0.182) = 0.5726 + 0.0546 = 0.6272 \\
P(X_2 = \text{Sun} \mid E_1) &= P(X_2 = \text{Sun} \mid X_1 = \text{Rain})P(X_1 = \text{Rain} \mid E_1) \\
&\quad + P(X_2 = \text{Sun} \mid X_1 = \text{Sun})P(X_1 = \text{Sun} \mid E_1) \\
&= (0.3)(0.818) + (0.7)(0.182) = 0.2454 + 0.1274 = 0.3728
\end{aligned}
$$

$$
\begin{aligned}
U(X_2 = \text{Rain}) &= P(E_2 = \text{Umbrella} \mid X_2 = \text{Rain})P(X_2 = \text{Rain} \mid E_1) \\
&= (0.9)(0.6272) \approx 0.5645 \\
U(X_2 = \text{Sun}) &= P(E_2 = \text{Umbrella} \mid X_2 = \text{Sun})P(X_2 = \text{Sun} \mid E_1) \\
&= (0.2)(0.3728) \approx 0.0746
\end{aligned}
$$

$$
\begin{aligned}
P(X_2 = \text{Rain} \mid E_{1:2}) &= \frac{0.5645}{0.6391} \approx 0.883 \\
P(X_2 = \text{Sun} \mid E_{1:2}) &= \frac{0.0746}{0.6391} \approx 0.117
\end{aligned}
$$

$$
\begin{aligned}
P(X_1 = A) &= P(X_1 = A \mid X_0 = A)P(X_0 = A) \\
&= (0.2)(1/3) = \frac{1}{15} \approx 0.0667 \\
P(X_1 = B) &= P(X_1 = B \mid X_0 = A)P(X_0 = A) + P(X_1 = B \mid X_0 = B)P(X_0 = B) \\
&= (0.8)(1/3) + (0.2)(1/3) = \frac{1}{3} \approx 0.3333 \\
P(X_1 = C) &= P(X_1 = C \mid X_0 = B)P(X_0 = B) + P(X_1 = C \mid X_0 = C)P(X_0 = C) \\
&= (0.8)(1/3) + (1.0)(1/3) = \frac{1.8}{3} = 0.6
\end{aligned}
$$

$$
\begin{aligned}
U(X_1 = A) &= P(\text{Red} \mid A)P(X_1 = A) = (0.7)\left(\frac{1}{15}\right) = \frac{0.7}{15} \approx 0.0467 \\
U(X_1 = B) &= P(\text{Red} \mid B)P(X_1 = B) = (0.3)\left(\frac{1}{3}\right) = 0.10 \\
U(X_1 = C) &= P(\text{Red} \mid C)P(X_1 = C) = (0.7)(0.6) = 0.42
\end{aligned}
$$

$$
\begin{aligned}
P(X_1 = A \mid E_1 = \text{Red}) &= \frac{0.7/15}{8.5/15} = \frac{0.7}{8.5} \approx 0.082 \\
P(X_1 = B \mid E_1 = \text{Red}) &= \frac{1.5/15}{8.5/15} = \frac{1.5}{8.5} \approx 0.176 \\
P(X_1 = C \mid E_1 = \text{Red}) &= \frac{6.3/15}{8.5/15} = \frac{6.3}{8.5} \approx 0.741
\end{aligned}
$$

$$
\begin{aligned}
U(X_1 = 1) &= 0 \times 0.25 = 0 \\
U(X_1 = 2) &= 1 \times 0.25 = 0.25 \\
U(X_1 = 3) &= 0 \times 0.25 = 0 \\
U(X_1 = 4) &= 1 \times 0.25 = 0.25
\end{aligned}
$$

$$
\begin{aligned}
P(X_2 = 1 \mid E_1) &= P(1 \mid 2)P(X_1 = 2) + P(1 \mid 4)P(X_1 = 4) = (0.5)(0.5) + (0)(0.5) = 0.25 \\
P(X_2 = 2 \mid E_1) &= P(2 \mid 2)P(X_1 = 2) + P(2 \mid 4)P(X_1 = 4) = (0)(0.5) + (0.5)(0.5) = 0.25 \\
P(X_2 = 3 \mid E_1) &= P(3 \mid 2)P(X_1 = 2) + P(3 \mid 4)P(X_1 = 4) = (0)(0.5) + (0.5)(0.5) = 0.25 \\
P(X_2 = 4 \mid E_1) &= P(4 \mid 2)P(X_1 = 2) + P(4 \mid 4)P(X_1 = 4) = (0.5)(0.5) + (0)(0.5) = 0.25
\end{aligned}
$$

$$
\begin{aligned}
U(X_2 = 1) &= 1 \times 0.25 = 0.25 \\
U(X_2 = 2) &= 0 \times 0.25 = 0 \\
U(X_2 = 3) &= 1 \times 0.25 = 0.25 \\
U(X_2 = 4) &= 0 \times 0.25 = 0
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{u}_1 &= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \\
&= \begin{bmatrix} 0.1 \times 0.8 \\ 0.8 \times 0.2 \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.16 \end{bmatrix}
\end{aligned}
$$

General Inference

$$
\begin{aligned}
f(0) &= P(X = 0)P(Y = 0 \mid X = 0) + P(X = 1)P(Y = 0 \mid X = 1) \\
&= (0.4)(0.9) + (0.6)(0.2) = 0.36 + 0.12 = 0.48 \\
f(1) &= P(X = 0)P(Y = 1 \mid X = 0) + P(X = 1)P(Y = 1 \mid X = 1) \\
&= (0.4)(0.1) + (0.6)(0.8) = 0.04 + 0.48 = 0.52
\end{aligned}
$$

$$
\begin{aligned}
P(Z = 1) &= \sum_{y \in \{0,1\}} P(Z = 1 \mid Y = y) f(y) \\
&= P(Z = 1 \mid Y = 0) f(0) + P(Z = 1 \mid Y = 1) f(1) \\
&= (0.3)(0.48) + (0.7)(0.52) \\
&= 0.144 + 0.364 = 0.508
\end{aligned}
$$

$$
\begin{aligned}
g(0) &= P(Y = 0 \mid X = 0)P(Z = 1 \mid Y = 0) + P(Y = 1 \mid X = 0)P(Z = 1 \mid Y = 1) \\
&= (0.9)(0.3) + (0.1)(0.7) = 0.27 + 0.07 = 0.34 \\
g(1) &= P(Y = 0 \mid X = 1)P(Z = 1 \mid Y = 0) + P(Y = 1 \mid X = 1)P(Z = 1 \mid Y = 1) \\
&= (0.2)(0.3) + (0.8)(0.7) = 0.06 + 0.56 = 0.62
\end{aligned}
$$

$$
\begin{aligned}
P(X = 0, Z = 1) &= P(X = 0) g(0) = (0.4)(0.34) = 0.136 \\
P(X = 1, Z = 1) &= P(X = 1) g(1) = (0.6)(0.62) = 0.372
\end{aligned}
$$

Adding Random Variables

$$
\begin{aligned}
P(Z = 4) &= P(X = 1)P(Y = 3) + P(X = 2)P(Y = 2) + P(X = 3)P(Y = 1) \\
&= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) \\
&= \frac{3}{36} = \frac{1}{12}
\end{aligned}
$$

$$
\begin{aligned}
\mu_Z &= \mu_X - \mu_Y = 10 - 10.5 = -0.5 \\
\sigma_Z^2 &= \sigma_X^2 + \sigma_Y^2 = 0.04 + 0.05 = 0.09
\end{aligned}
$$

$$
\begin{aligned}
M_X(t) &= e^{\lambda_1(e^t-1)} \\
M_Y(t) &= e^{\lambda_2(e^t-1)}
\end{aligned}
$$

$$
\begin{aligned}
M_Z(t) &= M_X(t)M_Y(t) \\
&= e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} \\
&= e^{\lambda_1(e^t-1)+\lambda_2(e^t-1)} \\
&= e^{(\lambda_1+\lambda_2)(e^t-1)}
\end{aligned}
$$

MAP — Maximum A Posteriori

$$
\begin{aligned}
\frac{k+a-1}{p} &= \frac{n-k+b-1}{1-p} \\
(k+a-1)(1-p) &= (n-k+b-1)p \\
k+a-1-(k+a-1)p &= (n-k+b-1)p \\
k+a-1 &= (n+a+b-2)p \\
p &= \frac{k+a-1}{n+a+b-2}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}_{\text{MAP}} = \mu_n &= \sigma_n^2\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \\
&= \frac{4}{7}\left(\frac{5\cdot 11}{4} + \frac{8}{2}\right) \\
&= \frac{4}{7}\left(\frac{55}{4} + 4\right) \\
&= \frac{4}{7}\left(\frac{71}{4}\right) = \frac{71}{7} \approx 10.14
\end{aligned}
$$

Naive Bayes

$$
\begin{aligned}
P(\text{word} = \text{"election"} \mid Y = \text{Sports}) &= \frac{\text{Count}(\text{"election"}, \text{Sports}) + 1}{\text{Total Words}(\text{Sports}) + |V|} \\
&= \frac{0+1}{500+1000} \\
&= \frac{1}{1500} \approx 0.00067
\end{aligned}
$$

$$
\begin{aligned}
\text{Sum}_A &= \log_{10} P(A) + \log_{10} P(X_1|A) + \log_{10} P(X_2|A) + \log_{10} P(X_3|A) \\
&= -1 + (-2) + (-1.301) + (-1.699) \\
&= -6.000
\end{aligned}
$$

$$
\begin{aligned}
\text{Sum}_B &= \log_{10} P(B) + \log_{10} P(X_1|B) + \log_{10} P(X_2|B) + \log_{10} P(X_3|B) \\
&= -0.0458 + (-3) + (-2.699) + (-2.301) \\
&= -8.0458
\end{aligned}
$$

Advanced Probability: DALL-E and GPT

$$P(w_1, w_2, \ldots, w_n) = \prod_{i=1}^{n} P(w_i \mid w_1, \ldots, w_{i-1})$$

$$\log P(w_1, \ldots, w_n) = \sum_{i=1}^{n} \log P(w_i \mid w_1, \ldots, w_{i-1})$$

$$P(\text{"I love math"}) = P(\text{"I"}) \times P(\text{"love"} \mid \text{"I"}) \times P(\text{"math"} \mid \text{"I love"})$$

$$P(\text{"I love math"}) = 0.4 \times 0.8 \times 0.6$$

$$P(\text{"I love math"}) = 0.192$$

$$p_i = \frac{e^{z_i/T}}{\sum_{j=1}^{k} e^{z_j/T}}$$

$$e^{2.0/1.0} = e^2 \approx 7.389$$

$$e^{1.0/1.0} = e^1 \approx 2.718$$

$$7.389 + 2.718 = 10.107$$

$$p_1 = \frac{7.389}{10.107} \approx 0.731, \quad p_2 = \frac{2.718}{10.107} \approx 0.269$$

$$e^{2.0/0.5} = e^4 \approx 54.598$$

$$e^{1.0/0.5} = e^2 \approx 7.389$$

$$54.598 + 7.389 = 61.987$$

$$p_1 = \frac{54.598}{61.987} \approx 0.881, \quad p_2 = \frac{7.389}{61.987} \approx 0.119$$

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I})$$

$$q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I})$$

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

$$\alpha_1 = 1 - 0.2 = 0.8$$

$$\alpha_2 = 1 - 0.5 = 0.5$$

$$\bar{\alpha}_2 = \alpha_1 \times \alpha_2 = 0.8 \times 0.5 = 0.4$$

$$\mu = \sqrt{\bar{\alpha}_2}x_0 = \sqrt{0.4} \times 5.0 \approx 0.632 \times 5.0 = 3.16$$

$$\sigma^2 = 1 - \bar{\alpha}_2 = 1 - 0.4 = 0.6$$

$$x_2 = \mu + \sqrt{\sigma^2}\epsilon = 3.16 + \sqrt{0.6}(0.5) \approx 3.16 + 0.774(0.5) = 3.547$$

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right)$$

$$x_{t-1} = \mu_\theta(x_t, t) + \sigma_t z$$

$$\beta_2 = 0.5, \quad \bar{\alpha}_2 = 0.4$$

$$\alpha_2 = 1 - \beta_2 = 1 - 0.5 = 0.5$$

$$\frac{\beta_2}{\sqrt{1 - \bar{\alpha}_2}} = \frac{0.5}{\sqrt{1 - 0.4}} = \frac{0.5}{\sqrt{0.6}} \approx \frac{0.5}{0.7746} \approx 0.6455$$

$$0.6455 \times 0.8 = 0.5164$$

$$3.0 - 0.5164 = 2.4836$$

$$\mu_\theta(x_2, 2) = \frac{1}{\sqrt{0.5}} \times 2.4836 \approx 1.4142 \times 2.4836 \approx 3.512$$

$$x_1 = \mu_\theta(x_2, 2) + \sigma_2 z = 3.512 + 0 = 3.512$$

The Future of Probability

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i))$$

$$P(R, S, W) = P(R)P(S \mid R)P(W \mid R, S)$$

$$P(W = \text{True}) = 0.00198 + 0.1782 + 0.288 + 0 = 0.46818$$

$$V_{k+1}(s) = \max_{a \in A} \sum_{s' \in S} P(s' \mid s, a) \left[R(s, a, s') + \gamma V_k(s') \right]$$

$$P(Y = y \mid do(X = x)) = \sum_{z \in Z} P(Y = y \mid X = x, Z = z) P(Z = z)$$

$$P(Y = 1 \mid do(X = 1)) = \sum_{z \in \{0,1\}} P(Y = 1 \mid X = 1, Z = z) P(Z = z)$$

$$P(Y = 1 \mid do(X = 1)) = 0.35 + 0.09 = 0.44$$

$$P(y_i) = \frac{\exp(z_i / T)}{\sum_j \exp(z_j / T)}$$

$$\text{Sum} = 54.60 + 7.39 + 0.14 = 62.13$$

Final Exam Review

$$\begin{aligned} \text{score(Spam)} &= P(Y = \text{Spam}) P(X_{\text{buy}} = 1 \mid Y = \text{Spam}) P(X_{\text{meeting}} = 0 \mid Y = \text{Spam}) \\ &= \frac{4}{7} \cdot \frac{3}{5} \cdot \frac{4}{5} = \frac{48}{175} \approx 0.274 \end{aligned}$$

$$\begin{aligned} \text{score(Ham)} &= P(Y = \text{Ham}) P(X_{\text{buy}} = 1 \mid Y = \text{Ham}) P(X_{\text{meeting}} = 0 \mid Y = \text{Ham}) \\ &= \frac{3}{7} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{56} \approx 0.054 \end{aligned}$$

CHAPTER 3

CS221 Ai

ML — Differentiable Programming and Generalization

$$
\begin{aligned}
f(\mathbf{x}_1) &= (3)(1) + (1)(0) = 3 \implies \text{loss}_1 = (3-3)^2 = 0 \\
f(\mathbf{x}_2) &= (3)(1) + (1)(2) = 5 \implies \text{loss}_2 = (5-5)^2 = 0
\end{aligned}
$$

$$
\begin{aligned}
f(\mathbf{x}_1) &= (1)(1) + (2)(0) = 1 \implies \text{loss}_1 = (3-1)^2 = 4 \\
f(\mathbf{x}_2) &= (1)(1) + (2)(2) = 5 \implies \text{loss}_2 = (5-5)^2 = 0
\end{aligned}
$$

ML — Best Practices and K-Means Clustering

$$
\begin{aligned}
\mu_1 &= \frac{1+2}{2} = 1.5 \\
\mu_2 &= \frac{6+7+8}{3} = 7
\end{aligned}
$$

$$
\begin{aligned}
\mu_1 &= \frac{1+3}{2} = 2 \\
\mu_2 &= \frac{8+10}{2} = 9
\end{aligned}
$$

Search 2 — A* Search

$$0 \le h(s) \le h^*(s) \quad \text{for all states } s$$

$$h(s) \le \text{Cost}(s,a) + h(s') \quad \text{for all states } s \text{ and actions } a, \text{ where } s' = \text{Succ}(s,a)$$

$$h(s_{\text{goal}}) = 0 \quad \text{for all goal states}$$

$$h_1(s) \ge h_2(s) \quad \text{for all states } s$$

MDPs 2 – Reinforcement Learning

$$
\begin{aligned}
N(s_t, a_t) &\leftarrow N(s_t, a_t) + 1 \\
Q(s_t, a_t) &\leftarrow Q(s_t, a_t) + \frac{1}{N(s_t, a_t)}(u_t - Q(s_t, a_t))
\end{aligned}
$$

$$
\begin{aligned}
Q(A, \text{Right}) &\leftarrow (1-0.1)(5) + (0.1)(9) = 4.5 + 0.9 = 5.4 \\
Q(B, \text{Down}) &\leftarrow (1-0.1)(0) + (0.1)(-2) = 0 - 0.2 = -0.2
\end{aligned}
$$

$$\text{Target} = r + \gamma Q(B, a_2) = 4 + 1 \times 3 = 7$$
$$Q(A, a_1) \leftarrow (1 - 0.5)(2) + 0.5(7) = 1 + 3.5 = 4.5$$

$$\text{Target} = r + \gamma \max_{a'} Q(B, a') = 4 + 1 \times \max(5, 3) = 4 + 5 = 9$$
$$Q(A, a_1) \leftarrow (1 - 0.5)(2) + 0.5(9) = 1 + 4.5 = 5.5$$

$$\text{Target} = r + \gamma \max_{a'} Q_{\mathbf{w}}(S_2, a') = 10 + 0.8(5) = 10 + 4 = 14$$
$$\text{Error} = \text{Target} - Q_{\mathbf{w}}(S_1, A_1) = 14 - 5 = 9$$

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{Error})\boldsymbol{\phi}(S_1, A_1)$$
$$\mathbf{w} \leftarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (0.1)(9) \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$
$$\mathbf{w} \leftarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2.7 \\ 0.9 \end{bmatrix} = \begin{bmatrix} 3.7 \\ 2.9 \end{bmatrix}$$

Game Playing 1: Minimax and Alpha-Beta Pruning

$$v = \max(v, \text{Minimax}(s'))$$

$$v = \min(v, \text{Minimax}(s'))$$

$$v(B) = \min(3, 5) = 3$$

$$v(C) = \min(2, 9) = 2$$

$$v(A) = \max(v(B), v(C)) = \max(3, 2) = 3$$

$$v = \max(v, \text{AlphaBeta}(s', \alpha, \beta))$$

$$v = \min(v, \text{AlphaBeta}(s', \alpha, \beta))$$

$$v = \max(v, \text{Expectimax}(s'))$$

$$v = v + P(s') \cdot \text{Expectimax}(s')$$

$$v(B) = (0.5 \times 4) + (0.5 \times 8) = 2 + 4 = 6$$

$$v(C) = (0.1 \times 10) + (0.9 \times 2) = 1 + 1.8 = 2.8$$

$$v(A) = \max(v(B), v(C)) = \max(6, 2.8) = 6$$

$$Eval(s) = w_1 f_1(s) + w_2 f_2(s) + \cdots + w_n f_n(s)$$

$$Eval(S_1) = 3(2) + 1(4) = 6 + 4 = 10$$

$$Eval(S_2) = 3(-1) + 1(10) = -3 + 10 = 7$$

$$Eval(S_3) = 3(0) + 1(8) = 0 + 8 = 8$$

$$v = \max(10, 7, 8) = 10$$

Bayesian Networks 2: Probabilistic Inference and Filtering

$$\begin{aligned} f_{new2}(C=1) &= f_{new1}(0)f_C(0,1) + f_{new1}(1)f_C(1,1) \\ &= (0.52)(0.1) + (0.48)(0.9) \\ &= 0.052 + 0.432 = 0.484 \end{aligned}$$

$$\begin{aligned} P(H_1 = Sun) &= P(Sun \mid Sun)P(H_0 = Sun) + P(Sun \mid Rain)P(H_0 = Rain) \\ &= (0.8)(0.5) + (0.4)(0.5) = 0.4 + 0.2 = 0.6 \\ P(H_1 = Rain) &= P(Rain \mid Sun)P(H_0 = Sun) + P(Rain \mid Rain)P(H_0 = Rain) \\ &= (0.2)(0.5) + (0.6)(0.5) = 0.1 + 0.3 = 0.4 \end{aligned}$$

$$\begin{aligned} F_1(Sun) &= P(U_1 = True \mid Sun)P(H_1 = Sun) = (0.2)(0.6) = 0.12 \\ F_1(Rain) &= P(U_1 = True \mid Rain)P(H_1 = Rain) = (0.9)(0.4) = 0.36 \end{aligned}$$

$$\begin{aligned} P(H_1 = Sun \mid U_1 = True) &= \frac{0.12}{0.48} = 0.25 \\ P(H_1 = Rain \mid U_1 = True) &= \frac{0.36}{0.48} = 0.75 \end{aligned}$$

Bayesian Networks 3: Supervised Learning, Smoothing and EM

$$\begin{aligned} P(H = \text{Head} \mid E = 1) &= \frac{0.48}{0.68} \approx 0.706 \\ P(H = \text{Tail} \mid E = 1) &= \frac{0.20}{0.68} \approx 0.294 \end{aligned}$$

$$\begin{aligned} P(H = \text{Head} \mid E = 0) &= \frac{0.12}{0.32} = 0.375 \\ P(H = \text{Tail} \mid E = 0) &= \frac{0.20}{0.32} = 0.625 \end{aligned}$$

Logic 1: Propositional Logic — Overview, Syntax and Semantics

$$\phi = (0 \lor \neg 1) \to (1 \land 0)$$

$$\neg 1 = 0 \implies \phi = (0 \lor 0) \to (1 \land 0)$$

$$(0 \lor 0) = 0$$

$$(1 \land 0) = 0$$

$$\phi = 0 \to 0$$

$$0 \to 0 = 1$$

$$\neg(\neg A \lor (B \land C))$$

$$\neg(\neg A) \land \neg(B \land C)$$

$$A \land \neg(B \land C)$$

$$A \land (\neg B \lor \neg C)$$

$$(A \lor X) \land (B \lor X)$$

$$(A \lor (C \land D)) \land (B \lor (C \land D))$$

$$(A \lor C) \land (A \lor D)$$

$$(B \lor C) \land (B \lor D)$$

$$(A \lor C) \land (A \lor D) \land (B \lor C) \land (B \lor D)$$

Logic 2: Inference Rules, Modus Ponens and Resolution

$$(A \to (B \lor C)) \land ((B \lor C) \to A)$$

$$(\neg A \lor B \lor C) \land (\neg(B \lor C) \lor A)$$

$$(\neg A \lor B \lor C) \land ((\neg B \land \neg C) \lor A)$$

$$((\neg B \land \neg C) \lor A) \equiv (A \lor \neg B) \land (A \lor \neg C)$$

$$(\neg A \lor B \lor C) \land (A \lor \neg B) \land (A \lor \neg C)$$

$$l_1 \lor \cdots \lor l_k \lor m_1 \lor \cdots \lor m_n$$

CHAPTER 4

CS224R Drl

PPO and Soft Actor-Critic

$$\begin{aligned}\mathrm{KL} &= 0.5\ln\left(\frac{0.5}{0.6}\right) + 0.5\ln\left(\frac{0.5}{0.4}\right)\\ &= 0.5\ln(0.8333) + 0.5\ln(1.25)\\ &\approx 0.5(-0.1823) + 0.5(0.2231)\\ &\approx -0.09115 + 0.11155\\ &= 0.0204\end{aligned}$$

$$\begin{aligned}\mathcal{H} &= -(0.8\ln(0.8) + 0.2\ln(0.2))\\ &= -(0.8\times(-0.2231) + 0.2\times(-1.6094))\\ &= -(-0.1785 - 0.3219)\\ &= 0.5004\end{aligned}$$

$$\begin{aligned}V_{\text{target}}(s_{t+1}) &= \min_{i=1,2} Q_{\text{target},i}(s_{t+1}, a_{t+1}) - \alpha \ln \pi(a_{t+1} \mid s_{t+1})\\ &= 9.5 + 0.4605 = 9.9605\end{aligned}$$

$$\begin{aligned}y_t &= r_t + \gamma V_{\text{target}}(s_{t+1})\\ &= 1.0 + 0.99\times 9.9605\\ &= 1.0 + 9.8609\\ &= 10.8609\end{aligned}$$

Offline Reinforcement Learning

$$\begin{aligned}w_1 &= 0.7\times\exp(2.0/1.0) = 0.7\times 7.389 = 5.172\\ w_2 &= 0.2\times\exp(5.0/1.0) = 0.2\times 148.413 = 29.683\\ w_3 &= 0.1\times\exp(1.0/1.0) = 0.1\times 2.718 = 0.272\end{aligned}$$

$$\begin{aligned}\pi^*(a_1) &= \frac{5.172}{35.127} \approx 0.147\\ \pi^*(a_2) &= \frac{29.683}{35.127} \approx 0.845\\ \pi^*(a_3) &= \frac{0.272}{35.127} \approx 0.008\end{aligned}$$

$$
\begin{aligned}
e^{Q(s,a_1)} &= e^{2.0} \approx 7.389 \\
e^{Q(s,a_2)} &= e^{4.0} \approx 54.598 \\
e^{Q(s,a_3)} &= e^{1.0} \approx 2.718
\end{aligned}
$$

$$
\begin{aligned}
L_{CQL} &= \alpha \left(\log \sum_a \exp(Q(s,a)) - Q(s,a_1) \right) \\
&= 0.5 \times (4.170 - 2.0) = 0.5 \times 2.170 = 1.085
\end{aligned}
$$

$$
\begin{aligned}
\text{Loss} &= -w_1 \log \pi(a_1|s) - w_2 \log \pi(a_2|s) \\
&= -2.138 \log(0.4) - 0.106 \log(0.6) \\
&= -2.138(-0.916) - 0.106(-0.511) \\
&= 1.958 + 0.054 = 2.012
\end{aligned}
$$

Model-Based RL (cont.) and Multi-Task RL

$$
\begin{aligned}
Q(s_1,a_1) &= Q(s_1,a_1) + \alpha \left(r + \gamma \max_{a'} Q(s_2,a') - Q(s_1,a_1) \right) \\
&= 2.0 + 0.5\,(3 + 0.9 \max(1.0, 5.0) - 2.0) \\
&= 2.0 + 0.5\,(3 + 4.5 - 2.0) \\
&= 2.0 + 0.5(5.5) = 4.75
\end{aligned}
$$

$$
\begin{aligned}
Q(s_1,a_1) &= 4.75 + 0.5 \left(3 + 0.9 \max_{a'} Q(s_2,a') - 4.75 \right) \\
&= 4.75 + 0.5\,(3 + 4.5 - 4.75) \\
&= 4.75 + 0.5(2.75) \\
&= 4.75 + 1.375 = 6.125
\end{aligned}
$$

$$
\begin{aligned}
p_1 &= \frac{1+1}{2} = 1.0 \\
p_2 &= \frac{0+1}{2} = 0.5
\end{aligned}
$$

$$
\begin{aligned}
g_1' &= g_1 - \frac{g_1 \cdot g_2}{\|g_2\|^2} g_2 \\
&= \begin{bmatrix} 2 \\ -1 \end{bmatrix} - \frac{-10}{25} \begin{bmatrix} -3 \\ 4 \end{bmatrix} \\
&= \begin{bmatrix} 2 \\ -1 \end{bmatrix} + 0.4 \begin{bmatrix} -3 \\ 4 \end{bmatrix} \\
&= \begin{bmatrix} 2 - 1.2 \\ -1 + 1.6 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
g_2' &= g_2 - \frac{g_2 \cdot g_1}{\|g_1\|^2} g_1 \\
&= \begin{bmatrix} -3 \\ 4 \end{bmatrix} - \frac{-10}{5} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} -3 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} -3 + 4 \\ 4 - 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\end{aligned}
$$

Exploration in Reinforcement Learning

$$c\sqrt{\frac{\ln t}{N(1)}} = 2\sqrt{\frac{\ln 4}{1}} \approx 2 \times 1.177 = 2.354$$
$$c\sqrt{\frac{\ln t}{N(2)}} = 2\sqrt{\frac{\ln 4}{2}} \approx 2 \times 0.832 = 1.664$$

$$UCB(1) = 5.0 + 2.354 = 7.354$$
$$UCB(2) = 6.0 + 1.664 = 7.664$$

$$r_i(s, \mathrm{Up}) = \frac{1.0}{\sqrt{4}} = 0.5$$
$$r_i(s, \mathrm{Down}) = \frac{1.0}{\sqrt{1}} = 1.0$$

$$r'(s, \mathrm{Up}) = 0 + 0.5 = 0.5$$
$$r'(s, \mathrm{Down}) = 0 + 1.0 = 1.0$$

$$\begin{aligned} Q(s, \mathrm{Down}) &\leftarrow Q(s, \mathrm{Down}) + \alpha \left[r'(s, \mathrm{Down}) + \gamma \max_{a'} Q(s', a') - Q(s, \mathrm{Down}) \right] \\ &\leftarrow 0 + 0.1\,[1.0 + 0.9(2.0) - 0] \\ &\leftarrow 0.1\,[1.0 + 1.8] = 0.1 \times 2.8 = 0.28 \end{aligned}$$

$$\ln \pi(a_1|s) = \ln(0.6) \approx -0.511$$
$$\ln \pi(a_2|s) = \ln(0.4) \approx -0.916$$

$$\begin{aligned} \text{Term for } a_1 : &\quad 2.0 - 0.5(-0.511) = 2.0 + 0.2555 = 2.2555 \\ \text{Term for } a_2 : &\quad 1.5 - 0.5(-0.916) = 1.5 + 0.458 = 1.958 \end{aligned}$$

$$\begin{aligned} V_{\mathrm{soft}}(s) &= \pi(a_1|s)(2.2555) + \pi(a_2|s)(1.958) \\ &= 0.6(2.2555) + 0.4(1.958) \\ &= 1.3533 + 0.7832 = 2.1365 \end{aligned}$$

$$N(A_t) \leftarrow N(A_t) + 1$$

$$Q(A_t) \leftarrow Q(A_t) + \frac{1}{N(A_t)}(R_t - Q(A_t))$$

$$A_8 = \arg\max_a \{2.5, 1.0, 3.2\} = 3$$

$$N(3) = 4 + 1 = 5$$

$$Q(3) = 3.2 + \frac{1}{5}(4.0 - 3.2) = 3.2 + 0.16 = 3.36$$

$$UCB(a) = Q(a) + c\sqrt{\frac{\ln t}{N(a)}}$$

$$N(A_t) \leftarrow N(A_t) + 1, \quad Q(A_t) \leftarrow Q(A_t) + \frac{1}{N(A_t)}(R_t - Q(A_t))$$

$$A_4 = \arg\max\{7.354, 7.664\} = 2$$

$$Q(2) = 6.0 + \frac{1}{3}(4.0 - 6.0) = 6.0 - 0.667 = 5.333$$

$$\theta_a \sim \text{Beta}(\alpha_a, \beta_a)$$

$$A_t = \arg\max\{0.45, 0.82\} = 2$$

$$\beta_2 = 1 + 1 = 2$$

$$N(s,a) \leftarrow N(s,a) + 1$$

$$r_i(s,a) = \frac{\beta}{\sqrt{N(s,a)}}$$

$$r'(s,a) = r_e(s,a) + r_i(s,a)$$

$$N(s, \text{Down}) = 1 + 1 = 2$$

$$r_i = \|f_\phi(s_t, a_t) - s_{t+1}\|_2^2$$

$$r_t = r_e + \lambda r_i$$

$$L(\phi) = \|f_\phi(s_t, a_t) - s_{t+1}\|_2^2$$

$$e = \begin{bmatrix} 1.2 \\ -0.4 \end{bmatrix} - \begin{bmatrix} 1.0 \\ -0.2 \end{bmatrix} = \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix}$$

$$r_i = \|e\|_2^2 = (0.2)^2 + (-0.2)^2 = 0.04 + 0.04 = 0.08$$

$$r_t = r_e + \lambda r_i = 0.1 + 0.5(0.08) = 0.1 + 0.04 = 0.14$$

$$r_i = \|\hat{f}_\theta(s_{t+1}) - f(s_{t+1})\|_2^2$$

$$L(\theta) = \mathbb{E}_{s\sim\mathcal{D}}\left[\|\hat{f}_\theta(s) - f(s)\|_2^2\right]$$

$$d = \hat{f}_\theta(s) - f(s) = \begin{bmatrix} 0.3 - 0.5 \\ 0.2 - (-0.1) \\ 0.8 - 0.8 \end{bmatrix} = \begin{bmatrix} -0.2 \\ 0.3 \\ 0.0 \end{bmatrix}$$

$$r_i = \|d\|_2^2 = (-0.2)^2 + (0.3)^2 + (0.0)^2 = 0.04 + 0.09 + 0 = 0.13$$

$$J(\pi) = \mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t \Big(r(s_t, a_t) + \alpha\mathcal{H}(\pi(\cdot|s_t))\Big)\right]$$

$$Q_{\text{soft}}(s_t, a_t) = r(s_t, a_t) + \gamma\mathbb{E}_{s_{t+1}\sim P}[V_{\text{soft}}(s_{t+1})]$$

$$V_{\text{soft}}(s_t) = \mathbb{E}_{a_t\sim\pi}[Q_{\text{soft}}(s_t, a_t) - \alpha \ln \pi(a_t|s_t)]$$

$$\pi^*(a_t|s_t) \propto \exp\left(\frac{1}{\alpha} Q_{\text{soft}}(s_t, a_t)\right)$$

Open Problems and How to Do RL Research

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) \cdot G_t$$

$$\theta_{t+1} = \theta_t + \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) \cdot G_t$$

Q-Learning and Value-Based Methods Review

$$Q_1(s_1, \text{Move}) = 1 + 0.9 \cdot V_0(s_2) = 1 + 0 = 1$$
$$Q_1(s_1, \text{Stay}) = 0 + 0.9 \cdot V_0(s_1) = 0 + 0 = 0$$

$$Q_1(s_2, \text{Move}) = 10 + 0.9 \cdot V_0(s_3) = 10 + 0 = 10$$
$$Q_1(s_2, \text{Stay}) = 0 + 0.9 \cdot V_0(s_2) = 0 + 0 = 0$$

$$Q_2(s_1, \text{Move}) = 1 + 0.9 \cdot V_1(s_2) = 1 + 0.9 \cdot 10 = 10$$
$$Q_2(s_1, \text{Stay}) = 0 + 0.9 \cdot V_1(s_1) = 0 + 0.9 \cdot 1 = 0.9$$

$$Q_2(s_2, \text{Move}) = 10 + 0.9 \cdot V_1(s_3) = 10 + 0.9 \cdot 0 = 10$$
$$Q_2(s_2, \text{Stay}) = 0 + 0.9 \cdot V_1(s_2) = 0 + 0.9 \cdot 10 = 9$$

$$Q(B, a_1) = 5.0$$
$$Q(B, a_2) = 8.0$$

$$y = r + \gamma \max_{a'} Q(B, a')$$
$$y = 3.0 + 0.8 \cdot 8.0 = 3.0 + 6.4 = 9.4$$

$$Q(A, a_1) \leftarrow Q(A, a_1) + \alpha \left[y - Q(A, a_1)\right]$$
$$Q(A, a_1) \leftarrow 4.0 + 0.5 \left[9.4 - 4.0\right]$$
$$Q(A, a_1) \leftarrow 4.0 + 0.5 \left[5.4\right]$$
$$Q(A, a_1) \leftarrow 4.0 + 2.7 = 6.7$$

$$y = r + \gamma Q(B, a_1)$$
$$y = 3.0 + 0.8 \cdot 5.0 = 3.0 + 4.0 = 7.0$$

$$Q(A, a_1) \leftarrow Q(A, a_1) + \alpha \left[y - Q(A, a_1)\right]$$
$$Q(A, a_1) \leftarrow 4.0 + 0.5 \left[7.0 - 4.0\right]$$
$$Q(A, a_1) \leftarrow 4.0 + 0.5 \left[3.0\right] = 5.5$$

$$\max_{a'} Q_{\theta^-}(s'_1, a') = \max(3.0, 5.0, 1.0) = 5.0$$
$$y_1 = r_1 + \gamma \max_{a'} Q_{\theta^-}(s'_1, a')$$
$$y_1 = 2.0 + 0.9 \cdot 5.0 = 2.0 + 4.5 = 6.5$$

$$\begin{aligned}\max_{a'} Q_{\theta^-}(s'_2, a') &= \max(0.0, 0.5, 0.2) = 0.5 \\ y_2 &= r_2 + \gamma \max_{a'} Q_{\theta^-}(s'_2, a') \\ y_2 &= -1.0 + 0.9 \cdot 0.5 = -1.0 + 0.45 = -0.55\end{aligned}$$

$$y_3 = r_3 = 10.0$$

$$\begin{aligned}\text{Error}_1^2 &= (y_1 - Q_\theta(s_1, a_1))^2 = (6.5 - 4.0)^2 = (2.5)^2 = 6.25 \\ \text{Error}_2^2 &= (y_2 - Q_\theta(s_2, a_2))^2 = (-0.55 - (-0.5))^2 = (-0.05)^2 = 0.0025 \\ \text{Error}_3^2 &= (y_3 - Q_\theta(s_3, a_3))^2 = (10.0 - 8.0)^2 = (2.0)^2 = 4.0\end{aligned}$$

$$\begin{aligned}L(\theta) &= \frac{1}{3}(6.25 + 0.0025 + 4.0) \\ L(\theta) &= \frac{1}{3}(10.2525) = 3.4175\end{aligned}$$

$$\begin{aligned}\max_{a'} Q_{\theta^-}(s', a') &= \max(7.0, 4.0, 9.0) = 9.0 \\ y_{\text{DQN}} &= r + \gamma \cdot 9.0 = 1.0 + 0.9 \cdot 9.0 = 1.0 + 8.1 = 9.1\end{aligned}$$

$$a^* = \arg\max_{a'} Q_\theta(s', a') = \arg\max(5.0, 8.0, 6.0) = a_2$$

$$Q_{\text{eval}} = Q_{\theta^-}(s', a_2) = 4.0$$

$$y_{\text{DDQN}} = r + \gamma \cdot Q_{\text{eval}} = 1.0 + 0.9 \cdot 4.0 = 1.0 + 3.6 = 4.6$$

CHAPTER 5

CS229 Machine Learning

Gaussian Discriminant Analysis and Naive Bayes

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

$$\phi = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\}$$

$$\mu_0 = \frac{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 0\}x^{(i)}}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 0\}}$$

$$\mu_1 = \frac{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\}x^{(i)}}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\}}$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$$

$$P_0 = (1-\phi)\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)$$

$$P_1 = \phi\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)$$

$$\phi = \frac{2}{4} = 0.5$$

$$\mu_0 = \frac{1+3}{2} = 2$$

$$\mu_1 = \frac{5+7}{2} = 6$$

$$\sigma^2 = \frac{1}{4}\left[(1-2)^2 + (3-2)^2 + (5-6)^2 + (7-6)^2\right] = \frac{1}{4}[1+1+1+1] = 1$$

$$P_0 = (1-0.5)\frac{1}{\sqrt{2\pi(1)}}\exp\left(-\frac{1}{2(1)}(4-2)^2\right) = \frac{0.5}{\sqrt{2\pi}}e^{-2} \approx 0.027$$

$$P_1 = 0.5\frac{1}{\sqrt{2\pi(1)}}\exp\left(-\frac{1}{2(1)}(4-6)^2\right) = \frac{0.5}{\sqrt{2\pi}}e^{-2} \approx 0.027$$

$$p(x_1, \ldots, x_d|y) = \prod_{j=1}^{d} p(x_j|y)$$

$$\phi_y = \frac{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\}}{n}$$

$$\phi_{j|y=1} = \frac{\sum_{i=1}^{n} \mathbb{I}\{x_j^{(i)} = 1 \text{ and } y^{(i)} = 1\}}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\}}$$

$$\phi_{j|y=0} = \frac{\sum_{i=1}^{n} \mathbb{I}\{x_j^{(i)} = 1 \text{ and } y^{(i)} = 0\}}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 0\}}$$

$$P_1 = \phi_y \prod_{j=1}^{d} (\phi_{j|y=1})^{x_j} (1 - \phi_{j|y=1})^{1-x_j}$$

$$P_0 = (1 - \phi_y) \prod_{j=1}^{d} (\phi_{j|y=0})^{x_j} (1 - \phi_{j|y=0})^{1-x_j}$$

$$\phi_y = \frac{2}{5} = 0.4$$

$$\phi_{1|y=1} = \frac{2}{2} = 1.0, \quad \phi_{2|y=1} = \frac{0}{2} = 0.0$$

$$\phi_{1|y=0} = \frac{1}{3} \approx 0.33, \quad \phi_{2|y=0} = \frac{3}{3} = 1.0$$

$$P_1 = \phi_y (\phi_{1|y=1})^1 (1 - \phi_{1|y=1})^0 (\phi_{2|y=1})^0 (1 - \phi_{2|y=1})^1 = 0.4 \times 1.0 \times 1.0 = 0.4$$

$$P_0 = (1 - \phi_y)(\phi_{1|y=0})^1 (1 - \phi_{1|y=0})^0 (\phi_{2|y=0})^0 (1 - \phi_{2|y=0})^1 = 0.6 \times \left(\frac{1}{3}\right) \times 0.0 = 0$$

$$p(x_j = v | y = c) = \frac{\sum_{i=1}^{n} \mathbb{I}\{x_j^{(i)} = v \text{ and } y^{(i)} = c\} + 1}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = c\} + K}$$

$$\phi_{j|y=1} = \frac{\sum_{i=1}^{n} \mathbb{I}\{x_j^{(i)} = 1 \text{ and } y^{(i)} = 1\} + 1}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\} + 2}$$

$$\phi_{2|y=1} = \frac{0+1}{2+2} = 0.25$$

$$\phi_{1|y=0} = \frac{1+1}{3+2} = \frac{2}{5} = 0.4$$

$$\phi_{k|y=1} = \frac{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\} \sum_{j=1}^{d_i} \mathbb{I}\{x_j^{(i)} = k\} + 1}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 1\} d_i + |V|}$$

$$\phi_{k|y=0} = \frac{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 0\} \sum_{j=1}^{d_i} \mathbb{I}\{x_j^{(i)} = k\} + 1}{\sum_{i=1}^{n} \mathbb{I}\{y^{(i)} = 0\} d_i + |V|}$$

$$P_1 = \phi_y \prod_{j=1}^{d} \phi_{x_j|y=1}$$

$$P_0 = (1 - \phi_y) \prod_{j=1}^{d} \phi_{x_j|y=0}$$

$$\phi_1 = \frac{1}{2} = 0.5, \quad \phi_0 = \frac{1}{2} = 0.5$$

$$\phi_{1|y=1} = \frac{2+1}{2+3} = \frac{3}{5} = 0.6$$

$$\phi_{2|y=1} = \frac{0+1}{2+3} = \frac{1}{5} = 0.2$$

$$\phi_{3|y=1} = \frac{0+1}{2+3} = \frac{1}{5} = 0.2$$

$$\phi_{1|y=0} = \frac{0+1}{2+3} = \frac{1}{5} = 0.2$$

$$\phi_{2|y=0} = \frac{1+1}{2+3} = \frac{2}{5} = 0.4$$

$$\phi_{3|y=0} = \frac{1+1}{2+3} = \frac{2}{5} = 0.4$$

$$P_1 = p(y=1) \cdot p(x_1 = 1|y=1) \cdot p(x_2 = 2|y=1) = 0.5 \times 0.6 \times 0.2 = 0.06$$

$$P_0 = p(y=0) \cdot p(x_1 = 1|y=0) \cdot p(x_2 = 2|y=0) = 0.5 \times 0.2 \times 0.4 = 0.04$$

Kernels

$$\begin{aligned}
\phi(x)^T\phi(z) &= \begin{bmatrix} x_1^2 & \sqrt{2}x_1x_2 & x_2^2 \end{bmatrix} \begin{bmatrix} z_1^2 \\ \sqrt{2}z_1z_2 \\ z_2^2 \end{bmatrix} \\
&= x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2
\end{aligned}$$

$$\begin{aligned}
K(x,z) &= (x^Tz)^2 \\
&= (x_1z_1 + x_2z_2)^2 \\
&= (x_1z_1)^2 + 2(x_1z_1)(x_2z_2) + (x_2z_2)^2 \\
&= x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2
\end{aligned}$$

$$\begin{aligned}
x^Tz &= (1)(3) + (2)(1) = 3 + 2 = 5 \\
||x - z||^2 &= (1-3)^2 + (2-1)^2 = (-2)^2 + (1)^2 = 4 + 1 = 5
\end{aligned}$$

$$\begin{aligned}
\min_{w,b} \quad & \frac{1}{2}||w||^2 \\
\text{s.t.} \quad & y^{(i)}(w^Tx^{(i)} + b) \geq 1 \quad \text{for } i = 1, \ldots, m
\end{aligned}$$

$$\begin{aligned}
\min_{w,b,\xi} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i \\
\text{s.t.} \quad & y^{(i)}(w^Tx^{(i)} + b) \geq 1 - \xi_i \\
& \xi_i \geq 0 \quad \text{for } i = 1, \ldots, m
\end{aligned}$$

Learning Theory

$$\begin{aligned}
\exp(-2\gamma^2 m) &\leq \frac{\delta}{2} \\
-2\gamma^2 m &\leq \log\left(\frac{\delta}{2}\right) \\
m &\geq -\frac{1}{2\gamma^2}\log\left(\frac{\delta}{2}\right) = \frac{1}{2\gamma^2}\log\left(\frac{2}{\delta}\right)
\end{aligned}$$

$$\begin{aligned}
m &\geq \frac{1}{2(0.05)^2}\log\left(\frac{2}{0.01}\right) \\
m &\geq \frac{1}{0.005}\log(200) \\
m &\geq 200 \times 5.298 = 1059.6
\end{aligned}$$

$$
\begin{aligned}
\gamma &= \sqrt{\frac{1}{2m}\log\left(\frac{2k}{\delta}\right)} \\
\gamma &= \sqrt{\frac{1}{2(5000)}\log\left(\frac{2(100)}{0.05}\right)} \\
\gamma &= \sqrt{\frac{1}{10000}\log\left(\frac{200}{0.05}\right)} \\
\gamma &= \sqrt{\frac{1}{10000}\log(4000)}
\end{aligned}
$$

$$
\begin{aligned}
m &\geq \frac{1}{2(0.02)^2}\log\left(\frac{2\times 10^4}{0.01}\right) \\
m &\geq \frac{1}{0.0008}\log(2\times 10^6)
\end{aligned}
$$

Expectation-Maximization Algorithms

$$
\begin{aligned}
\phi_k &= \frac{1}{m}\sum_{i=1}^{m} w_k^{(i)} \\
\mu_k &= \frac{\sum_{i=1}^{m} w_k^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_k^{(i)}} \\
\Sigma_k &= \frac{\sum_{i=1}^{m} w_k^{(i)} (x^{(i)}-\mu_k)(x^{(i)}-\mu_k)^T}{\sum_{i=1}^{m} w_k^{(i)}}
\end{aligned}
$$

$$
\begin{aligned}
\text{Expected Heads} &= Q_1(A)\times 3 + Q_2(A)\times 1 \\
&= 0.633\times 3 + 0.434\times 1 = 1.899 + 0.434 = 2.333 \\
\text{Expected Total Flips} &= Q_1(A)\times 3 + Q_2(A)\times 3 \\
&= 0.633\times 3 + 0.434\times 3 = 1.899 + 1.302 = 3.201
\end{aligned}
$$

$$
\begin{aligned}
\text{Expected Heads} &= Q_1(B)\times 3 + Q_2(B)\times 1 \\
&= 0.367\times 3 + 0.566\times 1 = 1.101 + 0.566 = 1.667 \\
\text{Expected Total Flips} &= Q_1(B)\times 3 + Q_2(B)\times 3 \\
&= 0.367\times 3 + 0.566\times 3 = 1.101 + 1.698 = 2.799
\end{aligned}
$$

Independent Component Analysis and Introduction to RL

$$
\begin{aligned}
w_1^T x &= (1)(1) + (0)(0.5) = 1 \\
w_2^T x &= (0)(1) + (1)(0.5) = 0.5
\end{aligned}
$$

$$
\begin{aligned}
g(1) &= \frac{1}{1+e^{-1}} \approx 0.7311 \\
1 - 2g(1) &= 1 - 2(0.7311) = -0.4622 \\
g(0.5) &= \frac{1}{1+e^{-0.5}} \approx 0.6225 \\
1 - 2g(0.5) &= 1 - 2(0.6225) = -0.2450
\end{aligned}
$$

$$
\begin{aligned}
W_{new} &= W + \alpha \left(vx^T + (W^T)^{-1} \right) \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 0.1 \left(\begin{bmatrix} -0.4622 & -0.2311 \\ -0.2450 & -0.1225 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 0.1 \begin{bmatrix} 0.5378 & -0.2311 \\ -0.2450 & 0.8775 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.0538 & -0.0231 \\ -0.0245 & 0.0878 \end{bmatrix} = \begin{bmatrix} 1.0538 & -0.0231 \\ -0.0245 & 1.0878 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
V_1(S_1) &= R(S_1) + \gamma \max \begin{cases} P_{S_1 a_1}(S_1) V_0(S_1) + P_{S_1 a_1}(S_2) V_0(S_2) \\ P_{S_1 a_2}(S_1) V_0(S_1) + P_{S_1 a_2}(S_2) V_0(S_2) \end{cases} \\
&= 1 + 0.5 \max(1.0(0) + 0.0(0),\ 0.0(0) + 1.0(0)) \\
&= 1 + 0.5 \max(0, 0) = 1
\end{aligned}
$$

$$
V_1(S_2) = 0 + 0.5 \max(0.5(0) + 0.5(0),\ 1.0(0) + 0.0(0)) = 0
$$

$$
\begin{aligned}
V_2(S_1) &= 1 + 0.5 \max \begin{cases} 1.0(V_1(S_1)) + 0.0(V_1(S_2)) & (\text{Action } a_1) \\ 0.0(V_1(S_1)) + 1.0(V_1(S_2)) & (\text{Action } a_2) \end{cases} \\
&= 1 + 0.5 \max(1.0(1) + 0,\ 0 + 1.0(0)) \\
&= 1 + 0.5 \max(1, 0) = 1 + 0.5 = 1.5
\end{aligned}
$$

$$
\begin{aligned}
V_2(S_2) &= 0 + 0.5 \max \begin{cases} 0.5(V_1(S_1)) + 0.5(V_1(S_2)) & (\text{Action } a_1) \\ 1.0(V_1(S_1)) + 0.0(V_1(S_2)) & (\text{Action } a_2) \end{cases} \\
&= 0 + 0.5 \max(0.5(1) + 0,\ 1.0(1) + 0) \\
&= 0 + 0.5 \max(0.5, 1.0) = 0 + 0.5(1.0) = 0.5
\end{aligned}
$$

$$
\begin{aligned}
V^\pi(S_1) &= R(S_1) + \gamma \left(P_{S_1 a_2}(S_1) V^\pi(S_1) + P_{S_1 a_2}(S_2) V^\pi(S_2) \right) \\
&= 1 + 0.5 \left(0 \cdot V^\pi(S_1) + 1.0 \cdot V^\pi(S_2) \right) = 1 + 0.5 V^\pi(S_2) \\
V^\pi(S_2) &= R(S_2) + \gamma \left(P_{S_2 a_1}(S_1) V^\pi(S_1) + P_{S_2 a_1}(S_2) V^\pi(S_2) \right) \\
&= 0 + 0.5 \left(0.5 \cdot V^\pi(S_1) + 0.5 \cdot V^\pi(S_2) \right) = 0.25 V^\pi(S_1) + 0.25 V^\pi(S_2)
\end{aligned}
$$

$$
\begin{aligned}
Q(S_1, a_1) &= 1 + 0.5(1.0 \times 1.2 + 0.0 \times 0.4) = 1 + 0.5(1.2) = 1.6 \\
Q(S_1, a_2) &= 1 + 0.5(0.0 \times 1.2 + 1.0 \times 0.4) = 1 + 0.5(0.4) = 1.2
\end{aligned}
$$

$$
\begin{aligned}
Q(S_2, a_1) &= 0 + 0.5(0.5 \times 1.2 + 0.5 \times 0.4) = 0.5(0.6 + 0.2) = 0.4 \\
Q(S_2, a_2) &= 0 + 0.5(1.0 \times 1.2 + 0.0 \times 0.4) = 0.5(1.2) = 0.6
\end{aligned}
$$

Reward Models and Linear Dynamical Systems

$$
\begin{aligned}
Q(1, x) &= R(1) + \gamma \left(1 \cdot V_0(1) \right) = 1 + 0.5(0) = 1 \\
Q(1, y) &= R(1) + \gamma \left(1 \cdot V_0(2) \right) = 1 + 0.5(0) = 1
\end{aligned}
$$

$$
\begin{aligned}
Q(2,x) &= R(2) + \gamma\,(1 \cdot V_0(1)) = 2 + 0.5(0) = 2 \\
Q(2,y) &= R(2) + \gamma\,(1 \cdot V_0(2)) = 2 + 0.5(0) = 2
\end{aligned}
$$

$$
\begin{aligned}
Q(1,x) &= 1 + 0.5(V_1(1)) = 1 + 0.5(1) = 1.5 \\
Q(1,y) &= 1 + 0.5(V_1(2)) = 1 + 0.5(2) = 2.0
\end{aligned}
$$

$$
\begin{aligned}
Q(2,x) &= 2 + 0.5(V_1(1)) = 2 + 0.5(1) = 2.5 \\
Q(2,y) &= 2 + 0.5(V_1(2)) = 2 + 0.5(2) = 3.0
\end{aligned}
$$

$$
\begin{aligned}
V^{\pi_0}(1) &= R(1) + 0.5 \cdot V^{\pi_0}(1) \implies V^{\pi_0}(1) = 1 + 0.5V^{\pi_0}(1) \implies V^{\pi_0}(1) = 2 \\
V^{\pi_0}(2) &= R(2) + 0.5 \cdot V^{\pi_0}(1) \implies V^{\pi_0}(2) = 2 + 0.5(2) = 3
\end{aligned}
$$

$$
\begin{aligned}
Q(1,x) &= 1 + 0.5(2) = 2 \\
Q(1,y) &= 1 + 0.5(3) = 2.5
\end{aligned}
$$

$$
\begin{aligned}
Q(2,x) &= 2 + 0.5(2) = 3 \\
Q(2,y) &= 2 + 0.5(3) = 3.5
\end{aligned}
$$

$$
\begin{aligned}
V^{\pi_1}(2) &= 2 + 0.5 \cdot V^{\pi_1}(2) \implies 0.5V^{\pi_1}(2) = 2 \implies V^{\pi_1}(2) = 4 \\
V^{\pi_1}(1) &= 1 + 0.5 \cdot V^{\pi_1}(2) \implies V^{\pi_1}(1) = 1 + 0.5(4) = 3
\end{aligned}
$$

$$
\begin{aligned}
Q(1,x) &= 1 + 0.5(3) = 2.5 \\
Q(1,y) &= 1 + 0.5(4) = 3.0
\end{aligned}
$$

$$
\begin{aligned}
Q(2,x) &= 2 + 0.5(3) = 3.5 \\
Q(2,y) &= 2 + 0.5(4) = 4.0
\end{aligned}
$$

$$
\begin{aligned}
K_t &= -(B^\top V_{t+1} B + W)^{-1} B^\top V_{t+1} A \\
V_t &= U + A^\top V_{t+1} A + A^\top V_{t+1} B K_t
\end{aligned}
$$

$$
\begin{aligned}
a_t &= K_t s_t \\
s_{t+1} &= A s_t + B a_t
\end{aligned}
$$

$$
\begin{aligned}
K_1 &= -(B^\top V_2 B + W)^{-1} B^\top V_2 A \\
&= -(1 \cdot 2 \cdot 1 + 1)^{-1}(1 \cdot 2 \cdot 2) = -\frac{1}{3} \cdot 4 = -\frac{4}{3} \\
V_1 &= U + A^\top V_2 A + A^\top V_2 B K_1 \\
&= 1 + (2 \cdot 2 \cdot 2) + (2 \cdot 2 \cdot 1 \cdot (-\frac{4}{3})) \\
&= 1 + 8 - \frac{16}{3} = 9 - 5.333 = \frac{11}{3}
\end{aligned}
$$

$$
\begin{aligned}
K_0 &= -(B^\top V_1 B + W)^{-1} B^\top V_1 A \\
&= -\left(1 \cdot \frac{11}{3} \cdot 1 + 1\right)^{-1} \left(1 \cdot \frac{11}{3} \cdot 2\right) = -\left(\frac{14}{3}\right)^{-1} \left(\frac{22}{3}\right) = -\frac{3}{14} \cdot \frac{22}{3} = -\frac{11}{7} \\
V_0 &= U + A^\top V_1 A + A^\top V_1 B K_0 \\
&= 1 + \left(2 \cdot \frac{11}{3} \cdot 2\right) + \left(2 \cdot \frac{11}{3} \cdot 1 \cdot \left(-\frac{11}{7}\right)\right) \\
&= 1 + \frac{44}{3} - \frac{242}{21} = \frac{21 + 308 - 242}{21} = \frac{87}{21} = \frac{29}{7}
\end{aligned}
$$

$$
\begin{aligned}
a_0 &= K_0 s_0 = -\frac{11}{7}(1) \approx -1.571 \\
s_1 &= 2s_0 + a_0 = 2(1) - \frac{11}{7} = \frac{3}{7} \approx 0.429 \\
a_1 &= K_1 s_1 = -\frac{4}{3}\left(\frac{3}{7}\right) = -\frac{4}{7} \approx -0.571 \\
s_2 &= 2s_1 + a_1 = 2\left(\frac{3}{7}\right) - \frac{4}{7} = \frac{2}{7} \approx 0.286
\end{aligned}
$$

$$
\begin{aligned}
s_{t+1} &= A s_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, Q) \\
y_t &= C s_t + \delta_t, \quad \delta_t \sim \mathcal{N}(0, R)
\end{aligned}
$$

$$
\begin{aligned}
\mu_{t|t-1} &= A \mu_{t-1} \\
\Sigma_{t|t-1} &= A \Sigma_{t-1} A^\top + Q
\end{aligned}
$$

$$
\begin{aligned}
K_t &= \Sigma_{t|t-1} C^\top (C \Sigma_{t|t-1} C^\top + R)^{-1} \\
\mu_t &= \mu_{t|t-1} + K_t (y_t - C \mu_{t|t-1}) \\
\Sigma_t &= (I - K_t C) \Sigma_{t|t-1}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{1|0} &= A \mu_0 = 1(0) = 0 \\
\Sigma_{1|0} &= A \Sigma_0 A^\top + Q = 1(1)(1) + 1 = 2
\end{aligned}
$$

$$
\begin{aligned}
K_1 &= \Sigma_{1|0} C^\top (C \Sigma_{1|0} C^\top + R)^{-1} = 2(1)(1(2)1 + 2)^{-1} = \frac{2}{4} = 0.5 \\
\mu_1 &= \mu_{1|0} + K_1 (y_1 - C \mu_{1|0}) = 0 + 0.5(3 - 0) = 1.5 \\
\Sigma_1 &= (I - K_1 C) \Sigma_{1|0} = (1 - 0.5(1))(2) = 1
\end{aligned}
$$

$$
\begin{aligned}
\mu_{2|1} &= A \mu_1 = 1(1.5) = 1.5 \\
\Sigma_{2|1} &= A \Sigma_1 A^\top + Q = 1(1)(1) + 1 = 2
\end{aligned}
$$

$$
\begin{aligned}
K_2 &= \Sigma_{2|1} C^\top (C \Sigma_{2|1} C^\top + R)^{-1} = 2(1)(1(2)1 + 2)^{-1} = \frac{2}{4} = 0.5 \\
\mu_2 &= \mu_{2|1} + K_2 (y_2 - C \mu_{2|1}) = 1.5 + 0.5(2.4 - 1.5) = 1.5 + 0.5(0.9) = 1.95 \\
\Sigma_2 &= (I - K_2 C) \Sigma_{2|1} = (1 - 0.5(1))(2) = 1
\end{aligned}
$$

RL Debugging & Diagnostics

$$\theta \leftarrow \theta + \alpha G_0 \nabla_\theta \log \pi_\theta(a_1|s)$$
$$\theta \leftarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} + (0.1)(10) \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$$

$$\begin{aligned} P(S_1|a,b) &= P(S_1|S_1,a)b(S_1) + P(S_1|S_2,a)b(S_2) \\ &= (0.7)(0.6) + (0.4)(0.4) = 0.42 + 0.16 = 0.58 \\ P(S_2|a,b) &= P(S_2|S_1,a)b(S_1) + P(S_2|S_2,a)b(S_2) \\ &= (0.3)(0.6) + (0.6)(0.4) = 0.18 + 0.24 = 0.42 \end{aligned}$$

$$\begin{aligned} \tilde{b}'(S_1) &= P(O_1|S_1)P(S_1|a,b) = 0.8 \times 0.58 = 0.464 \\ \tilde{b}'(S_2) &= P(O_1|S_2)P(S_2|a,b) = 0.3 \times 0.42 = 0.126 \end{aligned}$$

$$\begin{aligned} \text{Total} &= \tilde{b}'(S_1) + \tilde{b}'(S_2) = 0.464 + 0.126 = 0.590 \\ b'(S_1) &= \frac{0.464}{0.590} \approx 0.786 \\ b'(S_2) &= \frac{0.126}{0.590} \approx 0.214 \end{aligned}$$

CHAPTER 6

CS231N Deep Learning For Computer Vision

Loss Functions and Optimization

$$
\begin{aligned}
L_i &= \max(0, s_1 - s_0 + \Delta) + \max(0, s_2 - s_0 + \Delta) \\
&= \max(0, 5.1 - 3.2 + 1.0) + \max(0, -1.7 - 3.2 + 1.0) \\
&= \max(0, 2.9) + \max(0, -3.9) \\
&= 2.9 + 0 = 2.9
\end{aligned}
$$

$$
\begin{aligned}
v &\leftarrow \rho v - \alpha \nabla W \\
W &\leftarrow W + v
\end{aligned}
$$

$$
\begin{aligned}
\mathit{cache} &\leftarrow \mathit{cache} + (\nabla W)^2 \\
W &\leftarrow W - \frac{\alpha}{\sqrt{\mathit{cache}} + \epsilon} \nabla W
\end{aligned}
$$

$$
\begin{aligned}
\mathit{cache} &\leftarrow \gamma \times \mathit{cache} + (1 - \gamma)(\nabla W)^2 \\
W &\leftarrow W - \frac{\alpha}{\sqrt{\mathit{cache}} + \epsilon} \nabla W
\end{aligned}
$$

$$
\begin{aligned}
m &\leftarrow \beta_1 m + (1 - \beta_1) \nabla W \\
v &\leftarrow \beta_2 v + (1 - \beta_2)(\nabla W)^2
\end{aligned}
$$

$$
\begin{aligned}
\hat{m} &= \frac{m}{1 - \beta_1^t} \\
\hat{v} &= \frac{v}{1 - \beta_2^t}
\end{aligned}
$$

Neural Networks and Backpropagation

$$
\begin{aligned}
q &= x + y = -2 + 5 = 3 \\
f &= qz = 3 \times (-4) = -12
\end{aligned}
$$

$$
\begin{aligned}
\text{Local gradients:} \quad & \frac{\partial f}{\partial q} = z = -4, \quad \frac{\partial f}{\partial z} = q = 3 \\
\text{Downstream gradient for } q: \quad & \frac{\partial f}{\partial q} = \frac{\partial f}{\partial f} \times \frac{\partial f}{\partial q} = 1 \times (-4) = -4 \\
\text{Downstream gradient for } z: \quad & \frac{\partial f}{\partial z} = \frac{\partial f}{\partial f} \times \frac{\partial f}{\partial z} = 1 \times 3 = 3
\end{aligned}
$$

$$
\begin{aligned}
\text{Upstream gradient: } & \frac{\partial f}{\partial q} = -4 \\
\text{Local gradients: } & \frac{\partial q}{\partial x} = 1, \quad \frac{\partial q}{\partial y} = 1 \\
\text{Downstream gradient for } x: & \frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \times \frac{\partial q}{\partial x} = -4 \times 1 = -4 \\
\text{Downstream gradient for } y: & \frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \times \frac{\partial q}{\partial y} = -4 \times 1 = -4
\end{aligned}
$$

$$
\begin{aligned}
Z_1 &= XW_1 + \mathbf{b}_1 \\
H &= \max(0, Z_1) \\
Z_2 &= HW_2 + \mathbf{b}_2
\end{aligned}
$$

$$
\begin{aligned}
dW_2 &= H^T dZ_2 \\
d\mathbf{b}_2 &= \sum_{i=1}^{N} (dZ_2)_{i,:} \quad \text{(sum across batch rows)} \\
dH &= dZ_2 W_2^T
\end{aligned}
$$

$$
dZ_1 = dH \odot \mathbb{I}(Z_1 > 0)
$$

$$
\begin{aligned}
dW_1 &= X^T dZ_1 \\
d\mathbf{b}_1 &= \sum_{i=1}^{N} (dZ_1)_{i,:} \quad \text{(sum across batch rows)} \\
dX &= dZ_1 W_1^T \quad \text{(if needed for earlier layers)}
\end{aligned}
$$

Image Classification with CNNs

$$
\begin{aligned}
H_{out} &= \frac{H_{in} - F + 2P}{S} + 1 \\
W_{out} &= \frac{W_{in} - F + 2P}{S} + 1
\end{aligned}
$$

$$
\begin{aligned}
H_{out} &= \frac{3 - 2 + 0}{1} + 1 = 2 \\
W_{out} &= \frac{3 - 2 + 0}{1} + 1 = 2
\end{aligned}
$$

$$
\begin{aligned}
H_{out} &= \frac{32 - 5 + 2(2)}{1} + 1 = 32 \\
W_{out} &= \frac{32 - 5 + 2(2)}{1} + 1 = 32
\end{aligned}
$$

$$
\begin{aligned}
H_{out} &= \frac{H_{in} - F}{S} + 1 \\
W_{out} &= \frac{W_{in} - F}{S} + 1
\end{aligned}
$$

$$
\begin{aligned}
H_{out} &= \frac{4 - 2}{2} + 1 = 2 \\
W_{out} &= \frac{4 - 2}{2} + 1 = 2
\end{aligned}
$$

Video Understanding

$$T_{out} = \left\lfloor \frac{T + 2P_T - K_T}{S_T} \right\rfloor + 1$$

$$H_{out} = \left\lfloor \frac{H + 2P_H - K_H}{S_H} \right\rfloor + 1$$

$$W_{out} = \left\lfloor \frac{W + 2P_W - K_W}{S_W} \right\rfloor + 1$$

$$T_{out} = \left\lfloor \frac{16 + 2(1) - 3}{1} \right\rfloor + 1 = \lfloor 15 \rfloor + 1 = 16$$

$$H_{out} = \left\lfloor \frac{112 + 2(1) - 3}{2} \right\rfloor + 1 = \lfloor 55.5 \rfloor + 1 = 55 + 1 = 56$$

$$W_{out} = 56 \quad \text{(since width parameters match height)}$$

$$W_{3D}[:, :, t, :, :] = \frac{1}{K_T} W_{2D} \quad \text{for } t \in \{1, 2, \ldots, K_T\}$$

$$W_{2D} = \begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix}$$

$$\frac{1}{2} W_{2D} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$$

$$W_{3D}[:, :, 1, :, :] = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$$

$$W_{3D}[:, :, 2, :, :] = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$$

$$p_{final} = \alpha p_{spatial} + (1 - \alpha) p_{temporal}$$

$$p_{final} = 0.5 \times \begin{bmatrix} 0.10 \\ 0.70 \\ 0.20 \end{bmatrix} + 0.5 \times \begin{bmatrix} 0.30 \\ 0.40 \\ 0.30 \end{bmatrix}$$

$$p_{final} = \begin{bmatrix} 0.05 \\ 0.35 \\ 0.10 \end{bmatrix} + \begin{bmatrix} 0.15 \\ 0.20 \\ 0.15 \end{bmatrix} = \begin{bmatrix} 0.20 \\ 0.55 \\ 0.25 \end{bmatrix}$$

$$h_t = \sigma(W_x x_t + W_h h_{t-1} + b)$$

$$y = \text{softmax}(W_y h_T + b_y)$$

$$h_1 = \max\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \max\left(0, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$h_2 = \max\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \max\left(0, \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$$h_3 = \max\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix}\right) = \max\left(0, \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\theta = X W_\theta \quad \text{(Queries)}$$

$$\phi = X W_\phi \quad \text{(Keys)}$$

$$g = XW_g \quad \text{(Values)}$$

$$f = \theta\phi^T \quad \text{(Shape: } N \times N\text{)}$$

$$Y = Sg$$

$$Z = YW_z + X$$

$$f = XX^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$S = \frac{1}{e+1} \begin{bmatrix} e & 1 \\ 1 & e \end{bmatrix}$$

$$Y = Sg = \frac{1}{e+1} \begin{bmatrix} e & 1 \\ 1 & e \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{e+1} \begin{bmatrix} e & 1 \\ 1 & e \end{bmatrix}$$

Large Scale Distributed Training

$$2 + 2 + 4 + 4 + 4 = 16 \text{ bytes per parameter}$$

$$2 \times 10^{10} \text{ bytes} = 20 \text{ GB}$$

$$4 \times 10^{10} \text{ bytes} = 40 \text{ GB}$$

$$8 \times 10^{10} \text{ bytes} = 80 \text{ GB}$$

$$20 + 20 + 40 + 80 = 160 \text{ GB}$$

$$\text{Memory} = (10 \text{ boundaries} \times 100 \text{ MB}) + (10 \text{ internal layers} \times 100 \text{ MB}) = 1000 + 1000 = 2000 \text{ MB} = 2 \text{ GB}$$

$$\text{Time} = 100 \times 5 \text{ ms (Initial Forward)} + 100 \times 5 \text{ ms (Recompute)} + 100 \times 10 \text{ ms (Backward)}$$

$$\text{Time} = 500 + 500 + 1000 = 2000 \text{ ms}$$

$$V = 2 \times \frac{P-1}{P} \times S$$

$$T = \frac{V}{B}$$

$$S = 100 \times 10^6 \times 4 \text{ bytes} = 400 \text{ MB}$$

$$V = 2 \times \frac{8-1}{8} \times 400 \text{ MB} = 2 \times \frac{7}{8} \times 400 = 700 \text{ MB}$$

$$T = \frac{700 \text{ MB}}{50,000 \text{ MB/s}} = 0.014 \text{ seconds} = 14 \text{ ms}$$

$$\text{Memory per GPU} = 2\Phi \text{ (weights)} + 2\Phi \text{ (grads)} + \frac{12\Phi}{P}$$

$$\text{Memory per GPU} = 2\Phi \text{ (weights)} + \frac{14\Phi}{P}$$

$$\text{Memory per GPU} = \frac{16\Phi}{P}$$

$$\text{Memory} = 24 + 24 + 144 = 192 \text{ GB}$$

$$\text{Memory} = 24 \text{ (weights)} + 24 \text{ (grads)} + 36 \text{ (partitioned optim)} = 84 \text{ GB}$$

$$\text{Memory} = 24 \text{ (weights)} + 6 \text{ (partitioned grads)} + 36 \text{ (partitioned optim)} = 66 \text{ GB}$$

$$\text{Memory} = 6 \text{ (partitioned weights)} + 6 \text{ (partitioned grads)} + 36 \text{ (partitioned optim)} = 48 \text{ GB}$$

$$\text{Efficiency} = \frac{N}{N + P - 1}$$

$$t_f + t_b = 20 + 40 = 60 \text{ ms}$$

$$T_{\text{total}} = (16 + 4 - 1) \times 60 \text{ ms} = 19 \times 60 = 1140 \text{ ms}$$

$$T_{\text{bubble}} = (4 - 1) \times 60 \text{ ms} = 3 \times 60 = 180 \text{ ms}$$

$$\text{Efficiency} = \frac{16}{16 + 4 - 1} = \frac{16}{19} \approx 84.21\%$$

Vision and Language

$$\begin{aligned}
S_{1,1} &= 1\left(\frac{\sqrt{2}}{2}\right) + 0\left(\frac{\sqrt{2}}{2}\right) = 0.707 \\
S_{1,2} &= 1(0) + 0(1) = 0 \\
S_{2,1} &= 0\left(\frac{\sqrt{2}}{2}\right) + 1\left(\frac{\sqrt{2}}{2}\right) = 0.707 \\
S_{2,2} &= 0(0) + 1(1) = 1
\end{aligned}$$

$$\begin{aligned}
H_v &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} (1)(1) + 0 + (-1)(-1) & 0 & (1)(0) + 0 + (-1)(1) & (1)(-1) + 0 + (-1)(1) \\ (0)(1) + (2)(0) + (1)(-1) & 2(1) & (2)(-1) + (1)(1) & (1)(1) \\ (-1)(1) + 0 + 0 & 1(1) & 1(-1) & (-1)(-1) + 0 \\ 1 - 1 & 1 & -1 + 1 & -1 + 1 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 & -1 & -2 \\ -1 & 2 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}
\end{aligned}$$

$$M = \frac{\partial S_c(I)}{\partial I}$$

$$S(i,j) = \max_{k \in \{R,G,B\}} |M(i,j,k)|$$

$$I = \begin{bmatrix} 10 & 20 \\ 30 & 40 \end{bmatrix}$$

$$M = \frac{\partial S_c(I)}{\partial I} = \begin{bmatrix} 0.5 & -1.2 \\ 0.1 & 0.8 \end{bmatrix}$$

$$|M| = \begin{bmatrix} 0.5 & 1.2 \\ 0.1 & 0.8 \end{bmatrix}$$

$$S(i,j) = \frac{|M(i,j)| - 0.1}{1.2 - 0.1} = \frac{|M(i,j)| - 0.1}{1.1}$$

$$S = \begin{bmatrix} 0.36 & 1.0 \\ 0.0 & 0.64 \end{bmatrix}$$

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial S_c}{\partial A_{i,j}^k}$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k \right)$$

$$A^1 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

$$\frac{\partial S_c}{\partial A^1} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \quad \frac{\partial S_c}{\partial A^2} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\alpha_1^c = \frac{1}{4}(0.5 + 0.5 + 0.5 + 0.5) = 0.5$$

$$\alpha_2^c = \frac{1}{4}(-1 + 0 + 0 - 1) = -0.5$$

$$\sum_k \alpha_k^c A^k = 0.5 \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} - 0.5 \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 1.0 & 0.5 \end{bmatrix} - \begin{bmatrix} 0 & 1.5 \\ 0.5 & 0 \end{bmatrix} = \begin{bmatrix} 0.5 & -1.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\begin{bmatrix} 0.5 & -1.5 \\ 0.5 & 0.5 \end{bmatrix} \right) = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\text{DPD} = |P(\hat{Y} = 1 \mid A = 1) - P(\hat{Y} = 1 \mid A = 0)|$$

$$\text{EOD} = |P(\hat{Y} = 1 \mid Y = 1, A = 1) - P(\hat{Y} = 1 \mid Y = 1, A = 0)|$$

$$P(\hat{Y} = 1 \mid A = 1) = \frac{40}{100} = 0.40$$

$$P(\hat{Y} = 1 \mid A = 0) = \frac{90}{200} = 0.45$$

$$\text{DPD} = |0.40 - 0.45| = 0.05$$

$$P(\hat{Y} = 1 \mid Y = 1, A = 1) = \frac{30}{50} = 0.60$$

$$P(\hat{Y} = 1 \mid Y = 1, A = 0) = \frac{60}{80} = 0.75$$

$$\text{EOD} = |0.60 - 0.75| = 0.15$$

$$x^*_{LC} = \text{argmax}_x \left(1 - \max_i p(y_i \mid x)\right)$$

$$x^*_M = \text{argmin}_x \left(p(y_1 \mid x) - p(y_2 \mid x)\right)$$

$$x^*_E = \text{argmax}_x \left(-\sum_{i=1}^{C} p(y_i \mid x) \log_2 p(y_i \mid x)\right)$$

$$g_i = \nabla_\theta L(f_\theta(x_i), y_i)$$

$$\bar{g}_i = \frac{g_i}{\max\left(1, \frac{\|g_i\|_2}{C}\right)}$$

$$\tilde{g} = \frac{1}{B}\left(\sum_{i=1}^{B} \bar{g}_i + \mathcal{N}(0, \sigma^2 C^2 I)\right)$$

$$\theta_{t+1} = \theta_t - \eta \tilde{g}$$

$$g_1 = \begin{bmatrix} 3.0 \\ -4.0 \end{bmatrix}, \quad g_2 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$$

$$\bar{g}_1 = \frac{g_1}{\max\left(1, \frac{5.0}{2.0}\right)} = \frac{g_1}{2.5} = \begin{bmatrix} 1.2 \\ -1.6 \end{bmatrix}$$

$$\bar{g}_2 = \frac{g_2}{\max\left(1, \frac{1.414}{2.0}\right)} = \frac{g_2}{1} = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$$

$$\sum_{i=1}^{2} \bar{g}_i = \begin{bmatrix} 1.2 \\ -1.6 \end{bmatrix} + \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} = \begin{bmatrix} 2.2 \\ -0.6 \end{bmatrix}$$

$$\tilde{g} = \frac{1}{2}\left(\begin{bmatrix} 2.2 \\ -0.6 \end{bmatrix} + \begin{bmatrix} 0.5 \\ -1.0 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 2.7 \\ -1.6 \end{bmatrix} = \begin{bmatrix} 1.35 \\ -0.8 \end{bmatrix}$$

$$\theta_{t+1} = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} - 0.1 \begin{bmatrix} 1.35 \\ -0.8 \end{bmatrix} = \begin{bmatrix} -0.135 \\ 0.08 \end{bmatrix}$$

CHAPTER 7

CS234 Rl

Policy Search 1

$$
\begin{aligned}
\nabla_\theta \log \pi_\theta(a|s) &= \frac{(3.0 - 2.5)}{0.5}\begin{bmatrix}1\\2\end{bmatrix} \\
&= \frac{0.5}{0.5}\begin{bmatrix}1\\2\end{bmatrix} \\
&= \begin{bmatrix}1\\2\end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
h(s, a_1, \theta) = \theta^\top \phi(s, a_1) = \begin{bmatrix}\ln(2) & \ln(3)\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \ln(2) \\
h(s, a_2, \theta) = \theta^\top \phi(s, a_2) = \begin{bmatrix}\ln(2) & \ln(3)\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \ln(3)
\end{aligned}
$$

$$
\begin{aligned}
\pi_\theta(a_1|s) &= \frac{e^{\ln(2)}}{e^{\ln(2)} + e^{\ln(3)}} = \frac{2}{2+3} = 0.4 \\
\pi_\theta(a_2|s) &= \frac{e^{\ln(3)}}{e^{\ln(2)} + e^{\ln(3)}} = \frac{3}{2+3} = 0.6
\end{aligned}
$$

$$
\begin{aligned}
\sum_{a'} \pi_\theta(a'|s)\phi(s, a') &= \pi_\theta(a_1|s)\phi(s, a_1) + \pi_\theta(a_2|s)\phi(s, a_2) \\
&= 0.4\begin{bmatrix}1\\0\end{bmatrix} + 0.6\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0.4\\0.6\end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\nabla_\theta \log \pi_\theta(a_1|s) &= \phi(s, a_1) - \sum_{a'} \pi_\theta(a'|s)\phi(s, a') \\
&= \begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.4\\0.6\end{bmatrix} = \begin{bmatrix}0.6\\-0.6\end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
G_1 &= r_2 = 10 \\
G_0 &= r_1 + \gamma r_2 = 5 + 0.9(10) = 14
\end{aligned}
$$

$$
\begin{aligned}
\Delta\theta_0 &= \alpha G_0 \nabla_\theta \log \pi_\theta(a_0|s_0) \\
&= 0.1(14)\begin{bmatrix}0.2\\-0.1\end{bmatrix} \\
&= 1.4\begin{bmatrix}0.2\\-0.1\end{bmatrix} = \begin{bmatrix}0.28\\-0.14\end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\Delta\theta_1 &= \alpha G_1 \nabla_\theta \log \pi_\theta(a_1|s_1) \\
&= 0.1(10) \begin{bmatrix} -0.4 \\ 0.3 \end{bmatrix} \\
&= 1.0 \begin{bmatrix} -0.4 \\ 0.3 \end{bmatrix} = \begin{bmatrix} -0.4 \\ 0.3 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\theta_{\text{new}} &= \theta_{\text{old}} + \Delta\theta_0 + \Delta\theta_1 \\
&= \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} + \begin{bmatrix} 0.28 \\ -0.14 \end{bmatrix} + \begin{bmatrix} -0.4 \\ 0.3 \end{bmatrix} \\
&= \begin{bmatrix} 0.88 \\ 1.16 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\hat{g}] &= \pi_\theta(a_1|s)G(a_1)\nabla_\theta \log \pi_\theta(a_1|s) + \pi_\theta(a_2|s)G(a_2)\nabla_\theta \log \pi_\theta(a_2|s) \\
&= 0.5(110) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.5(90) \begin{bmatrix} -1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 55 \\ 0 \end{bmatrix} + \begin{bmatrix} -45 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
g^{(1)} &= 110 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 110 \\ 0 \end{bmatrix} \\
g^{(2)} &= 90 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -90 \\ 0 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{g}) &= 0.5 \left\| \begin{bmatrix} 110 \\ 0 \end{bmatrix} - \begin{bmatrix} 10 \\ 0 \end{bmatrix} \right\|^2 + 0.5 \left\| \begin{bmatrix} -90 \\ 0 \end{bmatrix} - \begin{bmatrix} 10 \\ 0 \end{bmatrix} \right\|^2 \\
&= 0.5 \left\| \begin{bmatrix} 100 \\ 0 \end{bmatrix} \right\|^2 + 0.5 \left\| \begin{bmatrix} -100 \\ 0 \end{bmatrix} \right\|^2 \\
&= 0.5(10000) + 0.5(10000) = 10000
\end{aligned}
$$

$$
\begin{aligned}
A(a_1) &= 110 - 100 = 10 \\
A(a_2) &= 90 - 100 = -10
\end{aligned}
$$

$$
\begin{aligned}
g_b^{(1)} &= 10 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix} \\
g_b^{(2)} &= -10 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{g}_b) &= 0.5 \left\| \begin{bmatrix} 10 \\ 0 \end{bmatrix} - \begin{bmatrix} 10 \\ 0 \end{bmatrix} \right\|^2 + 0.5 \left\| \begin{bmatrix} 10 \\ 0 \end{bmatrix} - \begin{bmatrix} 10 \\ 0 \end{bmatrix} \right\|^2 \\
&= 0 + 0 = 0
\end{aligned}
$$

MaxEnt IRL and Reinforcement Learning from Human Feedback

$$R(\tau_A) = \theta^T \Phi(\tau_A) = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3 - 1 = 2$$
$$R(\tau_B) = \theta^T \Phi(\tau_B) = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = 1 - 4 = -3$$

$$\begin{aligned} \text{KL\_est}(s, a) &= \log \pi_\phi(a|s) - \log \pi_{ref}(a|s) \\ &= \log(0.4) - \log(0.1) \\ &= \log\left(\frac{0.4}{0.1}\right) = \log(4) \approx 1.386 \end{aligned}$$

Course Review, Case Studies, and Open Challenges

$$\begin{aligned} T_{\text{PPO}} &= (10^7 \times 0.001) + (2 \times 10^6 \times 0.005) \\ &= 10,000 + 10,000 = 20,000 \text{ seconds.} \end{aligned}$$

$$\begin{aligned} T_{\text{SAC}} &= (2 \times 10^6 \times 0.001) + (2 \times 10^6 \times 0.005) \\ &= 2,000 + 10,000 = 12,000 \text{ seconds.} \end{aligned}$$

CHAPTER 8

CS236 Deep Generative Models

Autoregressive Models Part 2

$$M^W = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$M^V = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$M_A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M_B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$X = \begin{bmatrix} 2 & 4 & 1 \\ 3 & 5 & 8 \\ 1 & 2 & 9 \end{bmatrix}$$

$$W = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ -1 & -2 & 1 \end{bmatrix}$$

$$M_A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies W_A = W \odot M_A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\text{Output}_A = (1)(2) + (0)(4) + (-1)(1) + (2)(3) + (0)(5) + (0)(8) + (0)(1) + (0)(2) + (0)(9)$$

$$\text{Output}_A = 2 + 0 - 1 + 6 = 7$$

$$M_B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies W_B = W \odot M_B = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\text{Output}_B = (1)(2) + (0)(4) + (-1)(1) + (2)(3) + (1)(5) + (0)(8) + (0)(1) + (0)(2) + (0)(9)$$

$$\text{Output}_B = 2 + 0 - 1 + 6 + 5 = 12$$

$$R = 1 + L(K - 1)$$

$$R = 1 + \sum_{l=1}^{L} (K - 1) d_l$$

$$R = 1 + (2-1)(1) + (2-1)(2) + (2-1)(4) + (2-1)(8)$$

$$R = 1 + 1 + 2 + 4 + 8 = 16$$

$$P(x_1, \ldots, x_T) = \prod_{t=1}^{T} P(x_t \mid x_{1:t-1})$$

$$W = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad c = 0$$

$$h_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Wx_1 + Uh_0 + b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}(1) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$h_1 = \mathrm{ReLU}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\log P(x_1, x_2) = \log P(x_1) + \log P(x_2 \mid x_1) = -0.693 - 1.313 = -2.006$$

Variational Autoencoders Part 1

$$\begin{aligned} \mu_1 = 0.5, \quad \sigma_1^2 &= 0.8 \\ 1 + \log(\sigma_1^2) - \mu_1^2 - \sigma_1^2 &= 1 + \log(0.8) - (0.5)^2 - 0.8 \\ &= 1 - 0.2231 - 0.25 - 0.8 \\ &= -0.2731 \end{aligned}$$

$$\begin{aligned} \mu_2 = -1.0, \quad \sigma_2^2 &= 1.2 \\ 1 + \log(\sigma_2^2) - \mu_2^2 - \sigma_2^2 &= 1 + \log(1.2) - (-1.0)^2 - 1.2 \\ &= 1 + 0.1823 - 1.0 - 1.2 \\ &= -1.0177 \end{aligned}$$

$$\begin{aligned} D_{KL} &= -\frac{1}{2}(-0.2731 - 1.0177) \\ &= -\frac{1}{2}(-1.2908) = 0.6454 \end{aligned}$$

$$\begin{aligned} \mathcal{L}_{KL} &= -\frac{1}{2}(1 + \log\sigma^2 - \mu^2 - \exp(\log\sigma^2)) \\ &= -\frac{1}{2}(1 - 0.5 - (1.0)^2 - \exp(-0.5)) \\ &= -\frac{1}{2}(1 - 0.5 - 1.0 - 0.6065) \\ &= -\frac{1}{2}(-1.1065) = 0.55325 \end{aligned}$$

$$\begin{aligned} \sigma &= \exp(0.5\log\sigma^2) = \exp(-0.25) \approx 0.7788 \\ z &= \mu + \sigma\epsilon = 1.0 + (0.7788)(-0.8) = 1.0 - 0.623 = 0.377 \end{aligned}$$

$$
\begin{aligned}
\mathcal{L}_{recon} &= -\frac{1}{2}\log(2\pi(1)) - \frac{(3.0-2.5)^2}{2(1)} \\
&= -0.9189 - \frac{(0.5)^2}{2} \\
&= -0.9189 - \frac{0.25}{2} \\
&= -0.9189 - 0.125 = -1.0439
\end{aligned}
$$

$$
\begin{aligned}
\text{ELBO} &= \mathcal{L}_{recon} - \mathcal{L}_{KL} \\
&= -1.0439 - 0.55325 = -1.59715
\end{aligned}
$$

Variational Autoencoders Part 2

$$\mathcal{L}_{\text{recon}} = -\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]$$

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(q_\phi(z|x)\|p(z))$$

$$\mathcal{L}_{\text{KL}} = -0.5\sum_{j=1}^{D}\left(1+\log(\sigma_j^2)-\mu_j^2-\sigma_j^2\right)$$

$$\mathcal{J}_{\beta\text{-VAE}} = \mathcal{L}_{\text{recon}} + \beta\mathcal{L}_{\text{KL}}$$

$$\text{KL}_1 = -0.5(1+\ln(0.1)-0.5^2-0.1) = -0.5(1-2.302-0.25-0.1) = -0.5(-1.652) = 0.826$$

$$\text{KL}_2 = -0.5(1+\ln(0.5)-(-0.2)^2-0.5) = -0.5(1-0.693-0.04-0.5) = -0.5(-0.233) = 0.1165$$

$$\mathcal{J}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{KL}} = 12.5 + 0.9425 = 13.4425$$

$$\mathcal{J}_{\beta\text{-VAE}} = \mathcal{L}_{\text{recon}} + 4\cdot\mathcal{L}_{\text{KL}} = 12.5 + 4(0.9425) = 12.5 + 3.77 = 16.27$$

$$\nabla_\phi\mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)}[f(z)\nabla_\phi\log q_\phi(z)]$$

$$\nabla_\phi\mathcal{L} \approx \frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})\nabla_\phi\log q_\phi(z^{(l)})$$

$$\nabla_\phi\mathcal{L} \approx \frac{1}{L}\sum_{l=1}^{L}(f(z^{(l)})-b)\nabla_\phi\log q_\phi(z^{(l)})$$

$$\log q_\phi(z) = z\log p + (1-z)\log(1-p)$$

$$\nabla_\phi\log q_\phi(z) = \left(\frac{z}{p}-\frac{1-z}{1-p}\right)p(1-p) = z(1-p)-(1-z)p = z-p$$

$$\nabla_\phi\log q_\phi(z^{(1)}) = 1-0.5 = 0.5$$

$$f(z^{(1)}) = 3(1)+1 = 4$$

$$\nabla_\phi\mathbb{E}_{q_\phi(z)}[f(z)] \approx f(z^{(1)})\nabla_\phi\log q_\phi(z^{(1)}) = 4\times 0.5 = 2.0$$

$$g_i = -\log(-\log(u_i)) \quad \text{for } i = 1,\ldots,K$$

$$y_i = \frac{\exp((\log \pi_i + g_i)/\tau)}{\sum_{j=1}^{K} \exp((\log \pi_j + g_j)/\tau)}$$

$$\log \pi_1 = \log(0.2) \approx -1.609$$

$$\log \pi_2 = \log(0.5) \approx -0.693$$

$$\log \pi_3 = \log(0.3) \approx -1.204$$

$$g_1 = -\log(-\log(0.6)) = -\log(0.5108) \approx 0.672$$

$$g_2 = -\log(-\log(0.2)) = -\log(1.6094) \approx -0.476$$

$$g_3 = -\log(-\log(0.8)) = -\log(0.2231) \approx 1.500$$

$$v_1 = (-1.609 + 0.672)/0.5 = -0.937/0.5 = -1.874$$

$$v_2 = (-0.693 - 0.476)/0.5 = -1.169/0.5 = -2.338$$

$$v_3 = (-1.204 + 1.500)/0.5 = 0.296/0.5 = 0.592$$

$$\exp(v_1) = \exp(-1.874) \approx 0.153$$

$$\exp(v_2) = \exp(-2.338) \approx 0.096$$

$$\exp(v_3) = \exp(0.592) \approx 1.808$$

$$y_1 = 0.153/2.057 \approx 0.074$$

$$y_2 = 0.096/2.057 \approx 0.047$$

$$y_3 = 1.808/2.057 \approx 0.879$$

$$p(x, z_1, z_2) = p(x|z_1)p(z_1|z_2)p(z_2)$$

$$q(z_1, z_2|x) = q(z_1|x)q(z_2|z_1)$$

$$\text{ELBO} = \mathbb{E}_{q(z_1,z_2|x)}\left[\log \frac{p(x|z_1)p(z_1|z_2)p(z_2)}{q(z_1|x)q(z_2|z_1)}\right]$$

$$\text{ELBO} = \mathbb{E}_{q(z_1|x)}[\log p(x|z_1)] - \mathbb{E}_{q(z_1|x)}[\text{KL}(q(z_2|z_1)\|p(z_2))] + \mathbb{E}_{q(z_1,z_2|x)}[\log p(z_1|z_2) - \log q(z_1|x)]$$

$$\text{ELBO} \approx \log \frac{p(x|z_1)p(z_1|z_2)p(z_2)}{q(z_1|x)q(z_2|z_1)}$$

$$\text{ELBO} \approx \log p(x|z_1) + \log p(z_1|z_2) + \log p(z_2) - \log q(z_1|x) - \log q(z_2|z_1)$$

$$\text{ELBO} \approx -10.0 + (-2.5) + (-1.0) - (-2.0) - (-0.5)$$

$$\text{ELBO} \approx -13.5 + 2.5 = -11.0$$

$$\log q_K(z_K|x) = \log q_0(z_0|x) - \sum_{k=1}^{K} \log \left| \det \frac{\partial f_k}{\partial z_{k-1}} \right|$$

$$\text{ELBO} = \mathbb{E}_{q_0(z_0|x)} \left[\log p(x|z_K) + \log p(z_K) - \log q_0(z_0|x) + \sum_{k=1}^{K} \log \left| \det \frac{\partial f_k}{\partial z_{k-1}} \right| \right]$$

$$z_1 = 3(1) + 2 = 5$$

$$\log q_0(1) = -0.5 \log(2\pi) - 0.5(1)^2 \approx -0.9189 - 0.5 = -1.4189$$

$$\log \left| \det \frac{\partial f_1}{\partial z_0} \right| = \log(3) \approx 1.0986$$

$$\log q_1(z_1|x) = \log q_0(z_0|x) - \log \left| \det \frac{\partial f_1}{\partial z_0} \right| = -1.4189 - 1.0986 = -2.5175$$

$$\log p(z_1 = 5) = -0.5 \log(2\pi) - 0.5(5)^2 \approx -0.9189 - 12.5 = -13.4189$$

$$\text{ELBO} \approx \log p(x|z_1) + \log p(z_1) - \log q_1(z_1|x)$$

$$\text{ELBO} \approx -5.0 - 13.4189 - (-2.5175) = -18.4189 + 2.5175 = -15.9014$$

$$\mathcal{L}(\theta, \phi) = \mathcal{L}_{\text{recon}} + \lambda(t)\mathcal{L}_{\text{KL}}$$

$$\lambda(t) = \min\left(1, \frac{t}{T}\right)$$

$$\mathcal{L}_{\text{KL\_free\_bits}} = \sum_{j=1}^{D} \max\left(\lambda_{\text{target}}, \text{KL}(q(z_j|x) \| p(z_j))\right)$$

$$\mathcal{L}_{\text{KL\_standard}} = KL_1 + KL_2 + KL_3 = 0.5 + 2.1 + 0.1 = 2.7$$

$$\mathcal{L}_{\text{KL\_free\_bits}} = 1.0 + 2.1 + 1.0 = 4.1$$

Normalizing Flows Part 1

$$\begin{aligned} x &= e^{\frac{z}{2}} + 3 \\ x - 3 &= e^{\frac{z}{2}} \\ \ln(x-3) &= \frac{z}{2} \\ z &= f^{-1}(x) = 2\ln(x-3) \end{aligned}$$

$$\begin{aligned} \log p_Z(z) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}z^2 \\ \log p_Z(2\ln 2) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}(2\ln 2)^2 \\ &= -\frac{1}{2}\log(2\pi) - 2(\ln 2)^2 \end{aligned}$$

$$\frac{\partial f^{-1}(x)}{\partial x} = \frac{\partial}{\partial x}(2\ln(x-3)) = \frac{2}{x-3}$$

$$
\begin{aligned}
\log p_X(5) &= \log p_Z(f^{-1}(5)) + \log \left| \det \left(\frac{\partial f^{-1}(5)}{\partial x} \right) \right| \\
&= \left(-\frac{1}{2} \log(2\pi) - 2(\ln 2)^2 \right) + \log(1) \\
&= -\frac{1}{2} \log(2\pi) - 2(\ln 2)^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y}_1 &= \mathbf{x}_1 \\
\mathbf{y}_2 &= \mathbf{x}_2 \odot \exp(s(\mathbf{x}_1)) + t(\mathbf{x}_1)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x}_1 &= \mathbf{y}_1 \\
\mathbf{x}_2 &= (\mathbf{y}_2 - t(\mathbf{y}_1)) \odot \exp(-s(\mathbf{y}_1))
\end{aligned}
$$

$$
\begin{aligned}
s_1(x_1, x_2) &= \ln(2^2 + 1) = \ln(5) \\
s_2(x_1, x_2) &= \frac{1}{2}(4) = 2
\end{aligned}
$$

$$
\begin{aligned}
\log \left| \det \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right) \right| &= \sum_{i=1}^{2} s(\mathbf{x}_1)_i \\
&= \ln(5) + 2
\end{aligned}
$$

$$
\begin{aligned}
y_1 &= h_1 \cdot e^{\gamma y_2} + \delta y_2 \\
h_1 \cdot e^{\gamma y_2} &= y_1 - \delta y_2 \\
h_1 &= (y_1 - \delta y_2) \cdot e^{-\gamma y_2}
\end{aligned}
$$

$$
\begin{aligned}
h_2 &= x_2 \cdot e^{\alpha x_1} + \beta x_1^2 \\
x_2 \cdot e^{\alpha x_1} &= h_2 - \beta x_1^2 \\
x_2 &= (h_2 - \beta x_1^2) \cdot e^{-\alpha x_1}
\end{aligned}
$$

$$
\begin{aligned}
x_1 &= (y_1 - \delta y_2) e^{-\gamma y_2} \\
x_2 &= \left(y_2 - \beta \left((y_1 - \delta y_2) e^{-\gamma y_2} \right)^2 \right) e^{-\alpha (y_1 - \delta y_2) e^{-\gamma y_2}}
\end{aligned}
$$

Normalizing Flows Part 2

$$
\begin{aligned}
z_1 &= \frac{x_1 - 0}{\exp(0)} = x_1 = 1 \\
z_2 &= \frac{x_2 - \mu_2(x_1)}{\exp(\alpha_2(x_1))} = \frac{4 - 2(1)}{\exp(0.5 \cdot 1)} = \frac{2}{1.648} \approx 1.213
\end{aligned}
$$

$$
\begin{aligned}
x_1 &= z_1 \exp(0) + 0 = z_1 = 1 \\
x_2 &= z_2 \exp(\alpha_2(z_1)) + \mu_2(z_1) = 1.213 \exp(0.5 \cdot 1) + 2(1) \\
&= 1.213(1.648) + 2 \approx 2 + 2 = 4
\end{aligned}
$$

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 2e^{0.1(1)} + 1.5(1) = 2(1.105) + 1.5 = 3.71 \end{aligned}$$

$$\begin{aligned} z_{T1} &= 1 \\ z_{T2} &= (3.71 - 2(1))e^{-0.5(1)} = 1.71(0.6065) \approx 1.037 \end{aligned}$$

$$\begin{aligned} \frac{dz_2}{dt} &= -z_2 \implies z_2(t) = C_2 e^{-t} \\ \frac{dz_1}{dt} &= 2z_1 + z_2 \implies \frac{dz_1}{dt} - 2z_1 = C_2 e^{-t} \end{aligned}$$

$$\begin{aligned} \frac{d}{dt}(z_1 e^{-2t}) &= e^{2-t}e^{-2t} = e^{2-3t} \\ z_1(t)e^{-2t} &= \int e^{2-3t}dt = -\frac{1}{3}e^{2-3t} + C_1 \\ z_1(t) &= -\frac{1}{3}e^{2-t} + C_1 e^{2t} \end{aligned}$$

$$z_i = \frac{x_i - \mu_i(\mathbf{x}_{<i})}{\exp(\alpha_i(\mathbf{x}_{<i}))}$$

$$x_i = z_i \exp(\alpha_i(\mathbf{x}_{<i})) + \mu_i(\mathbf{x}_{<i})$$

$$x_i = z_i \exp(\alpha_i(\mathbf{z}_{<i})) + \mu_i(\mathbf{z}_{<i})$$

$$z_i = \frac{x_i - \mu_i(\mathbf{z}_{<i})}{\exp(\alpha_i(\mathbf{z}_{<i}))}$$

$$D_{KL}(p_S \parallel p_T) = \mathbb{E}_{\mathbf{x}\sim p_S}[\log p_S(\mathbf{x}) - \log p_T(\mathbf{x})]$$

$$L = \log p_S(\mathbf{x}) - \log p_T(\mathbf{x}) = -4.438 - (-3.376) = -1.062$$

$$\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t, \theta)$$

$$\frac{\partial \log p(\mathbf{z}(t))}{\partial t} = -\mathrm{Tr}\left(\frac{\partial f}{\partial \mathbf{z}(t)}\right)$$

$$\log p(\mathbf{x}) = \log p(\mathbf{z}(t_0)) - \int_{t_0}^{t_1} \mathrm{Tr}\left(\frac{\partial f}{\partial \mathbf{z}(t)}\right) dt$$

$$3 = -\frac{1}{3}e^{2-1} + C_1 e^2 \implies C_1 e^2 = 3 + \frac{e}{3} \implies C_1 = 3e^{-2} + \frac{e^{-1}}{3}$$

$$z_1(0) = -\frac{1}{3}e^2 + C_1 = -\frac{1}{3}e^2 + 3e^{-2} + \frac{1}{3e} \approx -2.463 + 0.406 + 0.123 = -1.934$$

$$\mathrm{Tr}(A) = 2 + (-1) = 1$$

$$\int_0^1 \mathrm{Tr}(A)dt = \int_0^1 1dt = 1$$

$$\log p_0(\mathbf{z}(0)) = -\log(2\pi) - \frac{1}{2}((-1.934)^2 + (7.389)^2) = -1.838 - \frac{1}{2}(3.74 + 54.60) = -31.008$$

$$\log p_1(\mathbf{x}) = \log p_0(\mathbf{z}(0)) - 1 = -32.008$$

$$\mathrm{Tr}(A) = \mathbb{E}_{p(\boldsymbol{\epsilon})}[\boldsymbol{\epsilon}^T A \boldsymbol{\epsilon}]$$

$$\mathrm{Tr}\left(\frac{\partial f}{\partial \mathbf{z}}\right) \approx \boldsymbol{\epsilon}^T \nabla_{\mathbf{z}}(\boldsymbol{\epsilon}^T f(\mathbf{z}, t))$$

$$\log p(\mathbf{x}) \approx \log p(\mathbf{z}(t_0)) - \int_{t_0}^{t_1} \boldsymbol{\epsilon}^T \nabla_{\mathbf{z}}(\boldsymbol{\epsilon}^T f(\mathbf{z}(t), t)) dt$$

$$J = \frac{\partial f}{\partial \mathbf{z}} = \begin{bmatrix} \cos(z_1) & 0 \\ z_2^2 & 2z_1 z_2 \end{bmatrix}$$

$$J = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$J\boldsymbol{\epsilon}^{(1)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$[1, 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (1)(1) + (1)(1) = 2$$

$$J\boldsymbol{\epsilon}^{(2)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$[1, -1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (1)(1) + (-1)(1) = 0$$

$$\text{Estimated Trace} = \frac{2+0}{2} = 1$$

$$z_i = \Phi^{-1}(F_i(\tilde{x}_i))$$

$$F_1(0.5) = 1 - e^{-0.5} \approx 1 - 0.6065 = 0.3935$$

$$z = \Phi^{-1}(0.3935) \approx -0.27$$

Generative Adversarial Networks (Part 1)

$$\begin{aligned} \mathbb{E}_{x \sim p_{\text{data}}}[\log D(x)] &= \frac{1}{3}\log D(1) + \frac{1}{3}\log D(2) + \frac{1}{3}\log D(3) \\ &= \frac{1}{3}\log(0.8) + \frac{1}{3}\log(0.5) + \frac{1}{3}\log(0.9) \\ &= \frac{1}{3}(-0.2231 - 0.6931 - 0.1054) \\ &= \frac{1}{3}(-1.0216) \approx -0.3405 \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] &= \frac{1}{2}\log(1 - D(2)) + \frac{1}{2}\log(1 - D(4)) \\ &= \frac{1}{2}\log(1 - 0.5) + \frac{1}{2}\log(1 - 0.1) \\ &= \frac{1}{2}\log(0.5) + \frac{1}{2}\log(0.9) \\ &= \frac{1}{2}(-0.6931 - 0.1054) \\ &= \frac{1}{2}(-0.7985) \approx -0.3993 \end{aligned}$$

$$
\begin{aligned}
V(D,G) &= \int_x p_{\text{data}}(x)\log D(x)dx + \int_z p_z(z)\log(1-D(G(z)))dz \\
&= \int_x \left[p_{\text{data}}(x)\log D(x) + p_g(x)\log(1-D(x))\right]dx
\end{aligned}
$$

$$
\begin{aligned}
KL(P \parallel M) &= 0.5\log_2\left(\frac{0.5}{0.25}\right) + 0.5\log_2\left(\frac{0.5}{0.5}\right) + 0\log_2\left(\frac{0}{0.25}\right) \\
&= 0.5\log_2(2) + 0.5\log_2(1) + 0 \\
&= 0.5(1) + 0.5(0) = 0.5 \text{ bits}
\end{aligned}
$$

$$
\begin{aligned}
KL(Q \parallel M) &= 0\log_2\left(\frac{0}{0.25}\right) + 0.5\log_2\left(\frac{0.5}{0.5}\right) + 0.5\log_2\left(\frac{0.5}{0.25}\right) \\
&= 0 + 0.5\log_2(1) + 0.5\log_2(2) \\
&= 0 + 0 + 0.5(1) = 0.5 \text{ bits}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}_G^{\text{minimax}}}{\partial \theta} &= \frac{\partial \log(1-y)}{\partial y} \cdot \frac{\partial D}{\partial x} \cdot \frac{\partial x}{\partial \theta} \\
&= \left(\frac{-1}{1-y}\right) \cdot [y(1-y)\cdot w] \cdot z \\
&= -y \cdot w \cdot z \\
&= -(4.54 \times 10^{-5}) \cdot (1) \cdot (1) = -4.54 \times 10^{-5}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}_G^{\text{NS}}}{\partial \theta} &= \frac{\partial(-\log y)}{\partial y} \cdot \frac{\partial D}{\partial x} \cdot \frac{\partial x}{\partial \theta} \\
&= \left(\frac{-1}{y}\right) \cdot [y(1-y)\cdot w] \cdot z \\
&= -(1-y) \cdot w \cdot z \\
&= -(1 - 4.54 \times 10^{-5}) \cdot (1) \cdot (1) \approx -1.0
\end{aligned}
$$

Generative Adversarial Networks (Part 2)

$$
\begin{aligned}
D_{\chi^2}(P \parallel Q) &= \sum_{x\in\{A,B,C\}} Q(x) f(t_x) \\
&= 0.4(0.25) + 0.2(2.25) + 0.4(0.0625) \\
&= 0.1 + 0.45 + 0.025 \\
&= 0.575
\end{aligned}
$$

$$
\begin{aligned}
W(P,Q) &= \int_0^1 \frac{x}{2}\,dx + \int_1^2 \frac{1}{2}\,dx + \int_2^3 \frac{3-x}{2}\,dx \\
&= \left[\frac{x^2}{4}\right]_0^1 + \left[\frac{x}{2}\right]_1^2 + \left[\frac{3x}{2} - \frac{x^2}{4}\right]_2^3 \\
&= \frac{1}{4} + \left(\frac{2}{2} - \frac{1}{2}\right) + \left(\left(\frac{9}{2} - \frac{9}{4}\right) - \left(\frac{6}{2} - \frac{4}{4}\right)\right) \\
&= 0.25 + 0.5 + (2.25 - 2.0) \\
&= 0.25 + 0.5 + 0.25 = 1
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_{GP} &= \lambda \left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2 \\
&= 10\left(\sqrt{5} - 1\right)^2 \\
&\approx 10(2.236 - 1)^2 \\
&= 10(1.236)^2 \\
&\approx 10(1.528) = 15.28
\end{aligned}
$$

$$D_f(P \parallel Q) = \sum_{x \in \mathcal{X}} Q(x) f\left(\frac{P(x)}{Q(x)}\right)$$

$$W(P, Q) = \int_{-\infty}^{\infty} |F_P(x) - F_Q(x)|\, dx$$

$$F_P(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

$$F_Q(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{2} & 1 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

$$W(P_{data}, P_G) = \sup_{\|D\|_L \leq 1} \left(\mathbb{E}_{x \sim P_{data}}[D(x)] - \mathbb{E}_{z \sim P_z}[D(G(z))]\right)$$

$$\mathcal{L}_D = -\frac{1}{m}\sum_{i=1}^{m} D(x^{(i)}) + \frac{1}{m}\sum_{i=1}^{m} D(G(z^{(i)}))$$

$$\mathcal{L}_G = -\frac{1}{m}\sum_{i=1}^{m} D(G(z^{(i)}))$$

$$\mathcal{L}_D = -\text{Mean}(D(\text{real})) + \text{Mean}(D(\text{fake})) = -0.5 + (-0.5) = -1.0$$

$$\mathcal{L}_G = -\text{Mean}(D(\text{fake})) = -(-0.5) = 0.5$$

$$w \leftarrow w - \alpha \nabla_w \mathcal{L}_D$$

$$w_i \leftarrow \max(-c, \min(c, w_i))$$

$$w_{\text{new}} = \begin{bmatrix} 0.08 \\ -0.05 \end{bmatrix} - 0.1 \begin{bmatrix} -0.5 \\ 0.4 \end{bmatrix}$$

$$w_{\text{new}} = \begin{bmatrix} 0.08 \\ -0.05 \end{bmatrix} + \begin{bmatrix} 0.05 \\ -0.04 \end{bmatrix} = \begin{bmatrix} 0.13 \\ -0.09 \end{bmatrix}$$

$$\hat{x} = \epsilon x + (1 - \epsilon)\tilde{x}$$

$$\nabla_{\hat{x}} D(\hat{x})$$

$$\mathcal{L}_{GP} = \lambda \left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2$$

$$\hat{x} = 0.25 \begin{bmatrix} 2.0 \\ 4.0 \end{bmatrix} + 0.75 \begin{bmatrix} -2.0 \\ 0.0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix} + \begin{bmatrix} -1.5 \\ 0.0 \end{bmatrix} = \begin{bmatrix} -1.0 \\ 1.0 \end{bmatrix}$$

$$\nabla_{\mathbf{v}} D(\mathbf{v}) = \begin{bmatrix} \frac{\partial D}{\partial v_1} \\ \frac{\partial D}{\partial v_2} \end{bmatrix} = \begin{bmatrix} 2v_1 \\ v_2 \end{bmatrix}$$

$$\nabla_{\hat{x}} D(\hat{x}) = \begin{bmatrix} 2(-1.0) \\ 1.0 \end{bmatrix} = \begin{bmatrix} -2.0 \\ 1.0 \end{bmatrix}$$

$$\|\nabla_{\hat{x}} D(\hat{x})\|_2 = \sqrt{(-2.0)^2 + (1.0)^2} = \sqrt{4.0 + 1.0} = \sqrt{5} \approx 2.236$$

$$\min_G \max_D V(D, G) = \mathbb{E}_{(x,y)\sim P_{data}}[\log D(x|y)] + \mathbb{E}_{z\sim P_z, y\sim P_{data}}[\log(1 - D(G(z|y)|y))]$$

$$\mathcal{L}_D = -\frac{1}{m} \sum_{i=1}^{m} \left[\log D(x^{(i)}|y^{(i)}) + \log(1 - D(G(z^{(i)}|y^{(i)})|y^{(i)}))\right]$$

$$\mathcal{L}_G = -\frac{1}{m} \sum_{i=1}^{m} \log D(G(z^{(i)}|y^{(i)})|y^{(i)})$$

$$\log D(x_1|1) = \log(0.8) \approx -0.223$$

$$\log(1 - D(G(z_1|1)|1)) = \log(1 - 0.4) = \log(0.6) \approx -0.511$$

$$\mathcal{L}_D = -(\log(0.8) + \log(0.6)) \approx -(-0.223 - 0.511) = 0.734$$

$$\mathcal{L}_G = -\log D(G(z_1|1)|1) = -\log(0.4) \approx 0.916$$

Energy-Based Models (Part 2)

$$\begin{aligned} x_1 &= x_0 - \frac{\epsilon}{2} \nabla_x E(x_0) + \sqrt{\epsilon} z_0 \\ &= 1.0 - \frac{0.1}{2}(0) + \sqrt{0.1}(0.5) \\ &\approx 1.0 - 0 + 0.316 \times 0.5 \\ &= 1.158 \end{aligned}$$

$$\begin{aligned} x_2 &= x_1 - \frac{\epsilon}{2} \nabla_x E(x_1) + \sqrt{\epsilon} z_1 \\ &= 1.158 - \frac{0.1}{2}(0.395) + \sqrt{0.1}(-0.2) \\ &\approx 1.158 - 0.05(0.395) - 0.316(0.2) \\ &\approx 1.158 - 0.020 - 0.063 \\ &= 1.075 \end{aligned}$$

$$\begin{aligned} \tilde{p}_\theta(x) &= \exp(-1 \times 0.5^2) = \exp(-0.25) \approx 0.7788 \\ \tilde{p}_\theta(\tilde{x}) &= \exp(-1 \times 0.8^2) = \exp(-0.64) \approx 0.5273 \end{aligned}$$

$$\begin{aligned} q(x) &= 0.5 \\ q(\tilde{x}) &= 0.5 \end{aligned}$$

$$\begin{aligned} \mathcal{L}_{NCE} &= \log P(y = 1|x; \theta) + \log P(y = 0|\tilde{x}; \theta) \\ &\approx \log(0.6090) + \log(0.4867) \\ &\approx -0.4959 - 0.7201 \\ &= -1.2160 \end{aligned}$$

Score-Based Diffusion Models (DDPM Perspective)

$$\begin{aligned} x_2 &= \sqrt{\bar{\alpha}_2}x_0 + \sqrt{1-\bar{\alpha}_2}\epsilon \\ &= 0.8485 \begin{bmatrix} 4.0 \\ -2.0 \end{bmatrix} + 0.5292 \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \\ &= \begin{bmatrix} 3.394 \\ -1.697 \end{bmatrix} + \begin{bmatrix} 0.2646 \\ -0.2646 \end{bmatrix} \\ &= \begin{bmatrix} 3.6586 \\ -1.9616 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} s_\theta(x_t, t) &= -\frac{1}{0.7} \begin{bmatrix} -1.4 \\ 0.7 \end{bmatrix} \\ &= - \begin{bmatrix} -2.0 \\ 1.0 \end{bmatrix} \\ &= \begin{bmatrix} 2.0 \\ -1.0 \end{bmatrix} \end{aligned}$$

Diffusion Models - Advanced Topics

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\epsilon}{\sqrt{\bar{\alpha}_t}}$$

$$d_t = \sqrt{1-\bar{\alpha}_{t-1} - \sigma_t^2}\epsilon$$

$$\epsilon_t \sim \mathcal{N}(0, I)$$

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_0 + d_t + \sigma_t \epsilon_t$$

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}$$

$$\sqrt{\bar{\alpha}_t} = \sqrt{0.64} = 0.8$$

$$\sqrt{1-\bar{\alpha}_t} = \sqrt{1-0.64} = \sqrt{0.36} = 0.6$$

$$\hat{x}_0 = \frac{0.5 - (0.6)(0.2)}{0.8} = \frac{0.5 - 0.12}{0.8} = \frac{0.38}{0.8} = 0.475$$

$$d_t = \sqrt{1-\bar{\alpha}_{t-1}}\epsilon_\theta(x_t, t)$$

$$\sqrt{1-\bar{\alpha}_{t-1}} = \sqrt{1-0.81} = \sqrt{0.19} \approx 0.4359$$

$$d_t = 0.4359 \times 0.2 = 0.08718$$

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_0 + d_t$$

$$\sqrt{\bar{\alpha}_{t-1}} = \sqrt{0.81} = 0.9$$

$$x_{t-1} = (0.9)(0.475) + 0.08718 = 0.4275 + 0.08718 = 0.51468$$

$$g_t = \nabla_{x_t} \log p_\phi(y|x_t, t)$$

$$\hat{\epsilon}_\theta(x_t, t) = \epsilon_\theta(x_t, t) - s\sqrt{1-\bar{\alpha}_t} g_t$$

$$\sqrt{1-\bar{\alpha}_t} = \sqrt{1-0.84} = \sqrt{0.16} = 0.4$$

$$\text{Adjustment} = -s\sqrt{1-\bar{\alpha}_t}\nabla_{x_t} \log p_\phi(y|x_t, t)$$

$$\text{Adjustment} = -2.0 \times 0.4 \times \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix} = -0.8 \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix} = \begin{bmatrix} -0.4 \\ -0.24 \end{bmatrix}$$

$$\hat{\epsilon}_\theta = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} + \begin{bmatrix} -0.4 \\ -0.24 \end{bmatrix} = \begin{bmatrix} -0.3 \\ -0.44 \end{bmatrix}$$

$$\hat{\epsilon}_\theta(x_t, t, c) = \epsilon_{\text{uncond}} + w(\epsilon_{\text{cond}} - \epsilon_{\text{uncond}})$$

$$\hat{\epsilon}_\theta(x_t, t, c) = (1-w)\epsilon_{\text{uncond}} + w\epsilon_{\text{cond}}$$

$$\Delta\epsilon = \epsilon_{\text{cond}} - \epsilon_{\text{uncond}} = \begin{bmatrix} 0.6 \\ -0.2 \end{bmatrix} - \begin{bmatrix} 0.4 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.2 \\ -0.3 \end{bmatrix}$$

$$w\Delta\epsilon = 7.5 \times \begin{bmatrix} 0.2 \\ -0.3 \end{bmatrix} = \begin{bmatrix} 1.5 \\ -2.25 \end{bmatrix}$$

$$\hat{\epsilon}_\theta = \begin{bmatrix} 0.4 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 1.5 \\ -2.25 \end{bmatrix} = \begin{bmatrix} 1.9 \\ -2.15 \end{bmatrix}$$

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

$$\sqrt{\bar{\alpha}_t} = \sqrt{0.49} = 0.7$$

$$\sqrt{1-\bar{\alpha}_t} = \sqrt{1-0.49} = \sqrt{0.51} \approx 0.714$$

$$z_t = 0.7 \begin{bmatrix} 1.0 \\ -1.0 \end{bmatrix} + 0.714 \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix}$$

$$z_t = \begin{bmatrix} 0.7 \\ -0.7 \end{bmatrix} + \begin{bmatrix} 0.357 \\ 0.1428 \end{bmatrix} = \begin{bmatrix} 1.057 \\ -0.5572 \end{bmatrix}$$

$$dx = f(x,t)dt + g(t)dw$$

$$dx = [f(x,t) - g(t)^2 \nabla_x \log p_t(x)]dt + g(t)d\bar{w}$$

$$s_\theta(x,t) \approx \nabla_x \log p_t(x)$$

$$dx = [0 - g(t)^2 s_\theta(x_t, t)]dt + g(t)d\bar{w} = -2.0(-0.5)dt + \sqrt{2.0}d\bar{w} = 1.0dt + 1.414d\bar{w}$$

$$\Delta x_{\text{det}} = 1.0 \times \Delta t = 1.0 \times (-0.1) = -0.1$$

$$\Delta x_{\text{stoch}} = g(t)z\sqrt{|\Delta t|} = 1.414 \times 0.3 \times \sqrt{0.1} \approx 1.414 \times 0.3 \times 0.316 \approx 0.134$$

$$x_{t+\Delta t} = x_t + \Delta x_{\text{det}} + \Delta x_{\text{stoch}}$$

$$x_{t+\Delta t} = 1.2 - 0.1 + 0.134 = 1.234$$

Part II

IITM BS Data Science

CHAPTER 9

Advanced Algorithmsprof Neeldhara

Dynamic Programming: Longest Increasing Subsequence

$$f(i) = 1 + \max_{0 \le j < i,\, A[j] < A[i]} f(j)$$

$$\text{LIS Length} = \max_{0 \le i < n} f(i)$$

$$L[i] = L[j] + 1$$

Approximation Algorithms

$$\begin{aligned} x_1 + x_2 &\ge 1 \quad (\text{edge } (1,2)) \\ x_2 + x_3 &\ge 1 \quad (\text{edge } (2,3)) \\ x_1 + x_3 &\ge 1 \quad (\text{edge } (1,3)) \\ x_1, x_2, x_3 \in \{0,1\} \end{aligned}$$

$$\begin{aligned} x_1 + x_2 &\ge 1 \\ x_2 + x_3 &\ge 1 \\ x_1 + x_3 &\ge 1 \\ 0 \le x_1 &\le 1 \\ 0 \le x_2 &\le 1 \\ 0 \le x_3 &\le 1 \end{aligned}$$

$$\begin{aligned} x_A + x_B &\ge 1 \\ x_B + x_C &\ge 1 \\ 0 \le x_A, x_B, x_C &\le 1 \end{aligned}$$

Exact Exponential Algorithms

$$\begin{aligned} H = {} & (-1)^0 W_2(\{1,2,3\}) + (-1)^1 \left[W_2(\{1,2\}) + W_2(\{2,3\}) + W_2(\{1,3\})\right] \\ & + (-1)^2 \left[W_2(\{1\}) + W_2(\{2\}) + W_2(\{3\})\right] + (-1)^3 W_2(\emptyset) \\ H = {} & 1(12) - 1(2+2+2) + 1(0) - 1(0) = 12 - 6 = 6. \end{aligned}$$

CHAPTER 10

Ai Search Methods For Problem Solving

State Space and Goal Stack Planning

$$
\begin{aligned}
G_1 &= (G_0 \setminus ADD(Stack(A,B))) \cup PRE(Stack(A,B)) \\
&= (\{On(A,B)\} \setminus \{On(A,B), \dots\}) \cup \{Holding(A), Clear(B)\} \\
&= \{Holding(A), Clear(B)\}
\end{aligned}
$$

$$
\begin{aligned}
G_2 &= (G_1 \setminus ADD(Pickup(A))) \cup PRE(Pickup(A)) \\
&= (\{Holding(A), Clear(B)\} \setminus \{Holding(A)\}) \cup \{OnTable(A), Clear(A), HandEmpty\} \\
&= \{Clear(B), OnTable(A), Clear(A), HandEmpty\}
\end{aligned}
$$

CHAPTER 11

Algorithmic Thinking In Bioinformatics

Local Alignment and Affine Gap Penalties

$$X_{i,j} = \max \begin{cases} X_{i-1,j} - \sigma & \text{(Extend deletion)} \\ M_{i-1,j} - \rho & \text{(Open deletion)} \end{cases}$$

$$Y_{i,j} = \max \begin{cases} Y_{i,j-1} - \sigma & \text{(Extend insertion)} \\ M_{i,j-1} - \rho & \text{(Open insertion)} \end{cases}$$

$$M_{i,j} = \max \begin{cases} X_{i,j} \\ Y_{i,j} \\ M_{i-1,j-1} + \text{score}(v_i, w_j) & \text{(Match or Mismatch)} \end{cases}$$

Heuristics for Non-Additive Phylogeny

$$\begin{aligned} \frac{\partial S}{\partial e_A} &= 2(e_A + e_B - 3) + 2(e_A + e_C - 4) = 0 \\ &\Rightarrow 2e_A + e_B + e_C = 7 \\ \frac{\partial S}{\partial e_B} &= 2(e_A + e_B - 3) + 2(e_B + e_C - 6) = 0 \\ &\Rightarrow e_A + 2e_B + e_C = 9 \\ \frac{\partial S}{\partial e_C} &= 2(e_A + e_C - 4) + 2(e_B + e_C - 6) = 0 \\ &\Rightarrow e_A + e_B + 2e_C = 10 \end{aligned}$$

$$\begin{aligned} w(E \to A) &= \frac{D_{AC}}{2} + \frac{R_A - R_C}{2(4-2)} \\ &= \frac{4}{2} + \frac{12 - 16}{4} \\ &= 2 - 1 = 1 \end{aligned}$$

$$\begin{aligned} w(E \to C) &= \frac{D_{AC}}{2} + \frac{R_C - R_A}{2(4-2)} \\ &= \frac{4}{2} + \frac{16 - 12}{4} \\ &= 2 + 1 = 3 \end{aligned}$$

$$\begin{aligned} D_{EB} &= \frac{1}{2}(D_{AB} + D_{CB} - D_{AC}) \\ &= \frac{1}{2}(3 + 5 - 4) = 2 \\ D_{ED} &= \frac{1}{2}(D_{AD} + D_{CD} - D_{AC}) \\ &= \frac{1}{2}(5 + 7 - 4) = 4 \end{aligned}$$

Character-Based Phylogeny (Parsimony)

$$\begin{aligned}
\text{Cost from left child (Leaf 1): } & \min(s_1(A) + c_{AA}, s_1(C) + c_{AC}) \\
& = \min(0 + 0, \infty + 2) = 0 \\
\text{Cost from right child (Leaf 2): } & \min(s_2(A) + c_{AA}, s_2(C) + c_{AC}) \\
& = \min(\infty + 0, 0 + 2) = 2 \\
s_u(A) & = 0 + 2 = 2
\end{aligned}$$

$$\begin{aligned}
\text{Cost from left child (Leaf 1): } & \min(s_1(A) + c_{CA}, s_1(C) + c_{CC}) \\
& = \min(0 + 2, \infty + 0) = 2 \\
\text{Cost from right child (Leaf 2): } & \min(s_2(A) + c_{CA}, s_2(C) + c_{CC}) \\
& = \min(\infty + 2, 0 + 0) = 0 \\
s_u(C) & = 2 + 0 = 2
\end{aligned}$$

Exact Pattern Matching with Tries and Trees

$$.Insert all these suffixes into a standard Trie starting from the longest suffix to the shortest.$$

Construct a Suffix Trie for a Text String Build a suffix trie for the text $T =$ "ACA".

Step 1: Append terminal character

$T\$ =$ "ACA$"

Step 2: Generate all suffixes

1. "ACA$"
2. "CA$"
3. "A$"
4. "$"

Step 3: Insert into standard Trie

- **Insert "ACA$"**: Root $\xrightarrow{\text{A}}$ Node 1 $\xrightarrow{\text{C}}$ Node 2 $\xrightarrow{\text{A}}$ Node 3 $\xrightarrow{\$}$ Node 4
- **Insert "CA$"**: Root $\xrightarrow{\text{C}}$ Node 5 $\xrightarrow{\text{A}}$ Node 6 $\xrightarrow{\$}$ Node 7
- **Insert "A$"**: Match 'A' from Root to Node 1. Create '$' edge from Node 1: Node 1 $\xrightarrow{\$}$ Node 8
- **Insert "$"**: Create '$' edge from Root: Root $\xrightarrow{\$}$ Node 9

Final Suffix Trie Edges:

Source Node	Edge Character	Target Node
Root	A	Node 1
Root	C	Node 5
Root	$	Node 9
Node 1	C	Node 2
Node 1	$	Node 8
Node 2	A	Node 3
Node 3	$	Node 4
Node 5	A	Node 6
Node 6	$	Node 7

11.1 Suffix Trees

Constructing a Suffix Tree from a Suffix Trie A Suffix Tree compresses the paths of a Suffix Trie to reduce space complexity.

1. Build a Suffix Trie for the text T

CHAPTER 12

Big Data And Biological Networks

Quantitative Genomics and GWAS Statistics

$$E = \frac{\text{Row Total} \times \text{Column Total}}{N}$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$\chi^2 = \frac{1200(450 \cdot 300 - 300 \cdot 150)^2}{750 \cdot 450 \cdot 600 \cdot 600}$$

$$\chi^2 = \frac{1200(135000 - 45000)^2}{121500000000}$$

$$\chi^2 = \frac{1200(90000)^2}{121500000000} = \frac{1200 \times 8100000000}{121500000000} = \frac{9720000000000}{121500000000} = 80$$

$$\alpha_{\text{GWAS}} = \frac{\alpha}{M}$$

$$\alpha_{\text{GWAS}} = \frac{0.05}{10^6} = 5 \times 10^{-8}$$

Network Centrality Measures

$$\begin{aligned} -2x_1 + x_2 + x_3 &= 0 \\ x_1 - 2x_2 + x_3 &= 0 \end{aligned}$$

CHAPTER 13

Business Analytics

Fitting Probability Distributions

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\hat{\sigma} = s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\lambda} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

$$\hat{\lambda} = \frac{5+8+4+6+5+7+9+5+6+4}{10}$$
$$\hat{\lambda} = \frac{59}{10} = 5.9$$

$$\begin{aligned} P(X=5) &= \frac{e^{-5.9}(5.9)^5}{5!} \\ &= \frac{0.00274 \times 7149.24}{120} \approx 0.163 \end{aligned}$$

$$E = n \times P(X=5)$$
$$E = 10 \times 0.163 = 1.63$$

$$\begin{aligned} \chi^2 &= \frac{(30-25)^2}{25} + \frac{(45-50)^2}{50} + \frac{(25-25)^2}{25} \\ &= \frac{25}{25} + \frac{25}{50} + \frac{0}{25} \\ &= 1.0 + 0.5 + 0 = 1.5 \end{aligned}$$

$$df = k - 1 - p$$
$$df = 3 - 1 - 1 = 1$$

Chi-Squared Test of Independence

$$P(A_i \text{ and } B_j) = \frac{O_{ij}}{N}$$

$$P(A_i) = \frac{R_i}{N}$$

$$P(B_j) = \frac{C_j}{N}$$

$$O_{11} = 20, \quad N = 100$$

$$P(\text{Action and Service X}) = \frac{20}{100} = 0.20$$

$$C_2 = 55, \quad N = 100$$

$$P(\text{Service Y}) = \frac{55}{100} = 0.55$$

$$E_{ij} = \frac{R_i \times C_j}{N}$$

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$df = (r-1) \times (c-1)$$

$$\chi^2 \approx 3.130 + 2.561 + 0.500 + 0.409 + 0.907 + 0.742 = 8.249$$

$$r = 3 \text{ (Action, Comedy, Drama)}$$

$$c = 2 \text{ (Service X, Service Y)}$$

$$df = (3-1) \times (2-1) = 2 \times 1 = 2$$

Linear Programming and Duality

$$\begin{aligned} \text{Maximize} \quad & Z = 40x_1 + 50x_2 \\ \text{Subject to} \quad & 2x_1 + 3x_2 \leq 12 \\ & x_1 + 2x_2 \leq 7 \\ & x_1, x_2 \geq 0 \end{aligned}$$

$$\begin{aligned} \text{Maximize} \quad & Z = 3x_1 + 2x_2 \\ \text{Subject to} \quad & x_1 + x_2 \leq 4 \\ & 2x_1 + x_2 \leq 5 \\ & x_1, x_2 \geq 0 \end{aligned}$$

$$\begin{aligned} \text{Maximize} \quad & Z = 5x_1 + 4x_2 \\ \text{Subject to} \quad & 6x_1 + 4x_2 \leq 24 \\ & x_1 + 2x_2 \leq 6 \\ & x_1, x_2 \geq 0 \end{aligned}$$

$$
\begin{aligned}
6x_1 + 4x_2 + s_1 &= 24 \\
x_1 + 2x_2 + s_2 &= 6 \\
Z - 5x_1 - 4x_2 &= 0
\end{aligned}
$$

$$
\begin{aligned}
\text{Maximize} \quad & Z = 3x_1 + 5x_2 \\
\text{Subject to} \quad & x_1 \leq 4 \\
& 2x_2 \leq 12 \\
& 3x_1 + 2x_2 \leq 18 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\text{Minimize} \quad & W = 4y_1 + 12y_2 + 18y_3 \\
\text{Subject to} \quad & 1y_1 + 0y_2 + 3y_3 \geq 3 \\
& 0y_1 + 2y_2 + 2y_3 \geq 5 \\
& y_1, y_2, y_3 \geq 0
\end{aligned}
$$

$$Z = 40x_1 + 50x_2$$

$$x_1 \geq 0, \quad x_2 \geq 0$$

$$\text{Minimize} \quad W = 4y_1 + 12y_2 + 18y_3$$

Logistic Regression Formulation and Prediction

$$
\begin{aligned}
z &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\
z &= -2.5 + (0.05)(60) + (-0.1)(12) \\
z &= -2.5 + 3.0 - 1.2 \\
z &= -0.7
\end{aligned}
$$

$$
\begin{aligned}
p &= \frac{1}{1 + e^{-(-0.7)}} \\
p &= \frac{1}{1 + e^{0.7}}
\end{aligned}
$$

Data Envelopment Analysis (DEA)

$$
\begin{aligned}
\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{ik} \quad & \text{for } i = 1, \ldots, m \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rk} \quad & \text{for } r = 1, \ldots, s \\
\lambda_j \geq 0 \quad & \text{for } j = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{ik} \quad & \text{for } i = 1, \ldots, m \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rk} \quad & \text{for } r = 1, \ldots, s \\
\sum_{j=1}^{n} \lambda_j = 1 \quad & \text{(Convexity Constraint)} \\
\lambda_j \geq 0 \quad & \text{for } j = 1, \ldots, n
\end{aligned}
$$

CHAPTER 14

Business Data Management

Pivot Tables and Multi-dimensional Data Summarization

$$\begin{aligned}\text{North Total} &= 100 + 150 + 50 = 300\\ \text{South Total} &= 200 + 300 = 500\end{aligned}$$

$$\begin{aligned}\text{North Count} &= 3 \text{ (Three records)}\\ \text{South Count} &= 2 \text{ (Two records)}\end{aligned}$$

$$\begin{aligned}\text{North Average} &= \frac{300}{3} = 100\\ \text{South Average} &= \frac{500}{2} = 250\end{aligned}$$

$$\begin{aligned}(\text{North}, \text{A}) &= 100 + 50 = 150\\ (\text{North}, \text{B}) &= 150\\ (\text{South}, \text{A}) &= 200\\ (\text{South}, \text{B}) &= 300\end{aligned}$$

$$\begin{aligned}\text{Electronics Profit} &= 5000 - 4000 = 1000\\ \text{Clothing Profit} &= 3000 - 1500 = 1500\end{aligned}$$

$$\begin{aligned}\text{Electronics Margin} &= \frac{1000}{5000} = 0.20 \text{ (20\%)}\\ \text{Clothing Margin} &= \frac{1500}{3000} = 0.50 \text{ (50\%)}\end{aligned}$$

CHAPTER 15

Computational Thinking

No prominent formulae extracted for this course.

CHAPTER 16

Computer Networks

Network Performance Metrics and Delay

$$
\begin{aligned}
d_{\text{trans1}} &= \frac{1000 \times 8}{10 \times 10^6} = \frac{8000}{10^7} = 0.8 \text{ ms} \\
d_{\text{prop1}} &= \frac{100 \times 10^3}{2 \times 10^8} = 0.5 \text{ ms} \\
d_{\text{proc1}} &= 0 \text{ ms} \\
d_{\text{link1}} &= 0.8 + 0.5 + 0 = 1.3 \text{ ms}
\end{aligned}
$$

$$
\begin{aligned}
d_{\text{trans2}} &= \frac{1000 \times 8}{100 \times 10^6} = \frac{8000}{10^8} = 0.08 \text{ ms} \\
d_{\text{prop2}} &= \frac{500 \times 10^3}{2 \times 10^8} = 2.5 \text{ ms} \\
d_{\text{proc2}} &= 1 \text{ ms} \\
d_{\text{link2}} &= 0.08 + 2.5 + 1 = 3.58 \text{ ms}
\end{aligned}
$$

TCP Timeout and Loss Recovery Mechanisms

$$
\begin{aligned}
\text{SRTT} &= R_1 \\
\text{RTTVAR} &= \frac{R_1}{2} \\
\text{RTO} &= \text{SRTT} + \max(G, 4 \times \text{RTTVAR})
\end{aligned}
$$

$$
\begin{aligned}
\text{RTTVAR} &= (1-\beta) \cdot \text{RTTVAR} + \beta \cdot |\text{SRTT} - R'| \\
\text{SRTT} &= (1-\alpha) \cdot \text{SRTT} + \alpha \cdot R' \\
\text{RTO} &= \text{SRTT} + \max(G, 4 \times \text{RTTVAR})
\end{aligned}
$$

$$
\begin{aligned}
\text{SRTT} &= 100 \text{ ms} \\
\text{RTTVAR} &= \frac{100}{2} = 50 \text{ ms} \\
\text{RTO} &= 100 + 4 \times 50 = 300 \text{ ms}
\end{aligned}
$$

$$
\begin{aligned}
\text{RTTVAR} &= (1-0.25) \times 50 + 0.25 \times |100 - 120| \\
&= 0.75 \times 50 + 0.25 \times 20 \\
&= 37.5 + 5 = 42.5 \text{ ms} \\
\text{SRTT} &= (1-0.125) \times 100 + 0.125 \times 120 \\
&= 0.875 \times 100 + 15 \\
&= 87.5 + 15 = 102.5 \text{ ms} \\
\text{RTO} &= 102.5 + 4 \times 42.5 \\
&= 102.5 + 170 = 272.5 \text{ ms}
\end{aligned}
$$

$$\text{ssthresh} = \frac{20}{2} = 10 \text{ MSS}$$
$$\text{cwnd} = 10 + 3 = 13 \text{ MSS}$$

$$\text{cwnd} = 13 + 1 = 14 \text{ MSS (after 4th DUP ACK)}$$
$$\text{cwnd} = 14 + 1 = 15 \text{ MSS (after 5th DUP ACK)}$$

$$\text{cwnd} = \text{ssthresh} = 10 \text{ MSS}$$

IPv4 Addressing and CIDR Subnetting

$$11111111.11111111.11111111.11100000$$

$$\text{Block Size} = 256 - \text{Subnet Mask Value}$$

$$11111111.11111111.11111111.11110000 \implies 255.255.255.240$$

$$2^b \geq 4 \implies b = 2$$

$$\text{New CIDR} = 16 + 2 = 18 \implies /18$$

$$11111111.11111111.11000000.00000000 \implies 255.255.192.0$$

$$\text{Block Size} = 256 - 192 = 64$$

Link State Routing (Dijkstra's Algorithm)

$$D(v) = \min(D(v), D(w) + c(w, v))$$

Distance Vector Routing

$$c(A, B) + D_B(D) = 2 + 8 = 10$$
$$c(A, C) + D_C(D) = 5 + 4 = 9$$

$$\begin{aligned} D_X(Y) &= \min\{c(X, Y) + D_Y(Y), c(X, Z) + D_Z(Y)\} \\ &= \min\{2 + 0, 7 + 1\} = 2 \quad \text{(via Y)} \\ D_X(Z) &= \min\{c(X, Y) + D_Y(Z), c(X, Z) + D_Z(Z)\} \\ &= \min\{2 + 1, 7 + 0\} = 3 \quad \text{(via Y)} \end{aligned}$$

$$\begin{aligned} D_Y(X) &= \min\{c(Y, X), c(Y, Z) + D_Z(X)\} \\ &= \min\{60, 1 + 5\} = 6 \end{aligned}$$

$$\begin{aligned} D_Z(X) &= \min\{c(Z, Y) + D_Y(X)\} \quad \text{(assuming no direct Z-X link)} \\ &= 1 + 6 = 7 \end{aligned}$$

$$D_Y(X) = \min\{60, 1 + 7\} = 8$$

$$D_Y(X) = \min\{c(Y, X), c(Y, Z) + D_Z(X)\}$$

$$D_Y(X) = \min\{60, 1 + \infty\} = 60$$

$$D_Z(X) = c(Z, Y) + D_Y(X) = 1 + 60 = 61$$

CHAPTER 17

Corporate Finance Sep 2024

Inflation and Yield Curves

$$1 + R = (1 + r)(1 + h)$$

$$r = \frac{1 + R}{1 + h} - 1$$

$$R = (1 + r)(1 + h) - 1$$

$$R \approx r + h$$

$$\begin{aligned} 1 + r &= \frac{1 + R}{1 + h} \\ 1 + r &= \frac{1.08}{1.03} \\ 1 + r &\approx 1.04854 \\ r &= 0.04854 \text{ or } 4.854\% \end{aligned}$$

$$\begin{aligned} r &\approx R - h \\ r &\approx 0.08 - 0.03 = 0.05 \text{ or } 5.00\% \end{aligned}$$

$$PV = \frac{C_1}{1 + r_1} + \frac{C_2}{(1 + r_2)^2} + \cdots + \frac{C_n}{(1 + r_n)^n}$$

$$PV = \frac{C_1}{1 + y} + \frac{C_2}{(1 + y)^2} + \cdots + \frac{C_n}{(1 + y)^n}$$

$$\begin{aligned} PV &= \frac{50}{1.04} + \frac{50}{(1.05)^2} + \frac{1050}{(1.06)^3} \\ PV &= 48.08 + \frac{50}{1.1025} + \frac{1050}{1.191016} \\ PV &= 48.08 + 45.35 + 881.60 \\ PV &= \$975.03 \end{aligned}$$

$$(1 + r_{n+1})^{n+1} = (1 + r_n)^n(1 + f_{n,n+1})$$

$$f_{n,n+1} = \frac{(1 + r_{n+1})^{n+1}}{(1 + r_n)^n} - 1$$

$$(1 + r_m)^m = (1 + r_n)^n(1 + f_{n,m})^{m-n}$$

$$f_{n,m} = \left(\frac{(1+r_m)^m}{(1+r_n)^n} \right)^{\frac{1}{m-n}} - 1$$

$$(1+r_3)^3 = (1+r_2)^2(1+f_{2,3})$$

$$\begin{aligned}
1 + f_{2,3} &= \frac{(1.045)^3}{(1.04)^2} \\
1 + f_{2,3} &= \frac{1.141166}{1.0816} \\
1 + f_{2,3} &\approx 1.05507 \\
f_{2,3} &= 0.05507 \text{ or } 5.51\%
\end{aligned}$$

$$(1+r_4)^4 = (1+r_2)^2(1+f_{2,4})^2$$

$$\begin{aligned}
(1 + f_{2,4})^2 &= \frac{(1.05)^4}{(1.04)^2} \\
(1 + f_{2,4})^2 &= \frac{1.215506}{1.0816} \approx 1.12380 \\
1 + f_{2,4} &= \sqrt{1.12380} \approx 1.05999 \\
f_{2,4} &\approx 0.0600 \text{ or } 6.00\%
\end{aligned}$$

Fundamentals of Portfolio Theory

$$\begin{aligned}
\text{Cov}(R_A, R_B) = {} & 0.3 \times (0.20 - 0.10) \times (0.10 - 0.035) \\
& + 0.5 \times (0.10 - 0.10) \times (0.05 - 0.035) \\
& + 0.2 \times (-0.05 - 0.10) \times (-0.10 - 0.035)
\end{aligned}$$

$$\begin{aligned}
\text{Boom:} \ & 0.3 \times (0.10) \times (0.065) = 0.00195 \\
\text{Normal:} \ & 0.5 \times (0) \times (0.015) = 0 \\
\text{Recession:} \ & 0.2 \times (-0.15) \times (-0.135) = 0.00405
\end{aligned}$$

$$\begin{aligned}
\sigma_p^2 &= (0.40)^2(0.15)^2 + (0.60)^2(0.20)^2 + 2(0.40)(0.60)(0.3)(0.15)(0.20) \\
&= (0.16)(0.0225) + (0.36)(0.04) + 2(0.24)(0.3)(0.03) \\
&= 0.0036 + 0.0144 + 0.00432 \\
&= 0.02232
\end{aligned}$$

Factor Models and Arbitrage Pricing Theory

$$\begin{aligned}
1) \quad w_A + w_B + w_C &= 0 \quad \text{(Zero investment)} \\
2) \quad 1.0w_A + 1.5w_B + 0.5w_C &= 0 \quad \text{(Zero beta)}
\end{aligned}$$

$$\begin{aligned}
1 + w_B + w_C &= 0 \implies w_C = -1 - w_B \\
1.0 + 1.5w_B + 0.5w_C &= 0
\end{aligned}$$

Exchange Rates and International Capital Markets

$$S_{A/C} = S_{A/B} \times S_{B/C}$$

$$S_{A/B} = \frac{S_{A/C}}{S_{B/C}}$$

$$S_{\text{INR/GBP}} = S_{\text{INR/USD}} \times S_{\text{USD/GBP}}$$

$$S_{\text{INR/GBP}} = 83.00 \times 1.25 = 103.75$$

$$S_{\text{CHF/CAD}} = \frac{S_{\text{CHF/USD}}}{S_{\text{CAD/USD}}}$$

$$S_{\text{CHF/CAD}} = \frac{0.90}{1.35} = 0.6667$$

$$S_{A/B} = \frac{P_A}{P_B}$$

$$S_{\text{INR/USD}} = \frac{P_{\text{INR}}}{P_{\text{USD}}}$$

$$S_{\text{INR/USD}} = \frac{100,000}{1,200} = 83.33$$

$$\frac{S_1}{S_0} = \frac{1+\pi_A}{1+\pi_B}$$

$$S_1 = S_0 \times \frac{1+\pi_A}{1+\pi_B}$$

$$\frac{S_1 - S_0}{S_0} \approx \pi_A - \pi_B$$

$$S_1 = S_0 \times \frac{1+\pi_{\text{USD}}}{1+\pi_{\text{EUR}}}$$

$$S_1 = 1.10 \times \frac{1+0.05}{1+0.02}$$

$$S_1 = 1.10 \times \frac{1.05}{1.02} = 1.10 \times 1.0294 = 1.1324$$

$$F_{A/B} = S_{A/B} \times \frac{1+r_A}{1+r_B}$$

$$F_{\text{JPY/USD}} = S_{\text{JPY/USD}} \times \frac{1+r_{\text{JPY}}}{1+r_{\text{USD}}}$$

$$F_{\text{JPY/USD}} = 145 \times \frac{1+0.005}{1+0.04}$$

$$F_{\text{JPY/USD}} = 145 \times \frac{1.005}{1.04} = 145 \times 0.966346 = 140.12$$

$$E[S_1] = S_0 \times \frac{1+r_A}{1+r_B}$$

$$E[S_1] = S_0 \times \frac{1+r_{\text{GBP}}}{1+r_{\text{USD}}}$$

$$E[S_1] = 0.85 \times \frac{1+0.06}{1+0.025}$$

$$E[S_1] = 0.85 \times \frac{1.06}{1.025} = 0.85 \times 1.034146 = 0.8790$$

$$E[S_1] = S_0 \times \frac{1 + r_{\text{MXN}}}{1 + r_{\text{USD}}}$$

$$18.00 = 17.00 \times \frac{1 + r_{\text{MXN}}}{1.04}$$

$$\frac{18.00}{17.00} = \frac{1 + r_{\text{MXN}}}{1.04}$$

$$1.0588 = \frac{1 + r_{\text{MXN}}}{1.04}$$

$$1 + r_{\text{MXN}} = 1.0588 \times 1.04 = 1.1012$$

$$r_{\text{MXN}} = 1.1012 - 1 = 0.1012$$

CHAPTER 18

Data Science Ai Lab

Numerical Computing and Plotting with NumPy and Matplotlib

$$
\begin{aligned}
C &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + 2 \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 5 & 8 \end{bmatrix}
\end{aligned}
$$

$$
D = \begin{bmatrix} 1(2) + 2(1) & 1(0) + 2(2) \\ 3(2) + 4(1) & 3(0) + 4(2) \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 10 & 8 \end{bmatrix}
$$

$$
E = \begin{bmatrix} 1(2) & 2(0) \\ 3(1) & 4(2) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 3 & 8 \end{bmatrix}
$$

Data Manipulation and Analysis with Pandas

$$
\begin{aligned}
(50000 - 62000)^2 &= (-12000)^2 = 144000000 \\
(60000 - 62000)^2 &= (-2000)^2 = 4000000 \\
(55000 - 62000)^2 &= (-7000)^2 = 49000000 \\
(80000 - 62000)^2 &= (18000)^2 = 324000000 \\
(65000 - 62000)^2 &= (3000)^2 = 9000000
\end{aligned}
$$

Image Processing and Convolutional Neural Networks

$$
\begin{aligned}
(0.0 - 0.5)/0.2 &= -2.5 \\
(0.498 - 0.5)/0.2 &= -0.01 \\
(0.8 - 0.5)/0.2 &= 1.5 \\
(1.0 - 0.5)/0.2 &= 2.5
\end{aligned}
$$

$$
\begin{aligned}
Y_{0,0} &= (10)(1) + (10)(0) + (0)(-1) \\
&\quad + (10)(1) + (10)(0) + (0)(-1) \\
&\quad + (10)(1) + (10)(0) + (0)(-1) \\
&= 10 + 10 + 10 = 30
\end{aligned}
$$

$$
\begin{aligned}
Y_{0,1} &= (10)(1) + (0)(0) + (0)(-1) \\
&\quad + (10)(1) + (0)(0) + (0)(-1) \\
&\quad + (10)(1) + (0)(0) + (0)(-1) \\
&= 10 + 10 + 10 = 30
\end{aligned}
$$

Advanced Computer Vision: Transfer Learning and Object Detection

$$v_k = \frac{1}{7 \times 7} \sum_{i=1}^{7} \sum_{j=1}^{7} A_{i,j,k}$$

$$\mathbf{z} = W^T \mathbf{v} + \mathbf{b}$$

$$P(y = i|\mathbf{v}) = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}}$$

$$e^{\mathbf{z}} = [e^{1.2}, e^{-0.5}, e^{2.8}] \approx [3.32, 0.61, 16.44]$$

$$\sum e^{z_j} = 3.32 + 0.61 + 16.44 = 20.37$$

$$\mathbf{p} = \left[\frac{3.32}{20.37}, \frac{0.61}{20.37}, \frac{16.44}{20.37}\right] \approx [0.163, 0.030, 0.807]$$

$$\theta_f^{(t+1)} = \theta_f^{(t)} - \eta \nabla_{\theta_f} L(\theta_f^{(t)})$$

$$\text{IoU}(B_p, B_g) = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$

$$xI_{min} = \max(x_{min}^p, x_{min}^g), \quad yI_{min} = \max(y_{min}^p, y_{min}^g)$$

$$xI_{max} = \min(x_{max}^p, x_{max}^g), \quad yI_{max} = \min(y_{max}^p, y_{max}^g)$$

$$I = \max(0, xI_{max} - xI_{min}) \times \max(0, yI_{max} - yI_{min})$$

$$U = \text{Area}(B_p) + \text{Area}(B_g) - I$$

$$\text{Area}(B_p) = (150 - 50) \times (150 - 50) = 100 \times 100 = 10000$$

$$\text{Area}(B_g) = (200 - 100) \times (200 - 100) = 100 \times 100 = 10000$$

$$xI_{min} = \max(50, 100) = 100$$

$$yI_{min} = \max(50, 100) = 100$$

$$xI_{max} = \min(150, 200) = 150$$

$$yI_{max} = \min(150, 200) = 150$$

$$I = \max(0, 150 - 100) \times \max(0, 150 - 100) = 50 \times 50 = 2500$$

$$U = 10000 + 10000 - 2500 = 17500$$

$$\text{IoU} = \frac{2500}{17500} = \frac{1}{7} \approx 0.1428$$

$$t_x = \frac{g_x - a_x}{a_w}, \quad t_y = \frac{g_y - a_y}{a_h}$$

$$t_w = \ln\left(\frac{g_w}{a_w}\right), \quad t_h = \ln\left(\frac{g_h}{a_h}\right)$$

$$p_x = (\hat{t}_x \times a_w) + a_x, \quad p_y = (\hat{t}_y \times a_h) + a_y$$

$$p_w = a_w \times e^{\hat{t}_w}, \quad p_h = a_h \times e^{\hat{t}_h}$$

$$t_x = \frac{60 - 50}{100} = \frac{10}{100} = 0.1$$

$$t_y = \frac{40 - 50}{50} = \frac{-10}{50} = -0.2$$

$$t_w = \ln\left(\frac{120}{100}\right) = \ln(1.2) \approx 0.182$$

$$t_h = \ln\left(\frac{60}{50}\right) = \ln(1.2) \approx 0.182$$

Data Visualization and Explainable AI

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

$$SPD = \mathbb{P}(Y = 1|U) - \mathbb{P}(Y = 1|P)$$

$$\mathbb{P}(Y = 1|P) = \frac{30}{60} = 0.50$$

$$\mathbb{P}(Y = 1|U) = \frac{10}{40} = 0.25$$

$$SPD = 0.25 - 0.50 = -0.25$$

$$DI = \frac{\mathbb{P}(Y = 1|U)}{\mathbb{P}(Y = 1|P)}$$

$$DI = \frac{0.45}{0.60}$$

$$DI = 0.75$$

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left(v(S \cup \{i\}) - v(S)\right)$$

$$f(x) = v(\emptyset) + \sum_{i=1}^{|N|} \phi_i$$

$$\phi_1 = \left(\frac{1}{2} \times 4\right) + \left(\frac{1}{2} \times 6\right) = 2 + 3 = 5$$

$$\phi_2 = \left(\frac{1}{2} \times 2\right) + \left(\frac{1}{2} \times 4\right) = 1 + 2 = 3$$

$$v(\emptyset) + \phi_1 + \phi_2 = 10 + 5 + 3 = 18 = v(\{1, 2\})$$

$$\mathcal{L} = \sum_{z \in Z} \pi_x(z) \left(f(z) - g(z)\right)^2$$

$$\mathcal{L}(w) = 0.5(f(-1) - w(-1))^2 + 1.0(f(0) - w(0))^2 + 0.5(f(1) - w(1))^2$$

$$\mathcal{L}(w) = 0.5(-2 + w)^2 + 1.0(0)^2 + 0.5(4 - w)^2$$

$$\mathcal{L}(w) = 0.5(w^2 - 4w + 4) + 0.5(w^2 - 8w + 16)$$

$$\mathcal{L}(w) = 0.5w^2 - 2w + 2 + 0.5w^2 - 4w + 8 = w^2 - 6w + 10$$

$$\frac{d\mathcal{L}}{dw} = 2w - 6 = 0 \implies w = 3$$

$$\hat{f}_{x_S}(x_S) = \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x_S, x_C^{(i)})$$

$$\hat{f}_{x_1}(5) = \frac{19 + 31}{2} = \frac{50}{2} = 25$$

$$TPR_P = \mathbb{P}(Y_{pred} = 1 | Y_{true} = 1, P)$$

$$TPR_U = \mathbb{P}(Y_{pred} = 1 | Y_{true} = 1, U)$$

$$EOD = TPR_U - TPR_P$$

$$TPR_P = \frac{80}{100} = 0.80$$

$$TPR_U = \frac{25}{50} = 0.50$$

$$EOD = 0.50 - 0.80 = -0.30$$

$$FPR_P = \mathbb{P}(Y_{pred} = 1 | Y_{true} = 0, P)$$

$$FPR_U = \mathbb{P}(Y_{pred} = 1 | Y_{true} = 0, U)$$

$$AOD = \frac{1}{2}\left[(FPR_U - FPR_P) + (TPR_U - TPR_P)\right]$$

$$FPR_U - FPR_P = 0.15 - 0.30 = -0.15$$

$$TPR_U - TPR_P = 0.65 - 0.85 = -0.20$$

$$AOD = \frac{1}{2}\left[-0.15 + (-0.20)\right] = \frac{-0.35}{2} = -0.175$$

Time Series Analysis and Forecasting

$$Y_t = T_t + S_t + R_t$$

$$Y_t = T_t \times S_t \times R_t$$

CHAPTER 19

Data Visualization Design

Representing Distributions and Outliers

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR}$$
$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR}$$

Advanced Geo-Visualization Techniques

$$d(P_1, H_1) = \sqrt{(0.5-0)^2 + (0.5-0)^2} = \sqrt{0.25+0.25} = \sqrt{0.5} \approx 0.707$$
$$d(P_1, H_2) = \sqrt{(0.5-2)^2 + (0.5-0)^2} = \sqrt{2.25+0.25} = \sqrt{2.5} \approx 1.581$$

$$d(P_2, H_1) = \sqrt{(1.5-0)^2 + (-0.5-0)^2} = \sqrt{2.25+0.25} = \sqrt{2.5} \approx 1.581$$
$$d(P_2, H_2) = \sqrt{(1.5-2)^2 + (-0.5-0)^2} = \sqrt{0.25+0.25} = \sqrt{0.5} \approx 0.707$$

$$d(P_3, H_1) = \sqrt{(-0.5-0)^2 + (0.1-0)^2} = \sqrt{0.25+0.01} = \sqrt{0.26} \approx 0.510$$
$$d(P_3, H_2) = \sqrt{(-0.5-2)^2 + (0.1-0)^2} = \sqrt{6.25+0.01} = \sqrt{6.26} \approx 2.502$$

$$d(P_4, H_1) = \sqrt{(1.2-0)^2 + (0.2-0)^2} = \sqrt{1.44+0.04} = \sqrt{1.48} \approx 1.216$$
$$d(P_4, H_2) = \sqrt{(1.2-2)^2 + (0.2-0)^2} = \sqrt{0.64+0.04} = \sqrt{0.68} \approx 0.824$$

$$d_1 = \sqrt{(5-5)^2 + (5-3)^2} = \sqrt{0+4} = 2$$
$$d_2 = \sqrt{(5-8)^2 + (5-5)^2} = \sqrt{9+0} = 3$$

$$\begin{aligned} K(d_1) &= \frac{3}{\pi(4)^2}\left(1 - \frac{2^2}{4^2}\right)^2 \\ &= \frac{3}{16\pi}\left(1 - \frac{4}{16}\right)^2 \\ &= \frac{3}{16\pi}\left(\frac{3}{4}\right)^2 = \frac{3}{16\pi} \times \frac{9}{16} = \frac{27}{256\pi} \approx 0.0336 \end{aligned}$$

$$\begin{aligned} K(d_2) &= \frac{3}{\pi(4)^2}\left(1 - \frac{3^2}{4^2}\right)^2 \\ &= \frac{3}{16\pi}\left(1 - \frac{9}{16}\right)^2 \\ &= \frac{3}{16\pi}\left(\frac{7}{16}\right)^2 = \frac{3}{16\pi} \times \frac{49}{256} = \frac{147}{4096\pi} \approx 0.0114 \end{aligned}$$

CHAPTER 20

Database Management Systems

Relational Model and Relational Algebra

$$R \div S = \pi_X(R) - \pi_X((\pi_X(R) \times S) - R)$$

Query Processing and Optimization

$$\pi_{\text{Name}}(\sigma_{\text{Salary}>50000 \wedge \text{DName}=\text{'IT'} \wedge E.\text{DID}=D.\text{DID}}(E \times D))$$

$$\pi_{\text{Name}}(\sigma_{\text{Salary}>50000}(\sigma_{\text{DName}=\text{'IT'}}(\sigma_{E.\text{DID}=D.\text{DID}}(E \times D))))$$

$$\pi_{\text{Name}}(\sigma_{\text{Salary}>50000}(\sigma_{\text{DName}=\text{'IT'}}(E \bowtie_{E.\text{DID}=D.\text{DID}} D)))$$

$$\pi_{\text{Name}}(\sigma_{\text{Salary}>50000}(E) \bowtie_{E.\text{DID}=D.\text{DID}} \sigma_{\text{DName}=\text{'IT'}}(D))$$

$$\pi_{\text{Name}}\left(\pi_{\text{Name,DID}}(\sigma_{\text{Salary}>50000}(E)) \bowtie_{E.\text{DID}=D.\text{DID}} \pi_{\text{DID}}(\sigma_{\text{DName}=\text{'IT'}}(D))\right)$$

$$\text{Cost} = B_R + (N_R \times B_S)$$

$$\text{Cost} = B_R + (B_R \times B_S)$$

$$\text{Cost} = B_R + \left(\left\lceil \frac{B_R}{M-2} \right\rceil \times B_S\right)$$

$$\text{Cost} = B_R + (N_R \times C)$$

$$\text{Cost} = \text{Sort}(R) + \text{Sort}(S) + (B_R + B_S)$$

$$\text{Cost} = 3(B_R + B_S)$$

$$\text{Cost} = B_R + \left(\left\lceil \frac{B_R}{100} \right\rceil \times B_S\right) = 1000 + \left(\left\lceil \frac{1000}{100} \right\rceil \times 500\right) = 1000 + (10 \times 500) = 6000 \text{ IOs}$$

$$\text{Cost} = B_S + \left(\left\lceil \frac{B_S}{100} \right\rceil \times B_R\right) = 500 + \left(\left\lceil \frac{500}{100} \right\rceil \times 1000\right) = 500 + (5 \times 1000) = 5500 \text{ IOs}$$

$$\text{Total Cost} = 3000 + 1500 + 1500 = 6000 \text{ IOs}$$

$$\text{Total Cost} = 3 \times (B_R + B_S) = 3 \times (1000 + 500) = 4500 \text{ IOs}$$

$$\text{Result Size} = \frac{N_R}{V(A, R)}$$

$$\text{Result Size} = \frac{N_R \times N_S}{\max(V(A, R), V(B, S))}$$

$$\text{Size}_{RS} = \frac{N_R \times N_S}{\max(V(b,R), V(b,S))} = \frac{1000 \times 2000}{\max(100, 200)} = \frac{2000000}{200} = 10000 \text{ tuples}$$

$$\text{Size}_{ST} = \frac{N_S \times N_T}{\max(V(c,S), V(c,T))} = \frac{2000 \times 3000}{\max(50, 300)} = \frac{6000000}{300} = 20000 \text{ tuples}$$

CHAPTER 21

Deep Learning

McCulloch-Pitts Neurons and Threshold Logic

$$\begin{aligned}
&\text{For } (0,0) \to 0: \quad 0 \cdot w_1 + 0 \cdot w_2 < \theta \implies 0 < \theta \\
&\text{For } (0,1) \to 0: \quad 0 \cdot w_1 + 1 \cdot w_2 < \theta \implies w_2 < \theta \\
&\text{For } (1,0) \to 0: \quad 1 \cdot w_1 + 0 \cdot w_2 < \theta \implies w_1 < \theta \\
&\text{For } (1,1) \to 1: \quad 1 \cdot w_1 + 1 \cdot w_2 \geq \theta \implies w_1 + w_2 \geq \theta
\end{aligned}$$

$$\begin{aligned}
&\text{For } (0,0) \to 1: \quad 0 \geq \theta \\
&\text{For } (0,1) \to 1: \quad w_2 \geq \theta \\
&\text{For } (1,0) \to 1: \quad w_1 \geq \theta \\
&\text{For } (1,1) \to 0: \quad w_1 + w_2 < \theta
\end{aligned}$$

$$\begin{aligned}
0 \cdot w_2 \geq \theta &\implies 0 \geq \theta \\
1 \cdot w_2 < \theta &\implies w_2 < \theta
\end{aligned}$$

Perceptron and Linear Separability

$$\begin{aligned}
w^T x_i + b > 0 \quad \text{for } y_i = 1 \\
w^T x_i + b < 0 \quad \text{for } y_i = -1
\end{aligned}$$

$$\begin{aligned}
(0)w_1 + (0)w_2 + b < 0 &\implies b < 0 \\
(0)w_1 + (1)w_2 + b < 0 &\implies w_2 + b < 0 \\
(1)w_1 + (0)w_2 + b < 0 &\implies w_1 + b < 0 \\
(1)w_1 + (1)w_2 + b > 0 &\implies w_1 + w_2 + b > 0
\end{aligned}$$

$$\begin{aligned}
b &< 0 \\
w_2 + b &> 0 \\
w_1 + b &> 0 \\
w_1 + w_2 + b &< 0
\end{aligned}$$

$$\begin{aligned}
w &\leftarrow w + \eta y_i x_i \\
b &\leftarrow b + \eta y_i
\end{aligned}$$

Feed Forward Networks and Activation Functions

$$
\begin{aligned}
\text{Input } (0,0) &\implies b < 0 \\
\text{Input } (0,1) &\implies w_2 + b < 0 \\
\text{Input } (1,0) &\implies w_1 + b < 0 \\
\text{Input } (1,1) &\implies w_1 + w_2 + b \geq 0
\end{aligned}
$$

Backpropagation: Forward and Backward Pass

$$
\begin{aligned}
z^{(1)} &= W^{(1)}x + b^{(1)} \\
&= \begin{bmatrix} 0.5 & -0.5 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} \\
&= \begin{bmatrix} (0.5)(1) + (-0.5)(-1) \\ (0.2)(1) + (0.8)(-1) \end{bmatrix} + \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} \\
&= \begin{bmatrix} 1.0 \\ -0.6 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} = \begin{bmatrix} 1.1 \\ -0.8 \end{bmatrix}
\end{aligned}
$$

$$
a^{(1)} = \mathrm{ReLU}(z^{(1)}) = \begin{bmatrix} \max(0, 1.1) \\ \max(0, -0.8) \end{bmatrix} = \begin{bmatrix} 1.1 \\ 0 \end{bmatrix}
$$

$$
\begin{aligned}
\hat{y} &= W^{(2)}a^{(1)} + b^{(2)} \\
&= \begin{bmatrix} 0.6 & -0.4 \end{bmatrix} \begin{bmatrix} 1.1 \\ 0 \end{bmatrix} + 0.3 \\
&= (0.6)(1.1) + (-0.4)(0) + 0.3 \\
&= 0.66 + 0 + 0.3 = 0.96
\end{aligned}
$$

$$
\begin{aligned}
L &= \frac{1}{2}(\hat{y} - y)^2 \\
&= \frac{1}{2}(0.96 - 0.5)^2 \\
&= \frac{1}{2}(0.46)^2 = 0.1058
\end{aligned}
$$

$$
\frac{\partial L}{\partial \hat{y}} = (\hat{y} - y) = 0.96 - 0.5 = 0.46
$$

$$
\begin{aligned}
\frac{\partial L}{\partial W^{(2)}} &= \delta^{(2)} a^{(1)T} = 0.46 \begin{bmatrix} 1.1 & 0 \end{bmatrix} = \begin{bmatrix} 0.506 & 0 \end{bmatrix} \\
\frac{\partial L}{\partial b^{(2)}} &= \delta^{(2)} = 0.46
\end{aligned}
$$

$$
\frac{\partial L}{\partial a^{(1)}} = W^{(2)T}\delta^{(2)} = \begin{bmatrix} 0.6 \\ -0.4 \end{bmatrix} (0.46) = \begin{bmatrix} 0.276 \\ -0.184 \end{bmatrix}
$$

$$
\begin{aligned}
\sigma'(z^{(1)}) &= \begin{bmatrix} 1 \text{ (since } 1.1 > 0) \\ 0 \text{ (since -0.8} < 0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
\delta^{(1)} &= \frac{\partial L}{\partial a^{(1)}} \odot \sigma'(z^{(1)}) = \begin{bmatrix} 0.276 \\ -0.184 \end{bmatrix} \odot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.276 \\ 0 \end{bmatrix}
\end{aligned}
$$

$$\frac{\partial L}{\partial W^{(1)}} = \delta^{(1)} x^T = \begin{bmatrix} 0.276 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 0.276 & -0.276 \\ 0 & 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial b^{(1)}} = \delta^{(1)} = \begin{bmatrix} 0.276 \\ 0 \end{bmatrix}$$

$$\begin{aligned} W^{(2)}_{\text{new}} &= W^{(2)} - \eta \frac{\partial L}{\partial W^{(2)}} \\ &= \begin{bmatrix} 0.6 & -0.4 \end{bmatrix} - 0.5 \begin{bmatrix} 0.506 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0.6 & -0.4 \end{bmatrix} - \begin{bmatrix} 0.253 & 0 \end{bmatrix} = \begin{bmatrix} 0.347 & -0.4 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} b^{(2)}_{\text{new}} &= b^{(2)} - \eta \frac{\partial L}{\partial b^{(2)}} \\ &= 0.3 - 0.5(0.46) \\ &= 0.3 - 0.23 = 0.07 \end{aligned}$$

$$\begin{aligned} W^{(1)}_{\text{new}} &= W^{(1)} - \eta \frac{\partial L}{\partial W^{(1)}} \\ &= \begin{bmatrix} 0.5 & -0.5 \\ 0.2 & 0.8 \end{bmatrix} - 0.5 \begin{bmatrix} 0.276 & -0.276 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0.5 & -0.5 \\ 0.2 & 0.8 \end{bmatrix} - \begin{bmatrix} 0.138 & -0.138 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0.362 & -0.362 \\ 0.2 & 0.8 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} b^{(1)}_{\text{new}} &= b^{(1)} - \eta \frac{\partial L}{\partial b^{(1)}} \\ &= \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} - 0.5 \begin{bmatrix} 0.276 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} - \begin{bmatrix} 0.138 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.038 \\ -0.2 \end{bmatrix} \end{aligned}$$

Computing Gradients using Matrix Calculus

$$\begin{aligned} e^{\mathbf{z}} &= \begin{bmatrix} e^1 \\ e^2 \end{bmatrix} \approx \begin{bmatrix} 2.718 \\ 7.389 \end{bmatrix} \\ \sum e^{z_k} &\approx 2.718 + 7.389 = 10.107 \\ y_1 &= \frac{2.718}{10.107} \approx 0.269 \\ y_2 &= \frac{7.389}{10.107} \approx 0.731 \end{aligned}$$

$$\begin{aligned} \frac{\partial y_1}{\partial z_1} &= y_1(1 - y_1) = 0.269(1 - 0.269) = 0.269(0.731) \approx 0.197 \\ \frac{\partial y_2}{\partial z_2} &= y_2(1 - y_2) = 0.731(1 - 0.731) = 0.731(0.269) \approx 0.197 \end{aligned}$$

$$\begin{aligned} \frac{\partial y_1}{\partial z_2} &= -y_1 y_2 = -(0.269)(0.731) \approx -0.197 \\ \frac{\partial y_2}{\partial z_1} &= -y_2 y_1 = -(0.731)(0.269) \approx -0.197 \end{aligned}$$

CHAPTER 22

Deep Learning For Computer Vision Prof Vineeth

Variational Autoencoders

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}))$$

$$D_{\mathrm{KL}} = -\frac{1}{2}\sum_{i=1}^{k}\left(1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2\right)$$

$$\text{Loss} = -\mathbb{E}[\log p_\theta(\mathbf{x}|\mathbf{z})] + \beta D_{\mathrm{KL}}$$

$$D_{\mathrm{KL}} = -\frac{1}{2}\left(-0.2677 - 0.0631\right) = -\frac{1}{2}(-0.3308) = 0.1654$$

$$\boldsymbol{\sigma} = \exp\left(\frac{1}{2}\log(\boldsymbol{\sigma}^2)\right)$$

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} 0.5 \\ 1.2 \\ -0.8 \end{bmatrix}$$

$$\mathcal{L}_{\text{recon}} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$$

$$\text{Total Loss} = \mathcal{L}_{\text{recon}} + \beta D_{\mathrm{KL}}$$

$$\begin{aligned} D_{\mathrm{KL}} &= -\frac{1}{2}(1 + \log(\sigma^2) - \mu^2 - \sigma^2) \\ &= -\frac{1}{2}(1 + \log(1.1) - 0.2^2 - 1.1) \\ &= -\frac{1}{2}(1 + 0.0953 - 0.04 - 1.1) \\ &= -\frac{1}{2}(-0.0447) = 0.02235 \end{aligned}$$

$$\begin{aligned} \mathcal{L}_{\text{recon}} &= (x - \hat{x})^2 \\ &= (0.8 - 0.6)^2 \\ &= 0.2^2 = 0.04 \end{aligned}$$

$$\text{Total Loss} = \mathcal{L}_{\text{recon}} + D_{\mathrm{KL}} = 0.04 + 0.02235 = 0.06235$$

$$\mathbf{z}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad \text{for } i = 1, 2, \ldots, N$$

Generative Adversarial Networks (GANs) Fundamentals

$$
\begin{aligned}
D_{KL}(P \parallel M) &= 0.5 \log_2\left(\frac{0.5}{0.35}\right) + 0.3 \log_2\left(\frac{0.3}{0.35}\right) + 0.2 \log_2\left(\frac{0.2}{0.3}\right) \\
&= 0.5 \log_2(1.428) + 0.3 \log_2(0.857) + 0.2 \log_2(0.667) \\
&= 0.5(0.514) + 0.3(-0.222) + 0.2(-0.585) \\
&= 0.257 - 0.0666 - 0.117 = 0.0734
\end{aligned}
$$

$$
\begin{aligned}
D_{KL}(Q \parallel M) &= 0.2 \log_2\left(\frac{0.2}{0.35}\right) + 0.4 \log_2\left(\frac{0.4}{0.35}\right) + 0.4 \log_2\left(\frac{0.4}{0.3}\right) \\
&= 0.2 \log_2(0.571) + 0.4 \log_2(1.143) + 0.4 \log_2(1.333) \\
&= 0.2(-0.808) + 0.4(0.193) + 0.4(0.415) \\
&= -0.1616 + 0.0772 + 0.166 = 0.0816
\end{aligned}
$$

Advanced GANs and Training Hacks

$$
\begin{aligned}
\sum_j \gamma_{i,j} &= P_r(i) \quad \forall i \\
\sum_i \gamma_{i,j} &= P_g(j) \quad \forall j \\
\gamma_{i,j} &\geq 0 \quad \forall i, j
\end{aligned}
$$

$$
\begin{aligned}
v &\leftarrow \frac{W^T u}{\|W^T u\|_2} \quad (v \text{ is size } M \times 1) \\
u &\leftarrow \frac{W v}{\|W v\|_2} \quad (u \text{ is size } N \times 1)
\end{aligned}
$$

Diffusion Models: Forward Process

$$
\begin{aligned}
\text{Mean scalar} &= \sqrt{1 - 0.1} = \sqrt{0.9} \approx 0.9487 \\
\text{Noise scalar} &= \sqrt{0.1} \approx 0.3162
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x}_t &= \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon} \\
\mathbf{x}_t &= (0.9487)(2.0) + (0.3162)(-0.5) \\
\mathbf{x}_t &= 1.8974 - 0.1581 \\
\mathbf{x}_t &= 1.7393
\end{aligned}
$$

$$
\begin{aligned}
\alpha_1 &= 1 - \beta_1 = 1 - 0.1 = 0.9 \\
\alpha_2 &= 1 - \beta_2 = 1 - 0.2 = 0.8
\end{aligned}
$$

$$
\begin{aligned}
\text{Mean scalar} &= \sqrt{\bar{\alpha}_2} = \sqrt{0.72} \approx 0.8485 \\
\text{Noise scalar} &= \sqrt{1 - \bar{\alpha}_2} = \sqrt{1 - 0.72} = \sqrt{0.28} \approx 0.5292
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x}_2 &= \sqrt{\bar{\alpha}_2}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_2}\boldsymbol{\epsilon} \\
\mathbf{x}_2 &= 0.8485 \begin{bmatrix} 4.0 \\ -2.0 \end{bmatrix} + 0.5292 \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix} \\
\mathbf{x}_2 &= \begin{bmatrix} 3.394 \\ -1.697 \end{bmatrix} + \begin{bmatrix} 0.5292 \\ 0.2646 \end{bmatrix} \\
\mathbf{x}_2 &= \begin{bmatrix} 3.9232 \\ -1.4324 \end{bmatrix}
\end{aligned}
$$

Classifier and Classifier-Free Diffusion Guidance

$$
\begin{aligned}
x_1 &= 0.0 - \frac{0.1}{2(1)}(0.0-2) + \sqrt{0.1}(0.5) \\
&= 0.0 + 0.05(2) + (\approx 0.316)(0.5) \\
&= 0.1 + 0.158 \\
&= 0.258
\end{aligned}
$$

$$
\begin{aligned}
\nabla_{x_t} \log p(y|x_t) &= \frac{d}{dx_t}\left[-\log(1+e^{-x_t})\right] \\
&= \frac{e^{-x_t}}{1+e^{-x_t}} \\
&= 1 - \sigma(x_t)
\end{aligned}
$$

$$
\epsilon_\theta(x_t, \emptyset) = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}
$$

$$
\epsilon_\theta(x_t, y) = \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix}
$$

$$
\begin{aligned}
\epsilon_\theta(x_t, y) - \epsilon_\theta(x_t, \emptyset) &= \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix} - \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} \\
&= \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}
\end{aligned}
$$

$$
w(\epsilon_\theta(x_t, y) - \epsilon_\theta(x_t, \emptyset)) = 3.0 \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 1.5 \end{bmatrix}
$$

$$
\begin{aligned}
\hat{\epsilon}_t &= \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} + \begin{bmatrix} 1.2 \\ 1.5 \end{bmatrix} \\
&= \begin{bmatrix} 1.3 \\ 1.3 \end{bmatrix}
\end{aligned}
$$

Text-Conditioned and Latent Diffusion Models

$$
h = \frac{512}{8} = 64
$$

$$
w = \frac{512}{8} = 64
$$

$$Q = XW_Q \quad \text{where} \quad W_Q \in \mathbb{R}^{c_{in} \times d}$$

$$K = CW_K \quad \text{where} \quad W_K \in \mathbb{R}^{d_c \times d}$$

$$V = CW_V \quad \text{where} \quad W_V \in \mathbb{R}^{d_c \times d}$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

$$\hat{\epsilon} = \epsilon_{uncond} + s \cdot (\epsilon_{cond} - \epsilon_{uncond})$$

$$\hat{\epsilon} = 0.2 + s \cdot (0.5 - 0.2) = 0.2 + s \cdot 0.3$$

$$\hat{\epsilon} = 0.2 + 1.0 \cdot 0.3 = 0.5$$

$$\hat{\epsilon} = 0.2 + 7.5 \cdot 0.3 = 0.2 + 2.25 = 2.45$$

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, I)$$

$$z_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(z_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \hat{\epsilon}\right) + \sigma_t z \quad \text{where} \quad z \sim \mathcal{N}(0, I)$$

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$$

$$z_t = 0.8 \begin{bmatrix} 1.2 \\ -0.5 \end{bmatrix} + 0.6 \begin{bmatrix} 0.1 \\ -0.8 \end{bmatrix}$$

$$z_t = \begin{bmatrix} 0.96 \\ -0.40 \end{bmatrix} + \begin{bmatrix} 0.06 \\ -0.48 \end{bmatrix}$$

$$z_t = \begin{bmatrix} 1.02 \\ -0.88 \end{bmatrix}$$

Vision Language Models and CLIP

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$\mathbf{w}_1 = \begin{bmatrix} 6 \\ 8 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\|\mathbf{v}_1\| = \sqrt{3^2 + 4^2} = 5 \implies \hat{\mathbf{v}}_1 = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$$
$$\|\mathbf{v}_2\| = \sqrt{1^2 + (-1)^2} = \sqrt{2} \implies \hat{\mathbf{v}}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \approx \begin{bmatrix} 0.707 \\ -0.707 \end{bmatrix}$$

$$\|\mathbf{w}_1\| = \sqrt{6^2 + 8^2} = 10 \implies \hat{\mathbf{w}}_1 = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$$
$$\|\mathbf{w}_2\| = \sqrt{(-1)^2 + 1^2} = \sqrt{2} \implies \hat{\mathbf{w}}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \approx \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$$

$$
\begin{aligned}
S_{1,1} &= \hat{\mathbf{v}}_1 \cdot \hat{\mathbf{w}}_1 = (0.6)(0.6) + (0.8)(0.8) = 1.0 \\
S_{1,2} &= \hat{\mathbf{v}}_1 \cdot \hat{\mathbf{w}}_2 = (0.6)(-0.707) + (0.8)(0.707) \approx -0.424 + 0.566 = 0.142 \\
S_{2,1} &= \hat{\mathbf{v}}_2 \cdot \hat{\mathbf{w}}_1 = (0.707)(0.6) + (-0.707)(0.8) \approx 0.424 - 0.566 = -0.142 \\
S_{2,2} &= \hat{\mathbf{v}}_2 \cdot \hat{\mathbf{w}}_2 = (0.707)(-0.707) + (-0.707)(0.707) \approx -0.5 - 0.5 = -1.0
\end{aligned}
$$

$$
\begin{aligned}
s_{\text{cat}} &= \hat{\mathbf{v}} \cdot \hat{\mathbf{w}}_{\text{cat}} = (0.5)(0.8) + (0.866)(0.6) = 0.4 + 0.5196 = 0.9196 \\
s_{\text{dog}} &= \hat{\mathbf{v}} \cdot \hat{\mathbf{w}}_{\text{dog}} = (0.5)(-0.6) + (0.866)(0.8) = -0.3 + 0.6928 = 0.3928
\end{aligned}
$$

$$
\begin{aligned}
p_{\text{cat}} &= \frac{\exp(0.9196)}{\exp(0.9196) + \exp(0.3928)} = \frac{2.508}{2.508 + 1.481} = \frac{2.508}{3.989} \approx 0.629 \\
p_{\text{dog}} &= \frac{\exp(0.3928)}{\exp(0.9196) + \exp(0.3928)} = \frac{1.481}{3.989} \approx 0.371
\end{aligned}
$$

CHAPTER 23

Deep Learning Practice Prof Mitesh Umesh Kaushik 2024 Sept

Text-To-Speech (TTS) Synthesis

$$e_{t,1} = \tanh(0.5 + 0.1 + 0.4) = \tanh(1.0) \approx 0.76$$
$$e_{t,2} = \tanh(0.5 + 0.4 + 0.5) = \tanh(1.4) \approx 0.88$$
$$e_{t,3} = \tanh(0.5 + 0.9 + 0.1) = \tanh(1.5) \approx 0.91$$

$$\ln(1 + 2) = \ln(3) \approx 1.10$$
$$\ln(1 + 0) = \ln(1) = 0.00$$
$$\ln(1 + 5) = \ln(6) \approx 1.79$$

$$e_1 = (1.10 - 1.0)^2 = (0.10)^2 = 0.01$$
$$e_2 = (0.00 - 0.2)^2 = (-0.20)^2 = 0.04$$
$$e_3 = (1.79 - 1.9)^2 = (-0.11)^2 = 0.0121$$

Object Detection Algorithms

$$x_c = \frac{x_{min} + x_{max}}{2}$$
$$y_c = \frac{y_{min} + y_{max}}{2}$$
$$w = x_{max} - x_{min}$$
$$h = y_{max} - y_{min}$$

$$x_{min} = x_c - \frac{w}{2}$$
$$y_{min} = y_c - \frac{h}{2}$$
$$x_{max} = x_c + \frac{w}{2}$$
$$y_{max} = y_c + \frac{h}{2}$$

$$x_c = \frac{10 + 50}{2} = \frac{60}{2} = 30$$
$$y_c = \frac{20 + 80}{2} = \frac{100}{2} = 50$$

$$w = 50 - 10 = 40$$
$$h = 80 - 20 = 60$$

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

$$b_x = 0.5 + 2 = 2.5$$
$$b_y = 0.5 + 3 = 3.5$$

$$b_w = 40 \times e^{0.5} \approx 40 \times 1.6487 = 65.95$$
$$b_h = 50 \times e^{-0.2} \approx 50 \times 0.8187 = 40.94$$

$$I_{x_{min}} = \max(A_{x_{min}}, B_{x_{min}})$$
$$I_{y_{min}} = \max(A_{y_{min}}, B_{y_{min}})$$
$$I_{x_{max}} = \min(A_{x_{max}}, B_{x_{max}})$$
$$I_{y_{max}} = \min(A_{y_{max}}, B_{y_{max}})$$

$$\text{Area}(A) = (A_{x_{max}} - A_{x_{min}}) \times (A_{y_{max}} - A_{y_{min}})$$
$$\text{Area}(B) = (B_{x_{max}} - B_{x_{min}}) \times (B_{y_{max}} - B_{y_{min}})$$

$$I_{x_{min}} = \max(10, 20) = 20$$
$$I_{y_{min}} = \max(10, 20) = 20$$
$$I_{x_{max}} = \min(40, 50) = 40$$
$$I_{y_{max}} = \min(40, 60) = 40$$

$$\text{Area}(A) = (40 - 10) \times (40 - 10) = 30 \times 30 = 900$$
$$\text{Area}(B) = (50 - 20) \times (60 - 20) = 30 \times 40 = 1200$$

CHAPTER 24

Deep Learning Workshop

Linear Algebra in PyTorch

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 22 & 28 \\ 49 & 64 \end{bmatrix}$$

$$\begin{bmatrix} 7 & 8 & 9 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 16 & 17 \\ 4 & 5 \end{bmatrix}$$

Loss Functions and Optimizers

$$\begin{aligned} 2.5 - 3.0 &= -0.5 \\ 0.0 - (-0.5) &= 0.5 \\ 2.1 - 2.0 &= 0.1 \\ 7.8 - 7.5 &= 0.3 \end{aligned}$$

$$\begin{aligned} (-0.5)^2 &= 0.25 \\ (0.5)^2 &= 0.25 \\ (0.1)^2 &= 0.01 \\ (0.3)^2 &= 0.09 \end{aligned}$$

$$\begin{aligned} e^{z_1} &= e^{2.0} \approx 7.389 \\ e^{z_2} &= e^{1.0} \approx 2.718 \\ e^{z_3} &= e^{0.1} \approx 1.105 \end{aligned}$$

Convolutional Neural Networks

$$\begin{aligned} H_{out} &= \left\lfloor \frac{224 + 2(3) - 7}{2} \right\rfloor + 1 \\ &= \left\lfloor \frac{224 + 6 - 7}{2} \right\rfloor + 1 \\ &= \left\lfloor \frac{223}{2} \right\rfloor + 1 \\ &= 111 + 1 = 112 \end{aligned}$$

$$
\begin{aligned}
H_{in} &= H_{in} + 2P - K + 1 \\
0 &= 2P - K + 1 \\
2P &= K - 1 \\
P &= \frac{K-1}{2}
\end{aligned}
$$

$$X = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 2 & 1 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$X_{0,0} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

$$(1)(1) + (2)(0) + (0)(0) + (1)(-1) = 1 + 0 + 0 - 1 = 0$$

$$X_{0,1} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$(2)(1) + (1)(0) + (1)(0) + (2)(-1) = 2 + 0 + 0 - 2 = 0$$

$$X_{1,0} = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$$

$$(0)(1) + (1)(0) + (2)(0) + (1)(-1) = 0 + 0 + 0 - 1 = -1$$

$$X_{1,1} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$$

$$(1)(1) + (2)(0) + (1)(0) + (0)(-1) = 1 + 0 + 0 + 0 = 1$$

$$Y = \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix}$$

$$H_{out} = \left\lfloor \frac{H_{in} + 2P - K}{S} \right\rfloor + 1$$

$$W_{out} = \left\lfloor \frac{W_{in} + 2P - K}{S} \right\rfloor + 1$$

$$H_{in} = \left\lfloor \frac{H_{in} + 2P - K}{1} \right\rfloor + 1$$

$$X = \begin{bmatrix} 1 & 3 & 2 & 4 \\ 5 & 6 & 1 & 2 \\ 0 & 2 & 8 & 3 \\ 1 & 1 & 4 & 5 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 1 & 3 \\ 5 & 6 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$$

$$W_3 = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$$

$$W_4 = \begin{bmatrix} 8 & 3 \\ 4 & 5 \end{bmatrix}$$

$$Y = \begin{bmatrix} 6 & 4 \\ 2 & 8 \end{bmatrix}$$

Sequential Modeling with RNNs

$$\mathbf{h}_t = \tanh(\mathbf{W}_{ih}\mathbf{x}_t + \mathbf{b}_{ih} + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_{hh})$$

$$\mathbf{f}_1 \odot \mathbf{c}_0 = \begin{bmatrix} 0.8 \\ 0.1 \end{bmatrix} \odot \begin{bmatrix} 0.5 \\ -0.2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ -0.02 \end{bmatrix}$$

$$\mathbf{i}_1 \odot \mathbf{g}_1 = \begin{bmatrix} 0.5 \\ 0.9 \end{bmatrix} \odot \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} = \begin{bmatrix} 0.20 \\ 0.54 \end{bmatrix}$$

$$\mathbf{c}_1 = \begin{bmatrix} 0.4 \\ -0.02 \end{bmatrix} + \begin{bmatrix} 0.20 \\ 0.54 \end{bmatrix} = \begin{bmatrix} 0.60 \\ 0.52 \end{bmatrix}$$

$$\tanh(\mathbf{c}_1) = \begin{bmatrix} 0.537 \\ 0.477 \end{bmatrix}$$

$$\mathbf{h}_1 = \mathbf{o}_1 \odot \tanh(\mathbf{c}_1) = \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix} \odot \begin{bmatrix} 0.537 \\ 0.477 \end{bmatrix} = \begin{bmatrix} 0.3759 \\ 0.1431 \end{bmatrix}$$

$$\mathbf{X}_{padded} = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.0 \\ 0.6 & 0.0 & 0.0 \end{bmatrix} \quad \text{Shape: } (3, 3, 1)$$

CHAPTER 25

Design Thinking For Data Driven App Development

No prominent formulae extracted for this course.

CHAPTER 26

English 1 Basic English

No prominent formulae extracted for this course.

CHAPTER 27

English Ii

Quantifiers, Contrast Words & Vocabulary Building

$$\text{Word} = [\text{Prefix}] + \text{Base Root} + [\text{Suffix}]$$

$$\text{structure} + \text{-ed} \rightarrow \text{structured (Adjective: having organization)}$$

$$\text{un-} + \text{structured} \rightarrow \text{unstructured}$$

CHAPTER 28

Financial Forensics

The Balance Sheet

$$\begin{aligned}\text{Accumulated Depreciation (Year 2)} &= 9,000 \times 2 = \$18,000\\ \text{Book Value} &= \text{Cost} - \text{Accumulated Depreciation}\\ \text{Book Value} &= 50,000 - 18,000 = \$32,000\end{aligned}$$

$$\begin{aligned}\text{Beginning Book Value} &= \$50,000\\ \text{Depreciation Expense} &= 50,000 \times 0.40 = \$20,000\\ \text{Ending Book Value} &= 50,000 - 20,000 = \$30,000\end{aligned}$$

$$\begin{aligned}\text{Beginning Book Value} &= \$30,000\\ \text{Depreciation Expense} &= 30,000 \times 0.40 = \$12,000\\ \text{Ending Book Value} &= 30,000 - 12,000 = \$18,000\end{aligned}$$

The Cash Flow Statement

$$\begin{aligned}\text{FCFF} &= \text{CFO} + \text{Interest} \times (1 - \text{Tax Rate}) - \text{CapEx}\\ &= 250000 + 40000 \times (1 - 0.25) - 60000\\ &= 250000 + 40000 \times 0.75 - 60000\\ &= 250000 + 30000 - 60000 = 220000\end{aligned}$$

$$\begin{aligned}\text{FCFE} &= \text{CFO} - \text{CapEx} + \text{Net Borrowing}\\ &= 250000 - 60000 + 15000\\ &= 190000 + 15000 = 205000\end{aligned}$$

Corporate Finance Fundamentals

$$\begin{aligned}PV &= 1000 \times \left[\frac{1 - (1 + 0.06)^{-5}}{0.06}\right]\\ &= 1000 \times \left[\frac{1 - (1.06)^{-5}}{0.06}\right]\\ &= 1000 \times \left[\frac{1 - 0.747258}{0.06}\right]\\ &= 1000 \times \left[\frac{0.252742}{0.06}\right]\\ &= 1000 \times 4.21236\\ &= \$4212.36\end{aligned}$$

$$
\begin{aligned}
FV &= 1000 \times \left[\frac{(1+0.06)^5 - 1}{0.06}\right] \\
&= 1000 \times \left[\frac{1.338225 - 1}{0.06}\right] \\
&= 1000 \times \left[\frac{0.338225}{0.06}\right] \\
&= 1000 \times 5.63709 \\
&= \$5637.09
\end{aligned}
$$

$$
\begin{aligned}
PV_{\text{due}} &= 500 \times \left[\frac{1 - (1+0.01)^{-12}}{0.01}\right] \times (1+0.01) \\
&= 500 \times \left[\frac{1 - 0.887449}{0.01}\right] \times 1.01 \\
&= 500 \times 11.2551 \times 1.01 \\
&= 5627.55 \times 1.01 \\
&= \$5683.83
\end{aligned}
$$

$$
\begin{aligned}
PV &= \frac{C_1}{r - g} \\
&= \frac{2.50}{0.08 - 0.03} \\
&= \frac{2.50}{0.05} \\
&= \$50.00
\end{aligned}
$$

Advanced Credit Risk Modeling

$$BR_i = \frac{\text{Number of Bad Accepts in Parcel } i}{\text{Total Accepts in Parcel } i}$$

$$\text{Assigned Bad Rate for Rejects in Parcel } i = \min(k \times BR_i, 1)$$

$$PD_{raw}(S) = \frac{1}{1 + e^{-L(S)}}$$

$$\text{Shift} = \ln\left(\frac{CT}{1 - CT}\right) - \ln\left(\frac{EDR}{1 - EDR}\right)$$

$$L_{calibrated}(S) = L(S) + \text{Shift}$$

$$PD_{calibrated}(S) = \frac{1}{1 + e^{-L_{calibrated}(S)}}$$

$$L(S) = \ln\left(\frac{0.05}{1 - 0.05}\right) = \ln\left(\frac{0.05}{0.95}\right) = \ln(0.0526) \approx -2.944$$

$$\text{Log-odds of EDR} = \ln\left(\frac{0.04}{0.96}\right) = \ln(0.0417) \approx -3.178$$

$$\text{Log-odds of CT} = \ln\left(\frac{0.06}{0.94}\right) = \ln(0.0638) \approx -2.752$$

$$\text{Shift} = \text{Log-odds}(CT) - \text{Log-odds}(EDR)$$

$$\text{Shift} = -2.752 - (-3.178) = 0.426$$

$$L_{calibrated}(S) = -2.944 + 0.426 = -2.518$$

$$PD_{calibrated}(S) = \frac{1}{1 + e^{-(-2.518)}} = \frac{1}{1 + 12.404} = \frac{1}{13.404} \approx 0.0746$$

Unsupervised Learning and Anomaly Detection

$$\begin{aligned}
\Sigma &= \frac{1}{4}\left(\begin{bmatrix} 9 & 6 \\ 6 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} + \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 64 & 64 \\ 64 & 64 \end{bmatrix}\right) \\
&= \frac{1}{4}\begin{bmatrix} 82 & 80 \\ 80 & 82 \end{bmatrix} = \begin{bmatrix} 20.5 & 20 \\ 20 & 20.5 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
D^2(x_1) &= \begin{bmatrix} -3 & -2 \end{bmatrix} \frac{1}{81} \begin{bmatrix} 82 & -80 \\ -80 & 82 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \end{bmatrix} \\
&= \frac{1}{81} \begin{bmatrix} -3 & -2 \end{bmatrix} \begin{bmatrix} -246 + 160 \\ 240 - 164 \end{bmatrix} \\
&= \frac{1}{81} \begin{bmatrix} -3 & -2 \end{bmatrix} \begin{bmatrix} -86 \\ 76 \end{bmatrix} \\
&= \frac{258 - 152}{81} = \frac{106}{81} \approx 1.31
\end{aligned}$$

$$\begin{aligned}
D^2(x_5) &= \begin{bmatrix} 8 & 8 \end{bmatrix} \frac{1}{81} \begin{bmatrix} 82 & -80 \\ -80 & 82 \end{bmatrix} \begin{bmatrix} 8 \\ 8 \end{bmatrix} \\
&= \frac{1}{81} \begin{bmatrix} 8 & 8 \end{bmatrix} \begin{bmatrix} 656 - 640 \\ -640 + 656 \end{bmatrix} \\
&= \frac{1}{81} \begin{bmatrix} 8 & 8 \end{bmatrix} \begin{bmatrix} 16 \\ 16 \end{bmatrix} \\
&= \frac{128 + 128}{81} = \frac{256}{81} \approx 3.16
\end{aligned}$$

$$\begin{aligned}
c(4) &= 2H(3) - \frac{2(3)}{4} \\
&= 2(1.833) - 1.5 \\
&= 3.666 - 1.5 = 2.166
\end{aligned}$$

CHAPTER 29

Game Theory And Strategy

Bayesian Games (Incomplete Information)

$$
\begin{aligned}
EU_2(L) &= 0.4(p) + 1.0(1-p) = 1 - 0.6p \\
EU_2(R) &= 1.2(p) + 0(1-p) = 1.2p
\end{aligned}
$$

$$
\begin{aligned}
EU_2(L) &= 1.0(r) + 0.6(1-r) = 0.6 + 0.4r \\
EU_2(R) &= 0(r) + 0.8(1-r) = 0.8 - 0.8r
\end{aligned}
$$

Herding and Information Cascades

$$
\begin{aligned}
P(R_U|R) &= \frac{P(R|R_U)P(R_U)}{P(R|R_U)P(R_U) + P(R|B_U)P(B_U)} \\
&= \frac{(\frac{3}{4})(0.5)}{(\frac{3}{4})(0.5) + (\frac{1}{4})(0.5)} \\
&= \frac{\frac{3}{8}}{\frac{3}{8} + \frac{1}{8}} = \frac{3}{4} = 0.75
\end{aligned}
$$

$$
\begin{aligned}
P(G|H) &= \frac{P(H|G)P(G)}{P(H|G)P(G) + P(H|B)P(B)} \\
&= \frac{(0.8)(0.5)}{(0.8)(0.5) + (0.2)(0.5)} = \frac{0.4}{0.4 + 0.1} = 0.8
\end{aligned}
$$

$$
\begin{aligned}
P(G|H,H) &= \frac{P(H|G)^2P(G)}{P(H|G)^2P(G) + P(H|B)^2P(B)} \\
&= \frac{(0.8)^2(0.5)}{(0.8)^2(0.5) + (0.2)^2(0.5)} = \frac{0.64}{0.64 + 0.04} \approx 0.941
\end{aligned}
$$

$$
\begin{aligned}
P(G|H,H,W) &= \frac{P(H|G)^2P(W|G)P(G)}{P(H|G)^2P(W|G)P(G) + P(H|B)^2P(W|B)P(B)} \\
&= \frac{(0.8)^2(0.2)(0.5)}{(0.8)^2(0.2)(0.5) + (0.2)^2(0.8)(0.5)} \\
&= \frac{(0.64)(0.1)}{(0.64)(0.1) + (0.04)(0.4)} = \frac{0.064}{0.064 + 0.016} = 0.8
\end{aligned}
$$

$$
\begin{aligned}
P(G|d=3) &= \frac{p^d}{p^d + (1-p)^d} \\
&= \frac{0.75^3}{0.75^3 + 0.25^3} \\
&= \frac{\frac{27}{64}}{\frac{27}{64} + \frac{1}{64}} \\
&= \frac{27}{28} \approx 0.964
\end{aligned}
$$

CHAPTER 30

Industry 40

Sourcing Decisions and Supplier Evaluation

$$r_{ij} = \frac{x_{ij}}{x_j^{\max}}$$

$$r_{ij} = \frac{x_j^{\min}}{x_{ij}}$$

$$S_i = \sum_{j=1}^{n} w_j r_{ij}$$

$$R = \begin{bmatrix} 0.900 & 0.944 & 0.700 \\ 0.720 & 1.000 & 1.000 \\ 1.000 & 0.833 & 0.500 \end{bmatrix}$$

$$r_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}$$

$$r_{ij} = \frac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}}$$

$$C_j = \sum_{i=1}^{n} a_{ij}$$

$$\bar{a}_{ij} = \frac{a_{ij}}{C_j}$$

$$w_i = \frac{1}{n} \sum_{j=1}^{n} \bar{a}_{ij}$$

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{bmatrix}$$

$$\bar{A} = \begin{bmatrix} 1/1.533 & 3/4.500 & 5/8.000 \\ 0.333/1.533 & 1/4.500 & 2/8.000 \\ 0.200/1.533 & 0.500/4.500 & 1/8.000 \end{bmatrix} \approx \begin{bmatrix} 0.652 & 0.667 & 0.625 \\ 0.217 & 0.222 & 0.250 \\ 0.130 & 0.111 & 0.125 \end{bmatrix}$$

Deterministic Inventory Models

$$\begin{aligned} TC &= \left(50 \times \frac{10000}{707.11}\right) + \left(2 \times \frac{707.11}{2}\right) + (10000 \times 10) \\ &= 707.11 + 707.11 + 100000 \\ &= \$101414.22 \end{aligned}$$

$$\begin{aligned} TC(1000) &= \left(\frac{1000}{1000} \times 20\right) + \left(\frac{1000}{2} \times 0.20 \times 3.90\right) + (1000 \times 3.90) \\ &= 20 + 390 + 3900 = \$4310 \end{aligned}$$

$$\begin{aligned}TC(500) &= \left(\frac{1000}{500} \times 20\right) + \left(\frac{500}{2} \times 0.20 \times 4.50\right) + (1000 \times 4.50) \\ &= 40 + 225 + 4500 = \$4765\end{aligned}$$

$$\begin{aligned}TC(200) &= \left(\frac{1000}{200} \times 20\right) + \left(\frac{200}{2} \times 0.20 \times 5.00\right) + (1000 \times 5.00) \\ &= 100 + 100 + 5000 = \$5200\end{aligned}$$

Stochastic Inventory Models

$$\begin{aligned}P_{ij} &= P(S - D_{n+1} = j) = P(D_{n+1} = S - j) = p_{S-j} && \text{for } 0 < j \leq S \\ P_{i0} &= P(S - D_{n+1} \leq 0) = P(D_{n+1} \geq S) = \sum_{d=S}^{\infty} p_d && \text{for } j = 0\end{aligned}$$

$$\begin{aligned}P_{ij} &= P(i - D_{n+1} = j) = P(D_{n+1} = i - j) = p_{i-j} && \text{for } 0 < j \leq i \\ P_{i0} &= P(i - D_{n+1} \leq 0) = P(D_{n+1} \geq i) = \sum_{d=i}^{\infty} p_d && \text{for } j = 0 \\ P_{ij} &= 0 && \text{for } j > i\end{aligned}$$

$$\begin{aligned}P_{i3} &= P(D = 3 - 3) = P(D = 0) = 0.2 \\ P_{i2} &= P(D = 3 - 2) = P(D = 1) = 0.5 \\ P_{i1} &= P(D = 3 - 1) = P(D = 2) = 0.3 \\ P_{i0} &= P(D \geq 3 - 0) = P(D \geq 3) = 0\end{aligned}$$

$$\begin{aligned}P_{23} &= P(D = -1) = 0 \quad \text{(cannot end higher than start)} \\ P_{22} &= P(D = 2 - 2) = P(D = 0) = 0.2 \\ P_{21} &= P(D = 2 - 1) = P(D = 1) = 0.5 \\ P_{20} &= P(D \geq 2 - 0) = P(D \geq 2) = P(D = 2) = 0.3\end{aligned}$$

$$\begin{aligned}P_{33} &= P(D = 3 - 3) = P(D = 0) = 0.2 \\ P_{32} &= P(D = 3 - 2) = P(D = 1) = 0.5 \\ P_{31} &= P(D = 3 - 1) = P(D = 2) = 0.3 \\ P_{30} &= P(D \geq 3) = 0\end{aligned}$$

$$\begin{aligned}\pi_0 &= 0\pi_0 + 0\pi_1 + 0.3\pi_2 + 0\pi_3 = 0.3\pi_2 \\ \pi_1 &= 0.3\pi_0 + 0.3\pi_1 + 0.5\pi_2 + 0.3\pi_3 \\ \pi_2 &= 0.5\pi_0 + 0.5\pi_1 + 0.2\pi_2 + 0.5\pi_3 \\ \pi_3 &= 0.2\pi_0 + 0.2\pi_1 + 0\pi_2 + 0.2\pi_3\end{aligned}$$

MDP Algorithms: Value and Policy Iteration

$$\begin{aligned}Q_1(s_1, a_1) &= R(s_1, a_1) + \gamma \left[P(s_1 \mid s_1, a_1)V_0(s_1) + P(s_2 \mid s_1, a_1)V_0(s_2)\right] \\ &= 5 + 0.5\,[0.8(0) + 0.2(0)] = 5 \\ Q_1(s_1, a_2) &= R(s_1, a_2) + \gamma \left[P(s_1 \mid s_1, a_2)V_0(s_1) + P(s_2 \mid s_1, a_2)V_0(s_2)\right] \\ &= 10 + 0.5\,[0.4(0) + 0.6(0)] = 10\end{aligned}$$

$$
\begin{aligned}
Q_1(s_2, a_1) &= R(s_2, a_1) + \gamma\,[0.1V_0(s_1) + 0.9V_0(s_2)] = -2 + 0 = -2 \\
Q_1(s_2, a_2) &= R(s_2, a_2) + \gamma\,[0.7V_0(s_1) + 0.3V_0(s_2)] = 4 + 0 = 4
\end{aligned}
$$

$$
\begin{aligned}
Q_2(s_1, a_1) &= 5 + 0.5\,[0.8(10) + 0.2(4)] = 5 + 0.5(8 + 0.8) = 5 + 4.4 = 9.4 \\
Q_2(s_1, a_2) &= 10 + 0.5\,[0.4(10) + 0.6(4)] = 10 + 0.5(4 + 2.4) = 10 + 3.2 = 13.2
\end{aligned}
$$

$$
\begin{aligned}
Q_2(s_2, a_1) &= -2 + 0.5\,[0.1(10) + 0.9(4)] = -2 + 0.5(1 + 3.6) = -2 + 2.3 = 0.3 \\
Q_2(s_2, a_2) &= 4 + 0.5\,[0.7(10) + 0.3(4)] = 4 + 0.5(7 + 1.2) = 4 + 4.1 = 8.1
\end{aligned}
$$

$$
\begin{aligned}
-0.05V^{\pi_0}(s_1) + 0.55(6V^{\pi_0}(s_1) - 50) &= -2 \\
-0.05V^{\pi_0}(s_1) + 3.3V^{\pi_0}(s_1) - 27.5 &= -2 \\
3.25V^{\pi_0}(s_1) = 25.5 \implies V^{\pi_0}(s_1) &= \frac{25.5}{3.25} = \frac{102}{13}
\end{aligned}
$$

$$
\begin{aligned}
Q(s_1, a_1) &= V^{\pi_0}(s_1) = \frac{102}{13} \\
Q(s_1, a_2) &= 10 + 0.5\,[0.4V^{\pi_0}(s_1) + 0.6V^{\pi_0}(s_2)] = 10 + 0.2\left(\frac{102}{13}\right) + 0.3\left(-\frac{38}{13}\right) \\
&= 10 + \frac{20.4}{13} - \frac{11.4}{13} = 10 + \frac{9}{13} = \frac{139}{13}
\end{aligned}
$$

$$
\begin{aligned}
Q(s_2, a_1) &= V^{\pi_0}(s_2) = -\frac{38}{13} \\
Q(s_2, a_2) &= 4 + 0.5\,[0.7V^{\pi_0}(s_1) + 0.3V^{\pi_0}(s_2)] = 4 + 0.35\left(\frac{102}{13}\right) + 0.15\left(-\frac{38}{13}\right) \\
&= 4 + \frac{35.7}{13} - \frac{5.7}{13} = 4 + \frac{30}{13} = \frac{82}{13}
\end{aligned}
$$

Time Series Forecasting

$$
\begin{aligned}
L_2 &= \alpha A_2 + (1 - \alpha)(L_1 + T_1) \\
&= 0.3(110) + 0.7(100 + 10) = 33 + 77 = 110 \\
T_2 &= \beta(L_2 - L_1) + (1 - \beta)T_1 \\
&= 0.2(110 - 100) + 0.8(10) = 0.2(10) + 8 = 2 + 8 = 10
\end{aligned}
$$

$$
\begin{aligned}
L_3 &= \alpha A_3 + (1 - \alpha)(L_2 + T_2) \\
&= 0.3(115) + 0.7(110 + 10) = 34.5 + 84 = 118.5 \\
T_3 &= \beta(L_3 - L_2) + (1 - \beta)T_2 \\
&= 0.2(118.5 - 110) + 0.8(10) = 0.2(8.5) + 8 = 1.7 + 8 = 9.7
\end{aligned}
$$

CHAPTER 31

Introduction To Big Data

Cloud Native Architectures and Resource Provisioning

$$\begin{aligned} Cost_{OD}(t) &= C_{OD} \times t \\ Cost_{RI}(t) &= U_{RI} + (C_{RI} \times t) \end{aligned}$$

$$\begin{aligned} 0.80t - 0.30t &= 1200 \\ 0.50t &= 1200 \\ t &= \frac{1200}{0.50} = 2400 \text{ hours} \end{aligned}$$

$$\begin{aligned} \text{Storage Cost} &= 50,000 \times 0.01 = \$500 \\ \text{Retrieval Cost} &= 5,000 \times 0.05 = \$250 \\ Cost_{cold} &= 500 + 250 = \$750 \text{ per month} \end{aligned}$$

Data Formats and Schema Design

$$\begin{aligned} y_1 &= 1000 \\ y_2 &= 1000 - 1000 = 0 \\ y_3 &= 1005 - 1000 = 5 \\ y_4 &= 1005 - 1005 = 0 \\ y_5 &= 1005 - 1005 = 0 \\ y_6 &= 1010 - 1005 = 5 \end{aligned}$$

Hadoop and HDFS Operations

$$\begin{aligned} \text{Cost} &= 1000 \times (0.70 \times 0 + 0.20 \times 2 + 0.10 \times 5) \\ &= 1000 \times (0 + 0.40 + 0.50) \\ &= 1000 \times 0.90 = 900 \text{ units} \end{aligned}$$

CHAPTER 32

Introduction To Computer System Design

Standard and Canonical Forms

$$\begin{aligned} A &= A(B+B')(C+C') \\ &= (AB+AB')(C+C') \\ &= ABC+ABC'+AB'C+AB'C' \end{aligned}$$

$$\begin{aligned} B'C &= B'C(A+A') \\ &= AB'C+A'B'C \end{aligned}$$

$$F(ABC) = (ABC+ABC'+AB'C+AB'C')+(AB'C+A'B'C)$$

$$F(ABC) = ABC+ABC'+AB'C+AB'C'+A'B'C$$

$$\begin{aligned} x+y &= x+y+zz' \\ &= (x+y+z)(x+y+z') \end{aligned}$$

$$\begin{aligned} x'+z &= x'+z+yy' \\ &= (x'+y+z)(x'+y'+z) \end{aligned}$$

$$F(xyz) = (x+y+z)(x+y+z')(x'+y+z)(x'+y'+z)$$

$$F(AB) = A'B'+AB'$$

$$F(AB) = (A+B')(A'+B')$$

Flip-Flops and Excitation Tables

$$Q_{n+1} = J\overline{Q_n} + \overline{K}Q_n$$

ALU and Computer Arithmetic Design

$$B_{inv} = Op_1 \cdot Op_0$$

$$B_{eff} = B \oplus B_{inv}$$

$$Y_{ADD} = A \oplus B_{eff} \oplus C_{in}$$

$$C_{out} = (A \cdot B_{eff}) + (C_{in} \cdot (A \oplus B_{eff}))$$

$$R = \overline{Op_1 Op_0} Y_{AND} + \overline{Op_1} Op_0 Y_{OR} + Op_1 \overline{Op_0} Y_{ADD} + Op_1 Op_0 Y_{ADD}$$

$$R = \overline{Op_1 Op_0}(A \cdot B) + \overline{Op_1} Op_0 (A + B) + Op_1 (A \oplus (B \oplus (Op_1 \cdot Op_0)) \oplus C_{in})$$

$$Z = \overline{R_{N-1} + R_{N-2} + \cdots + R_1 + R_0}$$

$$N = R_{N-1}$$

$$C = C_{out,N-1}$$

$$V = C_{in,N-1} \oplus C_{out,N-1}$$

$$V = (A_{N-1} \cdot B_{N-1} \cdot \overline{R_{N-1}}) + (\overline{A_{N-1}} \cdot \overline{B_{N-1}} \cdot R_{N-1})$$

$$Z = 0$$

$$N = 1$$

$$C = 0$$

$$V = C_{in,7} \oplus C_{out,7} = 1 \oplus 0 = 1$$

$$R = A + \overline{B} + 1$$

$$N = 0$$

$$C = 1$$

$$V = C_{in,7} \oplus C_{out,7} = 0 \oplus 1 = 1$$

CHAPTER 33

Introduction To Large Language Models

Normalization and Residual Connections

$$\begin{aligned}\sigma^2 &= \frac{(2-5)^2+(4-5)^2+(6-5)^2+(8-5)^2}{4} \\ &= \frac{(-3)^2+(-1)^2+(1)^2+(3)^2}{4} \\ &= \frac{9+1+1+9}{4} = \frac{20}{4} = 5\end{aligned}$$

$$\begin{aligned}\mathrm{RMS}(\mathbf{x}) &= \sqrt{\frac{3^2+(-4)^2+12^2+0^2}{4}} \\ &= \sqrt{\frac{9+16+144+0}{4}} = \sqrt{\frac{169}{4}} = \frac{13}{2} = 6.5\end{aligned}$$

Masked Language Modeling (BERT)

$$\begin{aligned}\sum_{j=1}^{4} \exp(z_{1,j}) &= \exp(1.0)+\exp(2.5)+\exp(0.5)+\exp(-1.0) \\ &\approx 2.7183+12.1825+1.6487+0.3679 \\ &= 16.9174\end{aligned}$$

$$\begin{aligned}L_1 &= -z_{1,2}+\log\left(\sum_{j=1}^{4}\exp(z_{1,j})\right) \\ &= -2.5+\log(16.9174) \\ &\approx -2.5+2.8283 = 0.3283\end{aligned}$$

$$\begin{aligned}\sum_{j=1}^{4} \exp(z_{2,j}) &= \exp(0.0)+\exp(-0.5)+\exp(1.0)+\exp(2.0) \\ &\approx 1.0000+0.6065+2.7183+7.3891 \\ &= 11.7139\end{aligned}$$

$$\begin{aligned}L_2 &= -z_{2,4}+\log\left(\sum_{j=1}^{4}\exp(z_{2,j})\right) \\ &= -2.0+\log(11.7139) \\ &\approx -2.0+2.4608 = 0.4608\end{aligned}$$

Causal Language Modeling (GPT)

$$\begin{aligned} P(\text{“The cat sat”}) &= P(\text{The}) \times P(\text{cat} \mid \text{The}) \times P(\text{sat} \mid \text{The cat}) \\ &= 0.2 \times 0.4 \times 0.5 \end{aligned}$$

$$P(\text{“The cat sat”}) = 0.08 \times 0.5 = 0.04$$

$$\begin{aligned} P(\text{“I love AI”}) &= P(\text{I}) \times P(\text{love} \mid \text{I}) \times P(\text{AI} \mid \text{I love}) \\ &= 0.5 \times 0.8 \times 0.25 \\ &= 0.4 \times 0.25 \\ &= 0.1 \end{aligned}$$

$$\begin{aligned} \text{Perplexity} &= P(W)^{-\frac{1}{n}} \\ &= 0.1^{-\frac{1}{3}} \\ &= \left(\frac{1}{0.1}\right)^{\frac{1}{3}} \\ &= \sqrt[3]{10} \approx 2.154 \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^{4} \ln P(w_i \mid w_{<i}) &= -1.2 + (-0.5) + (-2.1) + (-0.2) \\ &= -4.0 \end{aligned}$$

$$\frac{1}{n} \sum_{i=1}^{n} \ln P(w_i \mid w_{<i}) = \frac{-4.0}{4} = -1.0$$

$$\begin{aligned} \text{Perplexity} &= \exp\left(-(-1.0)\right) \\ &= \exp(1.0) \\ &\approx 2.718 \end{aligned}$$

Efficient Attention Mechanisms

$$M_{i,j} = 1 \text{ for } i \leq G \text{ or } j \leq G$$

$$M_{i,j} = 1 \text{ if } |i - j| \leq \lfloor W/2 \rfloor$$

$$M = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$M_{\text{causal}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\text{Total Computations} = 7 \times 16 = 112$$

CHAPTER 34

Introduction To Natural Language Processing I Nlp

Basic Text Processing and Normalization

$$D[i,j] = \min \begin{cases} D[i-1,j] + \text{cost}(\text{delete}) \\ D[i,j-1] + \text{cost}(\text{insert}) \\ D[i-1,j-1] + \text{cost}(\text{substitute}) & \text{if } X[i] \neq Y[j] \\ D[i-1,j-1] & \text{if } X[i] = Y[j] \end{cases}$$

$$\text{tf}_{t,d} = f_{t,d}$$

$$\text{idf}_t = \log_{10}\left(\frac{N}{df_t}\right)$$

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} \times \text{idf}_t$$

Syntax and Context-Free Grammars

$$\begin{aligned} S &\to ASA \mid aB \\ A &\to B \mid S \\ B &\to b \mid \epsilon \end{aligned}$$

$$\begin{aligned} S_0 &\to S \\ S &\to ASA \mid aB \\ A &\to B \mid S \\ B &\to b \mid \epsilon \end{aligned}$$

$$\begin{aligned} S_0 &\to S \\ S &\to ASA \mid aB \mid a \\ A &\to B \mid S \mid \epsilon \\ B &\to b \end{aligned}$$

$$\begin{aligned} S_0 &\to S \\ S &\to ASA \mid SA \mid AS \mid S \mid aB \mid a \\ A &\to B \mid S \\ B &\to b \end{aligned}$$

$$A \to b \mid S$$

$$A \to b \mid ASA \mid SA \mid AS \mid aB \mid a$$

$$S_0 \to ASA \mid SA \mid AS \mid aB \mid a$$

$$\begin{aligned} S_0 &\to ASA \mid SA \mid AS \mid aB \mid a \\ S &\to ASA \mid SA \mid AS \mid aB \mid a \\ A &\to b \mid ASA \mid SA \mid AS \mid aB \mid a \\ B &\to b \end{aligned}$$

$$\begin{aligned} S_0 &\to ASA \mid SA \mid AS \mid X_aB \mid a \\ S &\to ASA \mid SA \mid AS \mid X_aB \mid a \\ A &\to b \mid ASA \mid SA \mid AS \mid X_aB \mid a \\ B &\to b \\ X_a &\to a \end{aligned}$$

$$\begin{aligned} S_0 &\to AY_1 \mid SA \mid AS \mid X_aB \mid a \\ S &\to AY_1 \mid SA \mid AS \mid X_aB \mid a \\ A &\to b \mid AY_1 \mid SA \mid AS \mid X_aB \mid a \\ B &\to b \\ X_a &\to a \\ Y_1 &\to SA \end{aligned}$$

$$\begin{aligned} S &\to NP\ VP \\ VP &\to V\ NP \\ NP &\to Det\ N \mid \text{cats} \mid \text{fish} \\ V &\to \text{ate} \\ N &\to \text{cats} \mid \text{fish} \\ Det &\to \text{the} \end{aligned}$$

$$\begin{aligned} S &\to NP\ VP && (1.0) \\ VP &\to V\ NP && (0.7) \\ VP &\to VP\ PP && (0.3) \\ NP &\to NP\ PP && (0.4) \\ NP &\to \text{John} && (0.3) \\ NP &\to \text{Mary} && (0.3) \\ V &\to \text{saw} && (1.0) \\ P &\to \text{with} && (1.0) \\ PP &\to P\ NP && (1.0) \end{aligned}$$

Distributional Semantics and Word Embeddings

$$N = \sum_w \sum_c \text{count}(w, c)$$

$$\text{count}(w) = \sum_{c'} \text{count}(w, c')$$

$$\text{count}(c) = \sum_{w'} \text{count}(w', c)$$

$$\text{PMI}(w, c) = \log_2 \left(\frac{\text{count}(w, c) \cdot N}{\text{count}(w) \cdot \text{count}(c)} \right)$$

$$\text{PPMI}(w, c) = \max(0, \text{PMI}(w, c))$$

$$N = 10 + 2 + 8 + 5 + 15 + 0 + 0 + 1 + 4 = 45$$

$$\text{count}(\text{model}) = 10 + 2 + 8 = 20$$

$$\text{count}(\text{data}) = 10 + 5 + 0 = 15$$

$$\text{count}(\text{apple}) = 0 + 1 + 4 = 5$$

$$\text{count}(\text{data}) = 15$$

$$\text{PMI}(\text{model}, \text{data}) = \log_2 \left(\frac{10 \cdot 45}{20 \cdot 15} \right) = \log_2 \left(\frac{450}{300} \right) = \log_2(1.5) \approx 0.585$$

$$\text{PMI}(\text{apple}, \text{data}) = -\infty$$

$$\text{PPMI}(\text{model}, \text{data}) = \max(0, 0.585) = 0.585$$

$$\text{PPMI}(\text{apple}, \text{data}) = \max(0, -\infty) = 0$$

$$z_i = v_t \cdot u_i$$

$$\hat{y}_i = P(w_i | w_t) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

$$e_i = \hat{y}_i - y_i$$

$$u_i^{(new)} = u_i^{(old)} - \alpha \cdot e_i \cdot v_t$$

$$v_t^{(new)} = v_t^{(old)} - \alpha \sum_i (e_i \cdot u_i^{(old)})$$

$$v_{\text{cat}} = \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix}$$

$$u_{\text{cat}} = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix}, \quad u_{\text{sat}} = \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}, \quad u_{\text{mat}} = \begin{bmatrix} -0.2 \\ 0.3 \end{bmatrix}$$

$$z_{\text{cat}} = (0.5)(0.1) + (0.2)(-0.1) = 0.05 - 0.02 = 0.03$$

$$z_{\text{sat}} = (0.5)(0.4) + (0.2)(0.5) = 0.20 + 0.10 = 0.30$$

$$z_{\text{mat}} = (0.5)(-0.2) + (0.2)(0.3) = -0.10 + 0.06 = -0.04$$

$$\sum_j \exp(z_j) = \exp(0.03) + \exp(0.30) + \exp(-0.04) \approx 1.0305 + 1.3499 + 0.9608 = 3.3412$$

$$\hat{y}_{\text{cat}} = \frac{1.0305}{3.3412} \approx 0.308$$

$$\hat{y}_{\text{sat}} = \frac{1.3499}{3.3412} \approx 0.404$$

$$\hat{y}_{\text{mat}} = \frac{0.9608}{3.3412} \approx 0.288$$

$$e_{\text{cat}} = 0.308 - 0 = 0.308$$

$$e_{\text{sat}} = 0.404 - 1 = -0.596$$

$$e_{\text{mat}} = 0.288 - 0 = 0.288$$

$$u_{\text{cat}}^{(new)} = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix} - (0.1)(0.308) \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix} - \begin{bmatrix} 0.0154 \\ 0.0062 \end{bmatrix} = \begin{bmatrix} 0.0846 \\ -0.1062 \end{bmatrix}$$

$$u_{\text{sat}}^{(new)} = \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix} - (0.1)(-0.596) \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0.0298 \\ 0.0119 \end{bmatrix} = \begin{bmatrix} 0.4298 \\ 0.5119 \end{bmatrix}$$

$$u_{\text{mat}}^{(new)} = \begin{bmatrix} -0.2 \\ 0.3 \end{bmatrix} - (0.1)(0.288) \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} = \begin{bmatrix} -0.2 \\ 0.3 \end{bmatrix} - \begin{bmatrix} 0.0144 \\ 0.0058 \end{bmatrix} = \begin{bmatrix} -0.2144 \\ 0.2942 \end{bmatrix}$$

$$\sum_i e_i \cdot u_i^{(old)} = (0.308) \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix} + (-0.596) \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix} + (0.288) \begin{bmatrix} -0.2 \\ 0.3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0308 \\ -0.0308 \end{bmatrix} + \begin{bmatrix} -0.2384 \\ -0.2980 \end{bmatrix} + \begin{bmatrix} -0.0576 \\ 0.0864 \end{bmatrix} = \begin{bmatrix} -0.2652 \\ -0.2424 \end{bmatrix}$$

$$v_{\text{cat}}^{(new)} = \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} - (0.1) \begin{bmatrix} -0.2652 \\ -0.2424 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.0265 \\ 0.0242 \end{bmatrix} = \begin{bmatrix} 0.5265 \\ 0.2242 \end{bmatrix}$$

$$v_{\text{target}} = v_B - v_A + v_C$$

$$\cos(v_w, v_{\text{target}}) = \frac{v_w \cdot v_{\text{target}}}{\|v_w\| \|v_{\text{target}}\|}$$

$$v_{\text{Paris}} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad v_{\text{France}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad v_{\text{Tokyo}} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

$$v_{\text{Japan}} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}, \quad v_{\text{China}} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$$

$$v_{\text{target}} = v_{\text{France}} - v_{\text{Paris}} + v_{\text{Tokyo}}$$

$$v_{\text{target}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

$$\|v_{\text{target}}\| = \sqrt{6^2 + 2^2} = \sqrt{36 + 4} = \sqrt{40} \approx 6.32$$

$$\|v_{\text{Japan}}\| = \sqrt{6^2 + 2^2} = \sqrt{40} \approx 6.32$$

$$v_{\text{Japan}} \cdot v_{\text{target}} = (6)(6) + (2)(2) = 36 + 4 = 40$$

$$\cos(v_{\text{Japan}}, v_{\text{target}}) = \frac{40}{\sqrt{40} \times \sqrt{40}} = \frac{40}{40} = 1.0$$

$$\|v_{\text{China}}\| = \sqrt{7^2 + 5^2} = \sqrt{49 + 25} = \sqrt{74} \approx 8.60$$

$$v_{\text{China}} \cdot v_{\text{target}} = (7)(6) + (5)(2) = 42 + 10 = 52$$

$$\cos(v_{\text{China}}, v_{\text{target}}) = \frac{52}{8.60 \times 6.32} \approx \frac{52}{54.35} \approx 0.957$$

Recurrent Neural Networks (RNN) and LSTMs

$$\begin{aligned} z_2 &= W_{hx}x_2 + W_{hh}h_1 + b_h \\ &= \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} (2) + \begin{bmatrix} 0.1 & -0.1 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.4621 \\ 0.1974 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 1.0 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.1(0.4621) - 0.1(0.1974) \\ 0.0(0.4621) + 0.2(0.1974) \end{bmatrix} \\ &= \begin{bmatrix} 1.0 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.02647 \\ 0.03948 \end{bmatrix} = \begin{bmatrix} 1.02647 \\ 0.43948 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial W_{hx}} &= \left(\frac{\partial L}{\partial z_t}\right) x_t^\top \\ \frac{\partial L}{\partial W_{hh}} &= \left(\frac{\partial L}{\partial z_t}\right) h_{t-1}^\top \\ \frac{\partial L}{\partial b_h} &= \frac{\partial L}{\partial z_t} \end{aligned}$$

$$\begin{aligned} c_1 &= f_1 \odot c_0 + i_1 \odot \tilde{c}_1 \\ &= (0.5987)(0.5) + (0.7311)(-0.2913) \\ &= 0.29935 - 0.21297 = 0.08638 \end{aligned}$$

Large Language Models (LLMs) and Fine-Tuning

$$\text{Memory in GB} = \frac{P \times \text{bytes/parameter}}{10^9}$$

$$P_{\text{orig}} = d \times k = 1024 \times 512 = 524288$$

$$P_{\text{LoRA}} = r(d + k) = 8(1024 + 512) = 8 \times 1536 = 12288$$

$$\frac{12288}{524288} \times 100\% \approx 2.34\%$$

$$W = \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$h_{\text{orig}} = Wx = \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(2) + 2(1) \\ -1(2) + 0(1) \\ 0(2) + 1(1) \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}$$

$$z = Ax = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(2) + (-1)(1) \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix}$$

$$\Delta h = Bz = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$$

$$h_{\text{update}} = \frac{\alpha}{r} \Delta h = 2 \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ -4 \end{bmatrix}$$

$$h_{\text{final}} = h_{\text{orig}} + h_{\text{update}} = \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ -4 \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \\ -3 \end{bmatrix}$$

$$\Delta W = BA = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 2(1) & 2(-1) \\ 1(1) & 1(-1) \\ -2(1) & -2(-1) \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 1 & -1 \\ -2 & 2 \end{bmatrix}$$

$$2\Delta W = 2 \begin{bmatrix} 2 & -2 \\ 1 & -1 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 2 & -2 \\ -4 & 4 \end{bmatrix}$$

$$W_{\text{merged}} = W + 2\Delta W = \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & -4 \\ 2 & -2 \\ -4 & 4 \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ 1 & -2 \\ -4 & 5 \end{bmatrix}$$

$$h = W_{\text{merged}} x = \begin{bmatrix} 5 & -2 \\ 1 & -2 \\ -4 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5(2) + (-2)(1) \\ 1(2) + (-2)(1) \\ -4(2) + 5(1) \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \\ -3 \end{bmatrix}$$

$$\text{Optimizer Memory in GB} = \frac{P_{\text{train}} \times 8}{10^9}$$

CHAPTER 35

Linear Statistical Models

Formulating Simple Linear Models

$$\begin{aligned}
3x_1 + 2x_2 &= 12 \\
5x_1 + 4x_2 &= 22
\end{aligned}$$

$$\begin{aligned}
4x_A + 3x_B &= 120 \\
2x_A + 5x_B &= 150
\end{aligned}$$

$$\begin{aligned}
x_1 - x_2 + x_3 &= 50 \\
2x_1 + x_2 - x_4 &= 100 \\
x_2 - 3x_3 + x_4 &= -20
\end{aligned}$$

$$\begin{aligned}
1x_1 - 1x_2 + 1x_3 + 0x_4 &= 50 \\
2x_1 + 1x_2 + 0x_3 - 1x_4 &= 100 \\
0x_1 + 1x_2 - 3x_3 + 1x_4 &= -20
\end{aligned}$$

Calculating and Interpreting Residuals

$$\begin{aligned}
\text{RSS} = \mathbf{e}^T\mathbf{e} &= \begin{bmatrix} -0.5 & 1.0 & -0.5 \end{bmatrix} \begin{bmatrix} -0.5 \\ 1.0 \\ -0.5 \end{bmatrix} \\
&= (-0.5)^2 + (1.0)^2 + (-0.5)^2 \\
&= 0.25 + 1.0 + 0.25 = 1.5
\end{aligned}$$

The Method of Least Squares

$$\begin{aligned}
4\beta_2 = 4 &\implies \beta_2 = 1 \\
-10\beta_1 - 10(1) = -10 &\implies \beta_1 = 0 \\
2\beta_0 + 6(0) + 8(1) = 8 &\implies \beta_0 = 0
\end{aligned}$$

Testing General Linear Hypotheses

$$\begin{aligned}
\beta_1 - \beta_2 &= 0 \\
\beta_3 &= 0 \\
\beta_1 + \beta_2 + \beta_3 &= 1
\end{aligned}$$

$$\begin{aligned}
&\text{Equation 1:} \quad 0 \cdot \beta_0 + 1 \cdot \beta_1 - 1 \cdot \beta_2 + 0 \cdot \beta_3 = 0 \\
&\text{Equation 2:} \quad 0 \cdot \beta_0 + 0 \cdot \beta_1 + 0 \cdot \beta_2 + 1 \cdot \beta_3 = 0 \\
&\text{Equation 3:} \quad 0 \cdot \beta_0 + 1 \cdot \beta_1 + 1 \cdot \beta_2 + 1 \cdot \beta_3 = 1
\end{aligned}$$

Construction and Analysis of ANOVA Tables

$$19 = 3 + DF_{Err} \implies DF_{Err} = 16$$

$$40 = \frac{SS_{Trt}}{3} \implies SS_{Trt} = 120$$

$$5 = \frac{SS_{Err}}{16} \implies SS_{Err} = 80$$

$$SS_{Tot} = 120 + 80 = 200$$

$$F = \frac{40}{5} = 8$$

$$150 = 50 + 60 + SS_{Err} \implies SS_{Err} = 40$$

$$F_B = \frac{MS_B}{MS_{Err}} \implies 4 = \frac{20}{MS_{Err}} \implies MS_{Err} = 5$$

$$F_{AB} = \frac{MS_{AB}}{MS_{Err}} \implies 1 = \frac{MS_{AB}}{5} \implies MS_{AB} = 5$$

$$MS_{AB} = \frac{SS_{AB}}{DF_{AB}} \implies 5 = \frac{30}{DF_{AB}} \implies DF_{AB} = 6$$

$$6 = 3 \times DF_B \implies DF_B = 2$$

$$SS_B = MS_B \times DF_B = 20 \times 2 = 40$$

$$DF_{Tot} = DF_A + DF_B + DF_{AB} + DF_{Err} \implies 31 = 3 + 2 + 6 + DF_{Err} \implies DF_{Err} = 20$$

$$SS_{Err} = MS_{Err} \times DF_{Err} = 5 \times 20 = 100$$

$$MS_A = \frac{SS_A}{DF_A} = \frac{45}{3} = 15$$

$$F_A = \frac{MS_A}{MS_{Err}} = \frac{15}{5} = 3$$

$$SS_{Tot} = SS_A + SS_B + SS_{AB} + SS_{Err} = 45 + 40 + 30 + 100 = 215$$

CHAPTER 36

Machine Learning Foundations

Multivariate Calculus

$$
\begin{aligned}
L(3.1, 3.9) &= f(3,4) + \nabla f(3,4) \cdot (\mathbf{x} - \mathbf{a}) \\
&= 5 + \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix} \\
&= 5 + (0.6 \times 0.1) + (0.8 \times -0.1) \\
&= 5 + 0.06 - 0.08 \\
&= 4.98
\end{aligned}
$$

Four Fundamental Subspaces

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & -2 \\ 3 & 6 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & -2 \\ 0 & 0 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\text{RREF}(A) = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

$$x_1 + 2x_3 = 0 \implies x_1 = -2x_3$$

$$x_2 + 2x_4 = 0 \implies x_2 = -2x_4$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ -2x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 2 & 4 & 6 & | & 0 & 1 & 0 \\ 1 & 1 & 2 & | & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 0 & 0 & | & -2 & 1 & 0 \\ 0 & -1 & -1 & | & -1 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 1 & 1 & | & 1 & 0 & -1 \\ 0 & 0 & 0 & | & -2 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 1 & | & -1 & 0 & 2 \\ 0 & 1 & 1 & | & 1 & 0 & -1 \\ 0 & 0 & 0 & | & -2 & 1 & 0 \end{bmatrix}$$

$$\text{Basis for } C(A) = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \right\}$$

$$\text{Basis for } C(A^T) = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

$$\text{Basis for } N(A) = \left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

$$\text{Basis for } N(A^T) = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$$

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \end{bmatrix}$$

$$R_{A^T} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -2y_2 - 3y_3 \\ y_2 \\ y_3 \end{bmatrix} = y_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + y_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

Complex and Hermitian Matrices

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^H \mathbf{v} &= \begin{bmatrix} 1 - i & 2 \end{bmatrix} \begin{bmatrix} i \\ 1 - i \end{bmatrix} \\ &= (1 - i)(i) + (2)(1 - i) \\ &= (i - i^2) + (2 - 2i) \\ &= (i - (-1)) + 2 - 2i \\ &= (1 + i) + 2 - 2i = 3 - i \end{aligned}$$

$$
\begin{aligned}
\|\mathbf{u}\|^2 = \mathbf{u}^H\mathbf{u} &= \begin{bmatrix} 1-i & 2 \end{bmatrix} \begin{bmatrix} 1+i \\ 2 \end{bmatrix} \\
&= (1-i)(1+i) + (2)(2) \\
&= (1^2 + 1^2) + 4 = 2 + 4 = 6
\end{aligned}
$$

$$
\begin{aligned}
U^H U &= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \right) \\
&= \frac{1}{2} \begin{bmatrix} (1)(1) + (-i)(i) & (1)(i) + (-i)(1) \\ (-i)(1) + (1)(i) & (-i)(i) + (1)(1) \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} 1 - (-1) & i - i \\ -i + i & -(-1) + 1 \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I
\end{aligned}
$$

$$
\begin{aligned}
\det(A - \lambda I) &= \det \begin{bmatrix} 2-\lambda & 1-i \\ 1+i & 1-\lambda \end{bmatrix} \\
&= (2-\lambda)(1-\lambda) - (1-i)(1+i) \\
&= (2 - 3\lambda + \lambda^2) - (1^2 - i^2) \\
&= \lambda^2 - 3\lambda + 2 - 2 \\
&= \lambda^2 - 3\lambda = 0
\end{aligned}
$$

Singular Value Decomposition (SVD)

$$
\begin{aligned}
(1,1) &= \frac{1}{10} + \frac{1}{18} = \frac{9+5}{90} = \frac{14}{90} = \frac{7}{45} \\
(1,2) &= \frac{1}{10} - \frac{1}{18} = \frac{9-5}{90} = \frac{4}{90} = \frac{2}{45} \\
(1,3) &= 0 + \frac{4}{18} = \frac{10}{45} \\
(2,1) &= \frac{1}{10} - \frac{1}{18} = \frac{2}{45} \\
(2,2) &= \frac{1}{10} + \frac{1}{18} = \frac{7}{45} \\
(2,3) &= 0 - \frac{4}{18} = -\frac{10}{45}
\end{aligned}
$$

Convex Sets and Convex Functions

$$
\begin{aligned}
\mathbf{a}^T\mathbf{z} &= \mathbf{a}^T(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \\
&= \theta\mathbf{a}^T\mathbf{x} + (1-\theta)\mathbf{a}^T\mathbf{y}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{a}^T\mathbf{z} &= \theta b + (1-\theta)b \\
&= b(\theta + 1 - \theta) \\
&= b
\end{aligned}
$$

CHAPTER 37

Machine Learning Operations Mlops Rangarajan Vasudevan

Model Deployment and Continuous Deployment (CD)

$$Z = \frac{p_B - p_A}{\sqrt{p(1-p)\left(\frac{1}{N_A} + \frac{1}{N_B}\right)}}$$

$$SE = \sqrt{0.11(0.89)\left(\frac{1}{1000} + \frac{1}{1000}\right)} = \sqrt{0.11 \times 0.89 \times 0.002} = \sqrt{0.0001958} \approx 0.014$$

$$Z = \frac{0.12 - 0.10}{0.014} \approx 1.43$$

$$\text{Agreement} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\hat{y}^{(i)}_{\text{primary}} = \hat{y}^{(i)}_{\text{shadow}}\right)$$

$$\hat{Y}_{\text{primary}} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \hat{Y}_{\text{shadow}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

ML Systems Logging and Monitoring

$$(45 - 48)^2 = 9$$
$$(50 - 48)^2 = 4$$
$$(48 - 48)^2 = 0$$
$$(52 - 48)^2 = 16$$
$$(45 - 48)^2 = 9$$

CHAPTER 38

Machine Learning Practice Course

Feature Engineering and Dimensionality Reduction

$$x_i^{\text{new}} = \frac{x_i - \mu}{\sigma}$$

$$x_i^{\text{new}} = \frac{x_i - \min(X)}{\max(X) - \min(X)}$$

$$\min(X) = 10, \quad \max(X) = 50$$

$$\text{For } x_1 = 10: \quad x_1^{\text{new}} = \frac{10 - 10}{50 - 10} = \frac{0}{40} = 0.0$$

$$\text{For } x_2 = 20: \quad x_2^{\text{new}} = \frac{20 - 10}{50 - 10} = \frac{10}{40} = 0.25$$

$$\text{For } x_3 = 30: \quad x_3^{\text{new}} = \frac{30 - 10}{50 - 10} = \frac{20}{40} = 0.5$$

$$\text{For } x_4 = 40: \quad x_4^{\text{new}} = \frac{40 - 10}{50 - 10} = \frac{30}{40} = 0.75$$

$$\text{For } x_5 = 50: \quad x_5^{\text{new}} = \frac{50 - 10}{50 - 10} = \frac{40}{40} = 1.0$$

$$z_i = \frac{x_i - \mu}{\sigma}$$

$$\mu = \frac{2 + 4 + 5 + 6 + 40}{5} = \frac{57}{5} = 11.4$$

$$\sigma = \sqrt{\frac{(2 - 11.4)^2 + (4 - 11.4)^2 + (5 - 11.4)^2 + (6 - 11.4)^2 + (40 - 11.4)^2}{5}} \approx 14.33$$

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR}$$

$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR}$$

$$\text{Lower Bound} = 3 - 1.5(4) = -3$$

$$\text{Upper Bound} = 7 + 1.5(4) = 13$$

$$x_i^{\text{new}} = \begin{cases} \text{Upper Bound}, & \text{if } x_i > \text{Upper Bound} \\ \text{Lower Bound}, & \text{if } x_i < \text{Lower Bound} \\ x_i, & \text{otherwise} \end{cases}$$

$$\text{Lower Bound} = -3, \quad \text{Upper Bound} = 13$$

$$X^{\text{new}} = \{1, 3, 4, 5, 6, 7, 13\}$$

$$S = \frac{1}{n-1} X_c^T X_c$$

$$Z = X_c W$$

$$X = \begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \\ 5 & 6 \end{bmatrix}$$

$$X_c = \begin{bmatrix} 2-3.5 & 3-4.5 \\ 3-3.5 & 4-4.5 \\ 4-3.5 & 5-4.5 \\ 5-3.5 & 6-4.5 \end{bmatrix} = \begin{bmatrix} -1.5 & -1.5 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \end{bmatrix}$$

$$S = \frac{1}{4-1} X_c^T X_c = \frac{1}{3} \begin{bmatrix} -1.5 & -0.5 & 0.5 & 1.5 \\ -1.5 & -0.5 & 0.5 & 1.5 \end{bmatrix} \begin{bmatrix} -1.5 & -1.5 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \end{bmatrix}$$

$$S = \frac{1}{3} \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} 5/3 & 5/3 \\ 5/3 & 5/3 \end{bmatrix}$$

$$\det(S - \lambda I) = \begin{vmatrix} 5/3 - \lambda & 5/3 \\ 5/3 & 5/3 - \lambda \end{vmatrix} = 0$$

$$(5/3 - \lambda)^2 - (5/3)^2 = 0 \implies \lambda_1 = 10/3, \lambda_2 = 0$$

$$v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \approx \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$$

$$Z = X_c W = \begin{bmatrix} -1.5 & -1.5 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \end{bmatrix} \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} = \begin{bmatrix} -2.121 \\ -0.707 \\ 0.707 \\ 2.121 \end{bmatrix}$$

$$\text{PVE} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$V_{\text{total}} = 4.2 + 1.1 + 0.5 + 0.2 = 6.0$$

Regularization and Hyperparameter Tuning

$$J(\mathbf{w}, b) = \frac{1}{2n} \sum_{i=1}^{n} (h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^{d} w_j^2$$

$$w_j := w_j - \alpha \left(\frac{1}{n} \sum_{i=1}^{n} (h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)} + \lambda w_j \right) \quad \text{for } j \geq 1$$

$$b := b - \alpha \left(\frac{1}{n} \sum_{i=1}^{n} (h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}) \right)$$

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

$$\mathbf{X} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = [1^2 + 2^2 + 3^2] = [14]$$

$$\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I} = [14] + 1 \cdot [1] = [15]$$

$$(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1} = [15]^{-1} = \left[\frac{1}{15}\right]$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = [1(2) + 2(3) + 3(5)] = [2 + 6 + 15] = [23]$$

$$\mathbf{w} = \left[\frac{1}{15}\right][23] = \left[\frac{23}{15}\right] \approx 1.533$$

$$J(\mathbf{w}, b) = \frac{1}{2n}\sum_{i=1}^{n}(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{d}|w_j|$$

$$w_j := w_j - \alpha\left(\frac{1}{n}\sum_{i=1}^{n}(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)})x_j^{(i)} + \lambda\,\text{sign}(w_j)\right) \quad \text{for } j \geq 1$$

$$\text{sign}(w_1) = \text{sign}(0.5) = 1$$

$$\text{sign}(w_2) = \text{sign}(-0.2) = -1$$

$$\nabla_{w_1}J = \nabla_{w_1}MSE + \lambda\,\text{sign}(w_1) = 1.2 + 0.5(1) = 1.7$$

$$\nabla_{w_2}J = \nabla_{w_2}MSE + \lambda\,\text{sign}(w_2) = -0.8 + 0.5(-1) = -1.3$$

$$w_1 := w_1 - \alpha\nabla_{w_1}J = 0.5 - 0.1(1.7) = 0.5 - 0.17 = 0.33$$

$$w_2 := w_2 - \alpha\nabla_{w_2}J = -0.2 - 0.1(-1.3) = -0.2 + 0.13 = -0.07$$

$$J(\mathbf{w}, b) = \frac{1}{2n}\sum_{i=1}^{n}(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)})^2 + \lambda_1\sum_{j=1}^{d}|w_j| + \frac{\lambda_2}{2}\sum_{j=1}^{d}w_j^2$$

$$J(\mathbf{w}, b) = \frac{1}{2n}\sum_{i=1}^{n}(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)})^2 + \lambda r\sum_{j=1}^{d}|w_j| + \frac{\lambda(1-r)}{2}\sum_{j=1}^{d}w_j^2$$

$$w_j := w_j - \alpha\left(\nabla_{w_j}MSE + \lambda r\,\text{sign}(w_j) + \lambda(1-r)w_j\right)$$

$$\text{Avg MSE}(\lambda = 0.1) = \frac{4.5 + 5.2 + 4.8}{3} = \frac{14.5}{3} \approx 4.83$$

$$\text{Avg MSE}(\lambda = 1.0) = \frac{3.8 + 4.1 + 3.9}{3} = \frac{11.8}{3} \approx 3.93$$

Classification Metrics and Evaluation

$$C = \begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix}$$

$$C = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Accuracy} = \frac{40 + 50}{40 + 50 + 10 + 5} = \frac{90}{105} \approx 0.857$$

$$\text{Precision} = \frac{40}{40 + 10} = \frac{40}{50} = 0.800$$

$$\text{Recall} = \frac{40}{40 + 5} = \frac{40}{45} \approx 0.889$$

$$\text{F1-Score} = \frac{2 \times 0.800 \times 0.889}{0.800 + 0.889} = \frac{1.4224}{1.689} \approx 0.842$$

$$\text{Specificity} = \frac{50}{50 + 10} = \frac{50}{60} \approx 0.833$$

$$P_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k}, \quad R_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k}$$

$$\text{Macro Precision} = \frac{1}{K}\sum_{k=1}^{K} P_k, \quad \text{Macro Recall} = \frac{1}{K}\sum_{k=1}^{K} R_k$$

$$\text{Micro Precision} = \frac{\sum \text{TP}_k}{\sum \text{TP}_k + \sum \text{FP}_k}, \quad \text{Micro Recall} = \frac{\sum \text{TP}_k}{\sum \text{TP}_k + \sum \text{FN}_k}$$

$$C = \begin{bmatrix} 10 & 2 & 1 \\ 1 & 15 & 2 \\ 0 & 3 & 12 \end{bmatrix}$$

$$\text{Macro Precision} = \frac{0.909 + 0.750 + 0.800}{3} \approx 0.820$$

$$\text{Macro Recall} = \frac{0.769 + 0.833 + 0.800}{3} \approx 0.801$$

$$\text{Micro Precision} = \frac{37}{37 + 9} = \frac{37}{46} \approx 0.804$$

$$\text{Micro Recall} = \frac{37}{37 + 9} = \frac{37}{46} \approx 0.804$$

Probabilistic and Distance-Based Classifiers

$$\begin{aligned}\text{Score}(\text{Yes}) &= P(\text{Yes}) \times P(\text{Rain} \mid \text{Yes}) \times P(\text{Strong} \mid \text{Yes}) \\ &= 0.625 \times 0.4 \times 0.2 = 0.05\end{aligned}$$

$$\begin{aligned}\text{Score}(\text{No}) &= P(\text{No}) \times P(\text{Rain} \mid \text{No}) \times P(\text{Strong} \mid \text{No}) \\ &= 0.375 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{24} = \frac{1}{12} \approx 0.0833\end{aligned}$$

$$\begin{aligned}P(\text{Low} \mid \text{Yes}) &= \frac{\text{Count}(\text{Low and Yes})}{\text{Count}(\text{Yes})} \\ &= \frac{0}{5} = 0\end{aligned}$$

$$\begin{aligned}P(\text{Low} \mid \text{Yes}) &= \frac{0+1}{5+1\cdot 3} \\ &= \frac{1}{8} = 0.125\end{aligned}$$

Ensemble Learning I: Bagging and Random Forests

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_B(x)\}$$

$$\hat{y} = \frac{1}{B}\sum_{i=1}^{B} h_i(x)$$

$$h_1(x) = 1, \quad h_2(x) = 0, \quad h_3(x) = 1, \quad h_4(x) = 1, \quad h_5(x) = 0$$

$$h_1(x) = 250, \quad h_2(x) = 265, \quad h_3(x) = 240, \quad h_4(x) = 255$$

$$250 + 265 + 240 + 255 = 1010$$

$$\hat{y} = \frac{1010}{4} = 252.5$$

$$\Delta I(t) = \frac{N_t}{N}\left(I(t) - \frac{N_{t_L}}{N_t} I(t_L) - \frac{N_{t_R}}{N_t} I(t_R)\right)$$

$$\Delta I = \frac{100}{100}\left(0.5 - \frac{40}{100}(0.2) - \frac{60}{100}(0.3)\right) = 0.5 - 0.08 - 0.18 = 0.24$$

$$\Delta I = \frac{40}{100}\,(0.2 - 0 - 0) = 0.08$$

$$\Delta I = \frac{100}{100}\left(0.48 - \frac{50}{100}(0.1) - \frac{50}{100}(0.1)\right) = 0.48 - 0.05 - 0.05 = 0.38$$

Artificial Neural Networks

$$\mathbf{z}^{(1)} = \mathbf{W}^{(1)}\mathbf{a}^{(0)} + \mathbf{b}^{(1)} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1-2 \\ -1+2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$\mathbf{a}^{(1)} = \max(0, \mathbf{z}^{(1)}) = \begin{bmatrix} \max(0,-1) \\ \max(0,2) \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$$\mathbf{z}^{(2)} = \mathbf{W}^{(2)}\mathbf{a}^{(1)} + \mathbf{b}^{(2)} = \begin{bmatrix} 2 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} = (0-6) + 1 = -5$$

$$\mathbf{a}^{(2)} = \mathbf{z}^{(2)} = -5$$

$$P(y = i|\mathbf{x}) = a_i^{(L)} = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

$$\sum_{j=1}^{3} e^{z_j} \approx 2.718 + 1 + 0.368 = 4.086$$

$$\mathcal{L}(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

$$\mathcal{L}(y, \hat{y}) = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})]$$

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

$$\mathcal{L}(1, 0.8) = -[1 \cdot \log(0.8) + (1-1) \cdot \log(1-0.8)]$$

$$\mathcal{L}(1, 0.8) = -\log(0.8) \approx -(-0.223) = 0.223$$

$$\mathcal{L}(0, 0.8) = -[0 \cdot \log(0.8) + (1-0) \cdot \log(1-0.8)]$$

$$\mathcal{L}(0, 0.8) = -\log(0.2) \approx -(-1.609) = 1.609$$

CHAPTER 39

Machine Learning Practice Course Week 1 Amp 2

Hyperparameter Tuning

$$\text{param\_grid} = \Big[\\ \{\texttt{'kernel'} : [\texttt{'linear'}], \texttt{'C'} : [1, 10]\}, \\ \{\texttt{'kernel'} : [\texttt{'rbf'}], \texttt{'C'} : [1, 10], \texttt{'gamma'} : [0.01, 0.1]\} \\ \Big]$$

$$\text{param\_grid} = \Big[\\ \{\texttt{'clf'} : [\text{RandomForestClassifier}()], \texttt{'clf\_\_max\_depth'} : [3, 5]\}, \\ \{\texttt{'clf'} : [\text{LogisticRegression}()], \texttt{'clf\_\_C'} : [0.1, 1.0]\} \\ \Big]$$

CHAPTER 40

Machine Learning Techniques Prof Arunkumar

K-Means Clustering

$$C_1 = \{2, 3\}$$
$$C_2 = \{4, 10, 11, 12, 20, 25, 30\}$$

$$\mu_1^{(new)} = \frac{2+3}{2} = 2.5$$
$$\mu_2^{(new)} = \frac{4+10+11+12+20+25+30}{7} = \frac{112}{7} = 16$$

$$C_1 = \{(1, 1), (2, 1)\}$$
$$C_2 = \{(4, 3), (5, 4)\}$$

$$\mu_1^{(new)} = \left(\frac{1+2}{2}, \frac{1+1}{2}\right) = (1.5, 1)$$
$$\mu_2^{(new)} = \left(\frac{4+5}{2}, \frac{3+4}{2}\right) = (4.5, 3.5)$$

$$C_1 = \{1, 2\}$$
$$C_2 = \{6, 7\}$$
$$C_3 = \{10, 11\}$$

$$\mu_1^{(new)} = \frac{1+2}{2} = 1.5$$
$$\mu_2^{(new)} = \frac{6+7}{2} = 6.5$$
$$\mu_3^{(new)} = \frac{10+11}{2} = 10.5$$

Parameter Estimation: MLE and MAP

$$\begin{aligned}
\ell(p) &= \sum_{i=1}^{n} \log\left(p^{x_i}(1-p)^{1-x_i}\right) \\
&= \sum_{i=1}^{n} \left(x_i \log p + (1-x_i)\log(1-p)\right) \\
&= \left(\sum_{i=1}^{n} x_i\right)\log p + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p)
\end{aligned}$$

$$\begin{aligned}
\frac{k}{p} - \frac{n-k}{1-p} &= 0 \\
\frac{k}{p} &= \frac{n-k}{1-p} \\
k(1-p) &= p(n-k) \\
k - kp &= np - kp \\
p &= \frac{k}{n}
\end{aligned}$$

$$\begin{aligned}
\ell(\mu) &= \sum_{i=1}^{n} \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} \right) \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) &= 0 \\
\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu &= 0 \\
\sum_{i=1}^{n} x_i - n\mu &= 0 \\
\mu &= \frac{1}{n}\sum_{i=1}^{n} x_i
\end{aligned}$$

$$\begin{aligned}
\frac{k+\alpha-1}{p} &= \frac{n-k+\beta-1}{1-p} \\
(k+\alpha-1)(1-p) &= p(n-k+\beta-1) \\
(k+\alpha-1) - p(k+\alpha-1) &= p(n-k+\beta-1) \\
k+\alpha-1 &= p(n-k+\beta-1+k+\alpha-1) \\
k+\alpha-1 &= p(n+\alpha+\beta-2) \\
p &= \frac{k+\alpha-1}{n+\alpha+\beta-2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell_{MAP}(\mu)}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) - \frac{\mu-\mu_0}{\sigma_0^2} \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right) - \frac{\mu-\mu_0}{\sigma_0^2}
\end{aligned}$$

$$\begin{aligned}
\frac{n\mu}{\sigma^2} + \frac{\mu}{\sigma_0^2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i + \frac{\mu_0}{\sigma_0^2} \\
\mu\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right) &= \frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}
\end{aligned}$$

K-Nearest Neighbors and Decision Trees

$$\begin{aligned}
d(P_1, \mathbf{x}) &= \sqrt{(1-2)^2 + (2-2)^2} = \sqrt{(-1)^2 + 0^2} = 1 \\
d(P_2, \mathbf{x}) &= \sqrt{(2-2)^2 + (3-2)^2} = \sqrt{0^2 + 1^2} = 1 \\
d(P_3, \mathbf{x}) &= \sqrt{(3-2)^2 + (3-2)^2} = \sqrt{1^2 + 1^2} = \sqrt{2} \approx 1.414 \\
d(P_4, \mathbf{x}) &= \sqrt{(5-2)^2 + (4-2)^2} = \sqrt{3^2 + 2^2} = \sqrt{13} \approx 3.606 \\
d(P_5, \mathbf{x}) &= \sqrt{(1-2)^2 + (4-2)^2} = \sqrt{(-1)^2 + 2^2} = \sqrt{5} \approx 2.236
\end{aligned}$$

$$p_{\text{yes}} = \frac{9}{14} \approx 0.643$$
$$p_{\text{no}} = \frac{5}{14} \approx 0.357$$

$$\begin{aligned} H(S) &= -p_{\text{yes}} \log_2(p_{\text{yes}}) - p_{\text{no}} \log_2(p_{\text{no}}) \\ &= -\left(\frac{9}{14}\right) \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2\left(\frac{5}{14}\right) \end{aligned}$$

$$\begin{aligned} \log_2\left(\frac{9}{14}\right) &\approx -0.637 \\ \log_2\left(\frac{5}{14}\right) &\approx -1.485 \\ H(S) &\approx -(0.643 \times -0.637) - (0.357 \times -1.485) \\ &= 0.410 + 0.530 = 0.940 \text{ bits} \end{aligned}$$

$$\begin{aligned} p_{\text{yes}} &= \frac{4}{6} = \frac{2}{3}, \quad p_{\text{no}} = \frac{2}{6} = \frac{1}{3} \\ H(S_{\text{weak}}) &= -\left(\frac{2}{3}\right) \log_2\left(\frac{2}{3}\right) - \left(\frac{1}{3}\right) \log_2\left(\frac{1}{3}\right) \approx 0.918 \end{aligned}$$

$$\begin{aligned} p_{\text{yes}} &= \frac{2}{4} = \frac{1}{2}, \quad p_{\text{no}} = \frac{2}{4} = \frac{1}{2} \\ H(S_{\text{strong}}) &= -\left(\frac{1}{2}\right) \log_2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) \log_2\left(\frac{1}{2}\right) = 1.000 \end{aligned}$$

$$\begin{aligned} H(S|\text{Wind}) &= \frac{|S_{\text{weak}}|}{|S|} H(S_{\text{weak}}) + \frac{|S_{\text{strong}}|}{|S|} H(S_{\text{strong}}) \\ &= \frac{6}{10}(0.918) + \frac{4}{10}(1.000) \\ &= 0.551 + 0.400 = 0.951 \end{aligned}$$

$$\begin{aligned} IG(S, \text{Wind}) &= H(S) - H(S|\text{Wind}) \\ &= 0.971 - 0.951 = 0.020 \end{aligned}$$

$$\begin{aligned} p_{\text{yes}} &= \frac{6}{10} = 0.6, \quad p_{\text{no}} = \frac{4}{10} = 0.4 \\ G(S) &= 1 - (0.6^2 + 0.4^2) = 1 - (0.36 + 0.16) = 1 - 0.52 = 0.480 \end{aligned}$$

$$\begin{aligned} p_{\text{yes}} &= \frac{4}{6}, \quad p_{\text{no}} = \frac{2}{6} \\ G(S_{\text{weak}}) &= 1 - \left(\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right) = 1 - \left(\frac{16}{36} + \frac{4}{36}\right) = 1 - \frac{20}{36} = \frac{16}{36} \approx 0.444 \end{aligned}$$

$$\begin{aligned} p_{\text{yes}} &= \frac{2}{4}, \quad p_{\text{no}} = \frac{2}{4} \\ G(S_{\text{strong}}) &= 1 - \left(\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right) = 1 - (0.25 + 0.25) = 0.500 \end{aligned}$$

$$\begin{aligned} G(S|\text{Wind}) &= \frac{6}{10}G(S_{\text{weak}}) + \frac{4}{10}G(S_{\text{strong}}) \\ &= \frac{6}{10}\left(\frac{16}{36}\right) + \frac{4}{10}(0.500) \\ &= \frac{16}{60} + 0.200 \approx 0.267 + 0.200 = 0.467 \end{aligned}$$

$$\begin{aligned} \text{Gini Gain} &= G(S) - G(S|\text{Wind}) \\ &= 0.480 - 0.467 = 0.013 \end{aligned}$$

Naive Bayes Classifiers

$$\begin{aligned} P(\text{Rainy} \mid \text{Yes}) &= \frac{1}{3} \approx 0.333 \\ P(\text{Hot} \mid \text{Yes}) &= \frac{0}{3} = 0 \end{aligned}$$

$$\begin{aligned} P(\text{Rainy} \mid \text{No}) &= \frac{1}{2} = 0.5 \\ P(\text{Hot} \mid \text{No}) &= \frac{1}{2} = 0.5 \end{aligned}$$

$$\begin{aligned} P(\text{`cheap'} \mid \text{Spam}) &= \frac{4+1}{9+1\times 4} = \frac{5}{13} \approx 0.385 \\ P(\text{`viagra'} \mid \text{Spam}) &= \frac{2+1}{9+1\times 4} = \frac{3}{13} \approx 0.231 \end{aligned}$$

$$\begin{aligned} P(\text{`cheap'} \mid \text{Ham}) &= \frac{0+1}{6+1\times 4} = \frac{1}{10} = 0.1 \\ P(\text{`viagra'} \mid \text{Ham}) &= \frac{0+1}{6+1\times 4} = \frac{1}{10} = 0.1 \end{aligned}$$

$$\begin{aligned} \mu_{jk} &= \frac{1}{N_{C_k}} \sum_{i=1}^{N_{C_k}} x_{ij} \\ \sigma_{jk}^2 &= \frac{1}{N_{C_k}} \sum_{i=1}^{N_{C_k}} (x_{ij} - \mu_{jk})^2 \end{aligned}$$

The Perceptron Algorithm

$$\begin{aligned} &\text{From } (0,0,-1): && w_0 < 0 \\ &\text{From } (0,1,-1): && w_0 + w_2 < 0 \\ &\text{From } (1,0,-1): && w_0 + w_1 < 0 \\ &\text{From } (1,1,+1): && w_0 + w_1 + w_2 > 0 \end{aligned}$$

$$\begin{aligned} &\text{From } (0,0,-1): && w_0 < 0 \\ &\text{From } (0,1,+1): && w_0 + w_2 > 0 \\ &\text{From } (1,0,+1): && w_0 + w_1 > 0 \\ &\text{From } (1,1,-1): && w_0 + w_1 + w_2 < 0 \end{aligned}$$

Ensemble Methods and Loss Functions

$$H(x) = \text{mode}(\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T\})$$

$$\epsilon_t = \sum_{i=1}^{N} w_i^{(t)} \mathbb{I}(h_t(x_i) \neq y_i)$$

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$$

$$\tilde{w}_i^{(t+1)} = w_i^{(t)} \exp(-\alpha_t y_i h_t(x_i))$$

$$w_i^{(t+1)} = \frac{\tilde{w}_i^{(t+1)}}{Z_t}$$

$$H(x) = \text{sign} \left(\sum_{t=1}^{T} \alpha_t h_t(x) \right)$$

$$\epsilon_1 = \sum_{h_1(x_i) \neq y_i} w_i^{(1)} = w_2^{(1)} = \frac{1}{3}$$

$$\alpha_1 = \frac{1}{2} \ln \left(\frac{1 - 1/3}{1/3} \right) = \frac{1}{2} \ln(2) \approx 0.346$$

$$L_{\text{hinge}}(y, f(x)) = \max(0, 1 - yf(x))$$

$$L_{\text{log}}(y, f(x)) = \ln(1 + \exp(-yf(x)))$$

$$L_{\text{exp}}(y, f(x)) = \exp(-yf(x))$$

$$L_{\text{squared}}(y, f(x)) = (y - f(x))^2$$

$$L_{0-1} = \mathbb{I}(yf(x) \leq 0) = \mathbb{I}(0.5 \leq 0) = 0$$

$$L_{\text{hinge}} = \max(0, 1 - yf(x)) = \max(0, 1 - 0.5) = 0.5$$

$$L_{\text{log}} = \ln(1 + \exp(-yf(x))) = \ln(1 + \exp(-0.5)) \approx \ln(1 + 0.6065) = \ln(1.6065) \approx 0.474$$

$$L_{\text{exp}} = \exp(-yf(x)) = \exp(-0.5) \approx 0.6065$$

CHAPTER 41

Managerial Economics Jan 2024

Cost Concepts and Marginal Analysis

$$\begin{aligned} TR(Q) &= P(Q) \times Q = (100 - 2Q)Q = 100Q - 2Q^2 \\ TC(Q) &= 10 + 4Q + Q^2 \end{aligned}$$

$$MR(Q) = \frac{d}{dQ}(100Q - 2Q^2) = 100 - 4Q$$

$$MC(Q) = \frac{d}{dQ}(10 + 4Q + Q^2) = 4 + 2Q$$

$$\begin{aligned} 100 - 4Q &= 4 + 2Q \\ 96 &= 6Q \\ Q^* &= 16 \end{aligned}$$

$$\begin{aligned} \frac{d(MR)}{dQ} &= -4 \\ \frac{d(MC)}{dQ} &= 2 \end{aligned}$$

$$\begin{aligned} P^* &= 100 - 2(16) = 100 - 32 = 68 \\ TR(16) &= 68 \times 16 = 1088 \\ TC(16) &= 10 + 4(16) + (16)^2 = 10 + 64 + 256 = 330 \\ \pi &= TR(16) - TC(16) = 1088 - 330 = 758 \end{aligned}$$

Cost Minimization

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial L} &= w - \lambda \frac{\partial f}{\partial L} = 0 \\ \frac{\partial \mathcal{L}}{\partial K} &= r - \lambda \frac{\partial f}{\partial K} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= q - f(L, K) = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial L} &= 2 - \lambda \cdot 2(L^{0.5} + K^{0.5}) \cdot 0.5L^{-0.5} = 0 \\ \frac{\partial \mathcal{L}}{\partial K} &= 8 - \lambda \cdot 2(L^{0.5} + K^{0.5}) \cdot 0.5K^{-0.5} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 100 - (L^{0.5} + K^{0.5})^2 = 0 \end{aligned}$$

Consumer Optimization and Demand Functions

$$\frac{\partial \mathcal{L}}{\partial X} = MU_X - \lambda P_X = 0$$
$$\frac{\partial \mathcal{L}}{\partial Y} = MU_Y - \lambda P_Y = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = I - P_X X - P_Y Y = 0$$

$$\frac{\partial \mathcal{L}}{\partial X} = X^{-0.5} - \lambda(1) = 0$$
$$\frac{\partial \mathcal{L}}{\partial Y} = 1 - \lambda(2) = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 10 - X - 2Y = 0$$

$$MU_X = \alpha X^{\alpha-1} Y^{\beta}$$
$$MU_Y = \beta X^{\alpha} Y^{\beta-1}$$

Market Equilibrium

$$Q_d = 200 - 4P$$
$$Q_s = 20 + 2P$$

$$200 - 20 = 2P + 4P$$
$$180 = 6P$$
$$P^* = 30$$

$$P_d = 100 - 0.5Q$$
$$P_s = 10 + Q$$

$$100 - 10 = Q + 0.5Q$$
$$90 = 1.5Q$$
$$Q^* = \frac{90}{1.5} = 60$$

$$Q_d = 500 - 10P$$
$$Q_s = 50 + 5P$$

$$500 - 10P = 50 + 5P$$
$$450 = 15P \implies P_1^* = 30$$
$$Q_1^* = 50 + 5(30) = 200$$

$$650 - 10P = 50 + 5P$$
$$600 = 15P \implies P_2^* = 40$$

$$\begin{aligned}\Delta P^* &= 40 - 30 = +10 \\ \Delta Q^* &= 250 - 200 = +50\end{aligned}$$

$$\begin{aligned}Q_d &= 120 - P \\ Q_s &= 20 + P\end{aligned}$$

$$\begin{aligned}160 - 40 &= 2P \\ 120 &= 2P \implies P_2^* = 60\end{aligned}$$

$$\begin{aligned}\Delta P^* &= 60 - 50 = +10 \\ \Delta Q^* &= 100 - 70 = +30\end{aligned}$$

$$\begin{aligned}Q_d &= 300 - 3P^2 \\ Q_s &= 108 + P^2\end{aligned}$$

$$\begin{aligned}300 - 108 &= P^2 + 3P^2 \\ 192 &= 4P^2 \\ P^2 &= 48\end{aligned}$$

Market Interventions and Welfare Analysis

$$\begin{aligned}100 - 2P &= 3P \\ 5P &= 100 \\ P^* &= 20\end{aligned}$$

$$\begin{aligned}&Q_d = 100 - 2(15) = 100 - 30 = 70 \\ &Q_s = 3(15) = 45\end{aligned}$$

$$\begin{aligned}P_b - P_s &= 15 \\ (120 - Q) - (2Q) &= 15\end{aligned}$$

$$\begin{aligned}120 - 3Q &= 15 \\ 3Q &= 105 \\ Q_t &= 35\end{aligned}$$

$$\begin{aligned}P_b &= 120 - 35 = 85 \\ P_s &= 2(35) = 70\end{aligned}$$

$$\text{Consumer Incidence} = \frac{P_b - P^*}{t}$$
$$\text{Producer Incidence} = \frac{P^* - P_s}{t}$$

$$\begin{aligned} 120 - Q &= 2Q \\ 3Q &= 120 \\ Q^* &= 40 \\ P^* &= 120 - 40 = 80 \end{aligned}$$

$$\text{Consumer Burden} = P_b - P^* = 85 - 80 = 5$$
$$\text{Producer Burden} = P^* - P_s = 80 - 70 = 10$$

$$\text{Consumer Incidence} = \frac{5}{15} = \frac{1}{3} \approx 33.3\%$$
$$\text{Producer Incidence} = \frac{10}{15} = \frac{2}{3} \approx 66.7\%$$

$$\begin{aligned} P_{\max} &= 120 \text{ (from } P = 120 - Q) \\ P_{\min} &= 0 \text{ (from } P = 2Q) \\ CS_{\text{initial}} &= \frac{1}{2} \times (120 - 80) \times 40 = 800 \\ PS_{\text{initial}} &= \frac{1}{2} \times (80 - 0) \times 40 = 1600 \\ TW_{\text{initial}} &= 800 + 1600 = 2400 \end{aligned}$$

$$\begin{aligned} CS_{\text{new}} &= \frac{1}{2} \times (120 - 85) \times 35 = 612.5 \\ PS_{\text{new}} &= \frac{1}{2} \times (70 - 0) \times 35 = 1225 \\ \text{Tax Revenue} &= 15 \times 35 = 525 \\ TW_{\text{new}} &= 612.5 + 1225 + 525 = 2362.5 \end{aligned}$$

Price Discrimination Strategies

$$\begin{aligned} P_1 &= 100 - Q_1 \\ P_2 &= 300 - 2Q_2 \end{aligned}$$

$$\begin{aligned} MR_1 &= 100 - 2Q_1 \\ MR_2 &= 300 - 4Q_2 \end{aligned}$$

Game Theory Fundamentals

$$\begin{aligned} E_2(H) &= (-1)(p) + (1)(1 - p) = 1 - 2p \\ E_2(T) &= (1)(p) + (-1)(1 - p) = 2p - 1 \end{aligned}$$

$$
\begin{aligned}
1 - 2p &= 2p - 1 \\
2 &= 4p \\
p &= 0.5
\end{aligned}
$$

$$
\begin{aligned}
E_1(H) &= (1)(q) + (-1)(1-q) = 2q - 1 \\
E_1(T) &= (-1)(q) + (1)(1-q) = 1 - 2q
\end{aligned}
$$

$$
\begin{aligned}
2q - 1 &= 1 - 2q \\
4q &= 2 \\
q &= 0.5
\end{aligned}
$$

CHAPTER 42

Market Research

Advertising Pre-Testing

$$
\begin{aligned}
Z &= \frac{0.045 - 0.070}{\sqrt{0.0575(1 - 0.0575)\left(\frac{1}{1000} + \frac{1}{1000}\right)}} \\
&= \frac{-0.025}{\sqrt{0.0575(0.9425)(0.002)}} \\
&= \frac{-0.025}{\sqrt{0.05419375 \times 0.002}} \\
&= \frac{-0.025}{\sqrt{0.0001083875}} \\
&\approx \frac{-0.025}{0.01041} \approx -2.40
\end{aligned}
$$

CHAPTER 43

Mathematical Foundations Of Generative Ai

Divergence Minimization and Probability Foundations

$$P(0) = 0.4, \quad P(1) = 0.6$$
$$Q_\theta(0) = \theta, \quad Q_\theta(1) = 1 - \theta$$

$$\begin{aligned} f(u) - uf'(u) &= (u-1)^2 - u[2(u-1)] \\ &= u^2 - 2u + 1 - 2u^2 + 2u \\ &= 1 - u^2 \end{aligned}$$

Generative Adversarial Networks (GAN) Fundamentals

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

$$V(G, D) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{x \sim p_g}[\log(1 - D(x))]$$

$$\begin{aligned} \mathbb{E}_{x \sim p_{data}}[\log D(x)] &= p_{data}(1) \log D(1) + p_{data}(2) \log D(2) + p_{data}(3) \log D(3) \\ &= 0.5 \log(0.8) + 0.3 \log(0.5) + 0.2 \log(0.2) \\ &\approx 0.5(-0.223) + 0.3(-0.693) + 0.2(-1.609) \\ &\approx -0.1115 - 0.2079 - 0.3218 = -0.6412 \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{x \sim p_g}[\log(1 - D(x))] &= p_g(1) \log(1 - D(1)) + p_g(2) \log(1 - D(2)) + p_g(3) \log(1 - D(3)) \\ &= 0.1 \log(0.2) + 0.4 \log(0.5) + 0.5 \log(0.8) \\ &\approx 0.1(-1.609) + 0.4(-0.693) + 0.5(-0.223) \\ &\approx -0.1609 - 0.2772 - 0.1115 = -0.5496 \end{aligned}$$

$$V(G, D) = -0.6412 - 0.5496 = -1.1908$$

$$V(G, D) = \int (p_{data}(x) \log D(x) + p_g(x) \log(1 - D(x)))\, dx$$

$$f(y) = A \log(y) + B \log(1 - y)$$

$$\begin{aligned}\frac{df}{dy} &= \frac{A}{y} - \frac{B}{1-y} = 0\\ \frac{A}{y} &= \frac{B}{1-y}\\ A(1-y) &= By\\ A - Ay &= By\\ y(A+B) &= A\\ y &= \frac{A}{A+B}\end{aligned}$$

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

$$p_{data}(x) = \begin{cases} 0.5 & \text{for } x \in [0,2] \\ 0 & \text{otherwise} \end{cases}$$

$$p_g(x) = \begin{cases} 0.5 & \text{for } x \in [1,3] \\ 0 & \text{otherwise} \end{cases}$$

$$D^*(x) = \frac{0.5}{0.5+0} = 1$$

$$D^*(x) = \frac{0.5}{0.5+0.5} = \frac{1}{2}$$

$$D^*(x) = \frac{0}{0+0.5} = 0$$

$$V(G, D^*) = \mathbb{E}_{x\sim p_{data}}\left[\log\left(\frac{p_{data}(x)}{p_{data}(x)+p_g(x)}\right)\right] + \mathbb{E}_{x\sim p_g}\left[\log\left(\frac{p_g(x)}{p_{data}(x)+p_g(x)}\right)\right]$$

$$\begin{aligned}V(G, D^*) &= \mathbb{E}_{x\sim p_{data}}\left[\log\left(\frac{p_{data}(x)/2}{(p_{data}(x)+p_g(x))/2}\right)\right] + \mathbb{E}_{x\sim p_g}\left[\log\left(\frac{p_g(x)/2}{(p_{data}(x)+p_g(x))/2}\right)\right]\\ &= \mathbb{E}_{x\sim p_{data}}\left[\log\left(\frac{1}{2}\cdot\frac{p_{data}(x)}{p_m(x)}\right)\right] + \mathbb{E}_{x\sim p_g}\left[\log\left(\frac{1}{2}\cdot\frac{p_g(x)}{p_m(x)}\right)\right]\end{aligned}$$

$$\begin{aligned}V(G, D^*) = \mathbb{E}_{x\sim p_{data}}[\log(1/2)] &+ \mathbb{E}_{x\sim p_{data}}\left[\log\left(\frac{p_{data}(x)}{p_m(x)}\right)\right]\\ &+ \mathbb{E}_{x\sim p_g}[\log(1/2)] + \mathbb{E}_{x\sim p_g}\left[\log\left(\frac{p_g(x)}{p_m(x)}\right)\right]\\ &= -2\log(2) + KL(p_{data} \parallel p_m) + KL(p_g \parallel p_m)\end{aligned}$$

$$JSD(P \parallel Q) = \frac{1}{2}KL\left(P \parallel \frac{P+Q}{2}\right) + \frac{1}{2}KL\left(Q \parallel \frac{P+Q}{2}\right)$$

$$V(G, D^*) = -2\log(2) + 2\cdot JSD(p_{data} \parallel p_g)$$

$$\begin{aligned}D^*(1) &= \frac{0.8}{0.8+0.4} = \frac{0.8}{1.2} = \frac{2}{3} \approx 0.667\\ D^*(2) &= \frac{0.2}{0.2+0.6} = \frac{0.2}{0.8} = \frac{1}{4} = 0.250\end{aligned}$$

$$\begin{aligned}V(G, D^*) &= p_{data}(1)\log D^*(1) + p_{data}(2)\log D^*(2) + p_g(1)\log(1-D^*(1)) + p_g(2)\log(1-D^*(2))\\ &= 0.8\log(2/3) + 0.2\log(1/4) + 0.4\log(1/3) + 0.6\log(3/4)\\ &\approx 0.8(-0.405) + 0.2(-1.386) + 0.4(-1.099) + 0.6(-0.288)\\ &\approx -0.324 - 0.277 - 0.440 - 0.173 = -1.214\end{aligned}$$

$$
\begin{aligned}
p_m(1) &= \frac{0.8+0.4}{2} = 0.6 \\
p_m(2) &= \frac{0.2+0.6}{2} = 0.4
\end{aligned}
$$

$$
\begin{aligned}
KL(p_{data} \parallel p_m) &= 0.8 \log\left(\frac{0.8}{0.6}\right) + 0.2 \log\left(\frac{0.2}{0.4}\right) \\
&= 0.8 \log(4/3) + 0.2 \log(1/2) \\
&\approx 0.8(0.288) + 0.2(-0.693) = 0.230 - 0.139 = 0.091 \\
KL(p_g \parallel p_m) &= 0.4 \log\left(\frac{0.4}{0.6}\right) + 0.6 \log\left(\frac{0.6}{0.4}\right) \\
&= 0.4 \log(2/3) + 0.6 \log(3/2) \\
&\approx 0.4(-0.405) + 0.6(0.405) = -0.162 + 0.243 = 0.081
\end{aligned}
$$

$$
\begin{aligned}
JSD(p_{data} \parallel p_g) &= \frac{1}{2}(0.091) + \frac{1}{2}(0.081) = 0.0455 + 0.0405 = 0.086 \\
V(G, D^*) &= -2\log(2) + 2 \cdot JSD(p_{data} \parallel p_g) \\
&\approx -1.386 + 2(0.086) = -1.386 + 0.172 = -1.214
\end{aligned}
$$

Convolutional and Conditional GANs

$$
\begin{aligned}
D(x|y) &= 0.8 \\
D(G(z|y)|y) &= 0.3
\end{aligned}
$$

$$
\begin{aligned}
P(S = \text{real}|X_{\text{fake}}) &= 0.4 \\
P(C = 2|X_{\text{fake}}) &= 0.7
\end{aligned}
$$

Wasserstein GANs and Training Stability

$$
L_C = -\frac{1}{m}\sum_{i=1}^{m} f(x^{(i)}) + \frac{1}{m}\sum_{i=1}^{m} f(\tilde{x}^{(i)})
$$

$$
w \leftarrow w - \alpha \nabla_w L_C
$$

$$
L_G = -\frac{1}{m}\sum_{i=1}^{m} f(G(z^{(i)}))
$$

$$
\theta \leftarrow \theta - \alpha \nabla_\theta L_G
$$

$$
\hat{x}^{(i)} = \epsilon^{(i)} x^{(i)} + (1 - \epsilon^{(i)})\tilde{x}^{(i)}
$$

$$
\nabla_{\hat{x}} f(\hat{x}^{(i)})
$$

$$
\text{GP} = \frac{1}{m}\sum_{i=1}^{m} \Big(||\nabla_{\hat{x}} f(\hat{x}^{(i)})||_2 - 1\Big)^2
$$

$$
L_C = \frac{1}{m}\sum_{i=1}^{m} f(\tilde{x}^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} f(x^{(i)}) + \lambda \text{GP}
$$

$$
\hat{x} = 0.25 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + (1 - 0.25)\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} + \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}
$$

$$\nabla_{\hat{x}} f(-1,1) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$||\nabla_{\hat{x}} f(-1,1)||_2 = \sqrt{(-1)^2 + 1^2} = \sqrt{2} \approx 1.414$$

$$\text{GP} = (||\nabla_{\hat{x}} f||_2 - 1)^2 = (\sqrt{2} - 1)^2 \approx (1.414 - 1)^2 = (0.414)^2 \approx 0.171$$

$$\lambda \text{GP} = 10 \times 0.171 = 1.71$$

GAN Inversion and Generative Evaluation

$$\begin{aligned} D_{KL}(p(y|x_1) \parallel p(y)) &= 0.8 \ln\left(\frac{0.8}{0.5}\right) + 0.1 \ln\left(\frac{0.1}{0.35}\right) + 0.1 \ln\left(\frac{0.1}{0.15}\right) \\ &= 0.8 \ln(1.6) + 0.1 \ln\left(\frac{2}{7}\right) + 0.1 \ln\left(\frac{2}{3}\right) \\ &\approx 0.8(0.4700) + 0.1(-1.2528) + 0.1(-0.4055) \\ &\approx 0.3760 - 0.1253 - 0.0406 = 0.2101 \end{aligned}$$

$$\begin{aligned} D_{KL}(p(y|x_2) \parallel p(y)) &= 0.2 \ln\left(\frac{0.2}{0.5}\right) + 0.6 \ln\left(\frac{0.6}{0.35}\right) + 0.2 \ln\left(\frac{0.2}{0.15}\right) \\ &= 0.2 \ln(0.4) + 0.6 \ln\left(\frac{12}{7}\right) + 0.2 \ln\left(\frac{4}{3}\right) \\ &\approx 0.2(-0.9163) + 0.6(0.5390) + 0.2(0.2877) \\ &\approx -0.1833 + 0.3234 + 0.0575 = 0.1976 \end{aligned}$$

$$\begin{aligned} \nabla_z \mathcal{L}(z_0) &= -2(\nabla_z G(z_0))^T (x - G(z_0)) \\ &= -2 \begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= -2(3 \times 3 + 2 \times 4) \\ &= -2(9 + 8) = -34 \end{aligned}$$

$$\begin{aligned} W^+ = (W^T W)^{-1} W^T &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2(1) - 1(0) & 2(0) - 1(1) & 2(1) - 1(1) \\ -1(1) + 2(0) & -1(0) + 2(1) & -1(1) + 2(1) \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} z^* = W^+ x &= \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2(2) - 1(3) + 1(4) \\ -1(2) + 2(3) + 1(4) \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 4 - 3 + 4 \\ -2 + 6 + 4 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 5/3 \\ 8/3 \end{bmatrix} \end{aligned}$$

Variational Autoencoders (VAEs)

$$\begin{aligned}
D_{KL}^{(1)} &= -\frac{1}{2}\left(1+\ln(\sigma_1^2)-\mu_1^2-\sigma_1^2\right) \\
&= -\frac{1}{2}\left(1+(-0.1)-(0.5)^2-0.9048\right) \\
&= -\frac{1}{2}\left(1-0.1-0.25-0.9048\right) \\
&= -\frac{1}{2}\left(-0.2548\right) = 0.1274
\end{aligned}$$

$$\begin{aligned}
D_{KL}^{(2)} &= -\frac{1}{2}\left(1+\ln(\sigma_2^2)-\mu_2^2-\sigma_2^2\right) \\
&= -\frac{1}{2}\left(1+0.4-(-0.2)^2-1.4918\right) \\
&= -\frac{1}{2}\left(1+0.4-0.04-1.4918\right) \\
&= -\frac{1}{2}\left(-0.1318\right) = 0.0659
\end{aligned}$$

$$\begin{aligned}
&\sigma = \exp\left(\frac{v}{2}\right) = \exp\left(\frac{-0.6}{2}\right) = \exp(-0.3) \approx 0.7408 \\
&z = \mu + \sigma \cdot \epsilon = 1.2 + (0.7408)(0.5) = 1.2 + 0.3704 = 1.5704
\end{aligned}$$

$$\begin{aligned}
L_{KL} &= -\frac{1}{2}\left(1+\ln(\sigma_1^2)-\mu_1^2-\sigma_1^2\right) \\
&= -\frac{1}{2}\left(1-0.5-(0.2)^2-0.6065\right) \\
&= -\frac{1}{2}\left(1-0.5-0.04-0.6065\right) \\
&= -\frac{1}{2}\left(-0.1465\right) = 0.07325
\end{aligned}$$

Advanced VAE Architectures

$$\begin{aligned}
L_{\text{codebook}} &= \|\text{sg}[z_e(x)] - e_k\|^2 \\
&= (2.1-2.0)^2 + (-1.8-(-1.5))^2 \\
&= (0.1)^2 + (-0.3)^2 \\
&= 0.01 + 0.09 = 0.10
\end{aligned}$$

$$\begin{aligned}
L_{\text{commitment}} &= \beta\|z_e(x) - \text{sg}[e_k]\|^2 \\
&= 0.25 \times \left((2.1-2.0)^2 + (-1.8-(-1.5))^2\right) \\
&= 0.25 \times 0.10 \\
&= 0.025
\end{aligned}$$

$$\begin{aligned}
L_{\text{codebook}} &= \frac{1}{2}\left(\|\text{sg}[z_{e,1}] - e_1\|^2 + \|\text{sg}[z_{e,2}] - e_2\|^2\right) \\
&= \frac{1}{2}(0.05+0.05) = 0.05
\end{aligned}$$

$$\begin{aligned}
L_{\text{commitment}} &= \frac{1}{2}\left(\beta\|z_{e,1} - \text{sg}[e_1]\|^2 + \beta\|z_{e,2} - \text{sg}[e_2]\|^2\right) \\
&= \frac{0.5}{2}(0.05+0.05) = 0.25 \times 0.10 = 0.025
\end{aligned}$$

$$
\begin{aligned}
M_1^{(k)} &= \gamma M_1^{(k-1)} + (1-\gamma) m_1 \\
&= 0.9 \begin{bmatrix} 20.0 \\ 10.0 \end{bmatrix} + 0.1 \begin{bmatrix} 4.0 \\ 2.4 \end{bmatrix} \\
&= \begin{bmatrix} 18.0 \\ 9.0 \end{bmatrix} + \begin{bmatrix} 0.4 \\ 0.24 \end{bmatrix} = \begin{bmatrix} 18.4 \\ 9.24 \end{bmatrix}
\end{aligned}
$$

Reinforcement Learning for Generative AI (RLHF)

$$
\begin{aligned}
\log \pi_\theta(y_w \mid x) = -1.5, &\quad \log \pi_{\text{ref}}(y_w \mid x) = -2.0 \\
\log \pi_\theta(y_l \mid x) = -2.5, &\quad \log \pi_{\text{ref}}(y_l \mid x) = -2.0
\end{aligned}
$$

$$
\begin{aligned}
\hat{r}(x, y_w) &= \beta \left(\log \pi_\theta(y_w \mid x) - \log \pi_{\text{ref}}(y_w \mid x) \right) \\
&= 0.5 \times (-1.5 - (-2.0)) \\
&= 0.5 \times (0.5) = 0.25
\end{aligned}
$$

$$
\begin{aligned}
\hat{r}(x, y_l) &= \beta \left(\log \pi_\theta(y_l \mid x) - \log \pi_{\text{ref}}(y_l \mid x) \right) \\
&= 0.5 \times (-2.5 - (-2.0)) \\
&= 0.5 \times (-0.5) = -0.25
\end{aligned}
$$

$$
\begin{aligned}
L_{\text{DPO}}(\theta) &= -\log \sigma(\Delta \hat{r}) \\
&= -\log \left(\frac{1}{1 + e^{-0.50}} \right) \\
&= \log(1 + e^{-0.50}) \\
&\approx \log(1 + 0.6065) \\
&\approx \log(1.6065) \\
&\approx 0.474
\end{aligned}
$$

$$
\begin{aligned}
\log \pi_\theta(y_w \mid x) = -3.0, &\quad \log \pi_{\text{ref}}(y_w \mid x) = -2.0 \\
\log \pi_\theta(y_l \mid x) = -1.0, &\quad \log \pi_{\text{ref}}(y_l \mid x) = -2.0
\end{aligned}
$$

$$
\begin{aligned}
\hat{r}(x, y_w) &= 0.2 \times (-3.0 - (-2.0)) = 0.2 \times (-1.0) = -0.2 \\
\hat{r}(x, y_l) &= 0.2 \times (-1.0 - (-2.0)) = 0.2 \times (1.0) = 0.2
\end{aligned}
$$

$$
\begin{aligned}
L_{\text{DPO}}(\theta) &= -\log \sigma(-0.4) \\
&= \log(1 + e^{0.4}) \\
&\approx \log(1 + 1.4918) \\
&\approx \log(2.4918) \\
&\approx 0.913
\end{aligned}
$$

$$
r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\text{old}}(a_t \mid s_t)}
$$

$$
L_{\text{unclipped}} = r_t(\theta) \hat{A}_t
$$

$$L_{\text{clipped}} = \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t$$

$$L^{\text{CLIP}}(\theta) = \min(L_{\text{unclipped}}, L_{\text{clipped}})$$

$$r_t(\theta) = \frac{0.7}{0.5} = 1.4$$

$$L_{\text{unclipped}} = r_t(\theta)\hat{A}_t = 1.4 \times 2.0 = 2.8$$

$$L_{\text{clipped}} = 1.2 \times 2.0 = 2.4$$

$$L^{\text{CLIP}}(\theta) = \min(2.8, 2.4) = 2.4$$

$$r_t(\theta) = \frac{0.3}{0.8} = 0.375$$

$$L_{\text{unclipped}} = r_t(\theta)\hat{A}_t = 0.375 \times (-1.5) = -0.5625$$

$$L_{\text{clipped}} = 0.8 \times (-1.5) = -1.2$$

$$L^{\text{CLIP}}(\theta) = \min(-0.5625, -1.2) = -1.2$$

$$\log \pi_\theta(y_w \mid x), \quad \log \pi_{\text{ref}}(y_w \mid x)$$

$$\log \pi_\theta(y_l \mid x), \quad \log \pi_{\text{ref}}(y_l \mid x)$$

$$\hat{r}(x, y_w) = \beta \left(\log \pi_\theta(y_w \mid x) - \log \pi_{\text{ref}}(y_w \mid x)\right)$$

$$\hat{r}(x, y_l) = \beta \left(\log \pi_\theta(y_l \mid x) - \log \pi_{\text{ref}}(y_l \mid x)\right)$$

$$\Delta\hat{r} = \hat{r}(x, y_w) - \hat{r}(x, y_l)$$

$$\Delta\hat{r} = \beta \log\left(\frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)}\right) - \beta \log\left(\frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}\right)$$

$$L_{\text{DPO}}(\theta) = -\log \sigma(\Delta\hat{r})$$

$$\Delta\hat{r} = \hat{r}(x, y_w) - \hat{r}(x, y_l) = 0.25 - (-0.25) = 0.50$$

$$\Delta\hat{r} = -0.2 - 0.2 = -0.4$$

CHAPTER 44

Modern Application Development I

No prominent formulae extracted for this course.

CHAPTER 45

Modern Application Development Ii

No prominent formulae extracted for this course.

CHAPTER 46

Operating Systems Prof Chester

Storage Systems and Hard Disks

$$\begin{aligned}
\text{Total Capacity} &= 8 \text{ surfaces} \times 10000 \text{ tracks} \times 500 \text{ sectors} \times 512 \text{ bytes} \\
&= 8 \times 10000 \times 500 \times 512 \\
&= 8 \times 5000000 \times 512 \\
&= 40000000 \times 512 \\
&= 20480000000 \text{ bytes}
\end{aligned}$$

$$\begin{aligned}
\text{Seek Time} &= \text{Startup Time} + (C \times \text{Track Crossing Time}) \\
&= 2 \text{ ms} + (100 \times 0.05 \text{ ms}) \\
&= 2 \text{ ms} + 5 \text{ ms} \\
&= 7 \text{ ms}
\end{aligned}$$

$$\begin{aligned}
\text{Transfer Time} &= \text{Time for one rotation} \times \frac{40}{200} \\
&= \frac{1}{90} \times \frac{1}{5} \\
&= \frac{1}{450} \text{ seconds} \approx 0.00222 \text{ seconds} = 2.22 \text{ ms}
\end{aligned}$$

$$\begin{aligned}
T_{\text{access}} &= T_{\text{seek}} + T_{\text{rotation}} + T_{\text{transfer}} + T_{\text{controller}} \\
&= 5 \text{ ms} + 3 \text{ ms} + 0.12 \text{ ms} + 1 \text{ ms} \\
&= 9.12 \text{ ms}
\end{aligned}$$

Disk Scheduling Algorithms

$$\begin{aligned}
\text{Distance} &= |0 - 53| + |183 - 0| \\
&= 53 + 183 \\
&= 236 \text{ cylinders}
\end{aligned}$$

$$\begin{aligned}
\text{Distance} &= |199 - 53| + |0 - 199| + |37 - 0| \\
&= 146 + 199 + 37 \\
&= 382 \text{ cylinders}
\end{aligned}$$

$$\begin{aligned}
\text{Distance} &= |14 - 53| + |183 - 14| \\
&= 39 + 169 \\
&= 208 \text{ cylinders}
\end{aligned}$$

$$
\begin{aligned}
\text{Distance} &= |183 - 53| + |14 - 183| + |37 - 14| \\
&= 130 + 169 + 23 \\
&= 322 \text{ cylinders}
\end{aligned}
$$

CHAPTER 47

Privacy And Security In Online Social Media

Analyzing Misinformation and Fake News

$$
\begin{aligned}
S(t+1) &= S(t) - \beta S(t)\frac{I(t)}{N} \\
I(t+1) &= I(t) + \beta S(t)\frac{I(t)}{N} - \gamma I(t) \\
R(t+1) &= R(t) + \gamma I(t)
\end{aligned}
$$

$$
\begin{aligned}
S(1) &= 99 - \left(0.5 \times 99 \times \frac{1}{100}\right) = 99 - 0.495 = 98.505 \\
I(1) &= 1 + 0.495 - (0.2 \times 1) = 1.495 - 0.2 = 1.295 \\
R(1) &= 0 + (0.2 \times 1) = 0.2
\end{aligned}
$$

Analyzing Cybercrime on Social Media

$$
\begin{aligned}
\Delta t_1 &= 180 - 120 = 60 \\
\Delta t_2 &= 240 - 180 = 60 \\
\Delta t_3 &= 305 - 240 = 65 \\
\Delta t_4 &= 365 - 305 = 60
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= \frac{(60 - 61.25)^2 + (60 - 61.25)^2 + (65 - 61.25)^2 + (60 - 61.25)^2}{4} \\
&= \frac{(-1.25)^2 + (-1.25)^2 + (3.75)^2 + (-1.25)^2}{4} \\
&= \frac{1.5625 + 1.5625 + 14.0625 + 1.5625}{4} \\
&= \frac{18.75}{4} = 4.6875
\end{aligned}
$$

CHAPTER 48

Programming Concepts Using Java

No prominent formulae extracted for this course.

CHAPTER 49

Programming Data Structures And Algorithms Using Pythonprof Madhavan Mukund

Divide and Conquer Applications

$$\begin{aligned} \text{Total Inversions} &= \text{Inv}(L) + \text{Inv}(R) + \text{Split Inv} \\ &= 1 + 1 + 4 = 6 \end{aligned}$$

Graph Representations and Traversal

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Advanced Dynamic Programming

$$\begin{aligned} M[1][2] &= 0 + 0 + (10 \times 20 \times 30) = 6000, & S[1][2] &= 1 \\ M[2][3] &= 0 + 0 + (20 \times 30 \times 40) = 24000, & S[2][3] &= 2 \\ M[3][4] &= 0 + 0 + (30 \times 40 \times 30) = 36000, & S[3][4] &= 3 \end{aligned}$$

CHAPTER 50

Programming In Python

No prominent formulae extracted for this course.

CHAPTER 51

Software Engineering

Project Size and Effort Estimation

$$\begin{aligned}
\text{EI} &= 50 \times 4 = 200 \\
\text{EO} &= 40 \times 5 = 200 \\
\text{EQ} &= 35 \times 4 = 140 \\
\text{ILF} &= 6 \times 10 = 60 \\
\text{EIF} &= 4 \times 7 = 28
\end{aligned}$$

$$\begin{aligned}
&E = 2.4 \times (32)^{1.05} \\
&E \approx 2.4 \times 38.38 \\
&E \approx 92.12 \text{ Person-Months (PM)}
\end{aligned}$$

$$\begin{aligned}
&D = 2.5 \times (92.12)^{0.38} \\
&D \approx 2.5 \times 5.48 \\
&D \approx 13.7 \text{ Months}
\end{aligned}$$

Agile Project Tracking

$$\begin{aligned}
&W_{ideal}(4) = 80 - \left(\frac{80}{10}\right) \times 4 \\
&W_{ideal}(4) = 80 - (8 \times 4) \\
&W_{ideal}(4) = 80 - 32 = 48 \text{ hours}
\end{aligned}$$

CHAPTER 52

Software Testing

No prominent formulae extracted for this course.

CHAPTER 53

Special Topics In Ml Reinforcement Learning

Multi-Armed Bandits and Exploration Strategies

$$\begin{aligned}\mathrm{UCB}_{11}(1) &= 7.5 + 2\sqrt{\frac{2.398}{4}}\\ &= 7.5 + 2\sqrt{0.5995}\\ &\approx 7.5 + 2(0.774)\\ &= 7.5 + 1.548 = 9.048\end{aligned}$$

$$\begin{aligned}\mathrm{UCB}_{11}(2) &= 8.0 + 2\sqrt{\frac{2.398}{5}}\\ &= 8.0 + 2\sqrt{0.4796}\\ &\approx 8.0 + 2(0.692)\\ &= 8.0 + 1.384 = 9.384\end{aligned}$$

$$\begin{aligned}\mathrm{UCB}_{11}(3) &= 5.0 + 2\sqrt{\frac{2.398}{1}}\\ &= 5.0 + 2\sqrt{2.398}\\ &\approx 5.0 + 2(1.548)\\ &= 5.0 + 3.096 = 8.096\end{aligned}$$

Markov Decision Processes

$$P = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.3 & 0.0 & 0.7 \\ 0.0 & 1.0 & 0.0 \end{bmatrix}$$

$$v_n = v_0 P^n$$

$$v_k = v_{k-1} P \quad \text{for } k = 1 \ldots n$$

$$P = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.3 & 0.0 & 0.7 \\ 0.0 & 1.0 & 0.0 \end{bmatrix}$$

$$v_0 = \begin{bmatrix} 0.0 & 0.0 & 1.0 \end{bmatrix}$$

$$v_1 = v_0 P = \begin{bmatrix} 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.3 & 0.0 & 0.7 \\ 0.0 & 1.0 & 0.0 \end{bmatrix} = \begin{bmatrix} 0.0 & 1.0 & 0.0 \end{bmatrix}$$

$$v_2 = v_1 P = \begin{bmatrix} 0.0 & 1.0 & 0.0 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.3 & 0.0 & 0.7 \\ 0.0 & 1.0 & 0.0 \end{bmatrix} = \begin{bmatrix} 0.3 & 0.0 & 0.7 \end{bmatrix}$$

$$P^\pi = \begin{bmatrix} 0.0 & 1.0 \\ 0.5 & 0.5 \end{bmatrix}, \quad R^\pi = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$V^\pi = R^\pi + \gamma P^\pi V^\pi$$

$$V^\pi = (I - \gamma P^\pi)^{-1} R^\pi$$

$$P^\pi = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}, \quad R^\pi = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\gamma P^\pi = 0.5 \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.3 \end{bmatrix}$$

$$(I - \gamma P^\pi) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.1 \\ -0.2 & 0.7 \end{bmatrix}$$

$$(I - \gamma P^\pi)^{-1} = \frac{1}{0.4} \begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.6 \end{bmatrix} = \begin{bmatrix} 1.75 & 0.25 \\ 0.50 & 1.50 \end{bmatrix}$$

$$V^\pi = \begin{bmatrix} 1.75 & 0.25 \\ 0.50 & 1.50 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \end{bmatrix} = \begin{bmatrix} (1.75)(5) + (0.25)(-1) \\ (0.50)(5) + (1.50)(-1) \end{bmatrix} = \begin{bmatrix} 8.75 - 0.25 \\ 2.50 - 1.50 \end{bmatrix} = \begin{bmatrix} 8.5 \\ 1.0 \end{bmatrix}$$

Temporal Difference (TD) Learning Prediction

$$\begin{aligned} \text{Target} &= R_{t+1} + \gamma V(S_{t+1}) \\ &= 10 + 0.9 \times 6.0 \\ &= 10 + 5.4 = 15.4 \end{aligned}$$

$$\begin{aligned} \delta_t &= \text{Target} - V(S_t) \\ &= 15.4 - 4.0 = 11.4 \end{aligned}$$

$$\begin{aligned} V(A) &\leftarrow V(A) + \alpha \delta_t \\ &= 4.0 + 0.1 \times 11.4 \\ &= 4.0 + 1.14 = 5.14 \end{aligned}$$

$$\begin{aligned} \delta &= R_{t+1} + \gamma V(S_2) - V(S_1) = 1 + 0.5 \times 0 - 0 = 1 \\ V(S_1) &\leftarrow V(S_1) + \alpha \delta = 0 + 1.0 \times 1 = 1 \end{aligned}$$

$$\begin{aligned} \delta &= R_{t+1} + \gamma V(T) - V(S_2) = 2 + 0.5 \times 0 - 0 = 2 \\ V(S_2) &\leftarrow V(S_2) + \alpha \delta = 0 + 1.0 \times 2 = 2 \end{aligned}$$

$$\begin{aligned} \delta &= R_{t+1} + \gamma V(T) - V(S_2) \\ &= 4 + 0.5 \times 0 - 2 = 2 \\ V(S_2) &\leftarrow V(S_2) + \alpha \delta = 2 + 1.0 \times 2 = 4 \end{aligned}$$

Off-Policy TD Control: Q-Learning

$$
\begin{aligned}
Q(S_1, \text{Right}) &\leftarrow Q(S_1, \text{Right}) + \alpha \left[R + \gamma \max_{a'} Q(S_2, a') - Q(S_1, \text{Right}) \right] \\
&\leftarrow 2.1 + 0.1 \left[-1 + 0.9(3.0) - 2.1 \right] \\
&\leftarrow 2.1 + 0.1 \left[-1 + 2.7 - 2.1 \right] \\
&\leftarrow 2.1 + 0.1 \left[-0.4 \right] \\
&\leftarrow 2.1 - 0.04 \\
&\leftarrow 2.06
\end{aligned}
$$

$$
\begin{aligned}
Q(X, a_1) &\leftarrow Q(X, a_1) + \alpha \left[R + \gamma \max_{a} Q(Y, a) - Q(X, a_1) \right] \\
&\leftarrow 0 + 0.5 \left[10 + 0.8(0) - 0 \right] \\
&\leftarrow 0.5[10] = 5
\end{aligned}
$$

$$
\begin{aligned}
Q(Y, a_2) &\leftarrow Q(Y, a_2) + \alpha \left[R + \gamma \max_{a} Q(Z, a) - Q(Y, a_2) \right] \\
&\leftarrow 0 + 0.5 \left[-5 + 0.8(0) - 0 \right] \\
&\leftarrow 0.5[-5] = -2.5
\end{aligned}
$$

CHAPTER 54

Speech Technologyprof Umesh

Gaussian Mixture Models (GMM)

$$\begin{aligned}
\pi_k^{new} &= \frac{N_k}{N} \\
\boldsymbol{\mu}_k^{new} &= \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{nk}\mathbf{x}_n \\
\boldsymbol{\Sigma}_k^{new} &= \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{nk}(\mathbf{x}_n-\boldsymbol{\mu}_k^{new})(\mathbf{x}_n-\boldsymbol{\mu}_k^{new})^T
\end{aligned}$$

$$\begin{aligned}
\mathcal{N}(x|\mu_1,\sigma_1^2) &= \frac{1}{\sqrt{2\pi\cdot 1}}\exp\left(-\frac{(1-0)^2}{2\cdot 1}\right) \\
&= \frac{1}{\sqrt{2\pi}}e^{-0.5} \approx 0.3989\times 0.6065 \approx 0.2420 \\
\mathcal{N}(x|\mu_2,\sigma_2^2) &= \frac{1}{\sqrt{2\pi\cdot 0.5}}\exp\left(-\frac{(1-3)^2}{2\cdot 0.5}\right) \\
&= \frac{1}{\sqrt{\pi}}e^{-4} \approx 0.5642\times 0.0183 \approx 0.0103
\end{aligned}$$

$$\begin{aligned}
w_1 &= \pi_1\mathcal{N}(1|\mu_1,\sigma_1^2) = 0.4\times 0.2420 = 0.0968 \\
w_2 &= \pi_2\mathcal{N}(1|\mu_2,\sigma_2^2) = 0.6\times 0.0103 = 0.0062
\end{aligned}$$

$$\begin{aligned}
\gamma_1 &= \frac{w_1}{p(x=1)} = \frac{0.0968}{0.1030} \approx 0.940 \\
\gamma_2 &= \frac{w_2}{p(x=1)} = \frac{0.0062}{0.1030} \approx 0.060
\end{aligned}$$

$$\begin{aligned}
w_{11} &= 0.5\times\frac{1}{\sqrt{2\pi}}e^{-0} \approx 0.5\times 0.3989 = 0.19945 \\
w_{12} &= 0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(1-3)^2}{2}} \approx 0.5\times 0.0540 = 0.0270 \\
p(x_1) &= 0.19945 + 0.0270 = 0.22645 \\
\gamma_{11} &= \frac{0.19945}{0.22645} \approx 0.881, \quad \gamma_{12} = \frac{0.0270}{0.22645} \approx 0.119
\end{aligned}$$

$$\begin{aligned}
w_{21} &= 0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(2-1)^2}{2}} \approx 0.5\times 0.2420 = 0.1210 \\
w_{22} &= 0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(2-3)^2}{2}} \approx 0.5\times 0.2420 = 0.1210 \\
p(x_2) &= 0.1210 + 0.1210 = 0.2420 \\
\gamma_{21} &= \frac{0.1210}{0.2420} = 0.500, \quad \gamma_{22} = \frac{0.1210}{0.2420} = 0.500
\end{aligned}$$

$$
\begin{aligned}
w_{31} &= 0.5 \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(4-1)^2}{2}} \approx 0.5 \times 0.0044 = 0.0022 \\
w_{32} &= 0.5 \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(4-3)^2}{2}} \approx 0.5 \times 0.2420 = 0.1210 \\
p(x_3) &= 0.0022 + 0.1210 = 0.1232 \\
\gamma_{31} &= \frac{0.0022}{0.1232} \approx 0.018, \quad \gamma_{32} = \frac{0.1210}{0.1232} \approx 0.982
\end{aligned}
$$

$$
\begin{aligned}
N_1 &= 0.881 + 0.500 + 0.018 = 1.399 \\
N_2 &= 0.119 + 0.500 + 0.982 = 1.601
\end{aligned}
$$

$$
\begin{aligned}
\pi_1^{new} &= \frac{1.399}{3} \approx 0.466 \\
\pi_2^{new} &= \frac{1.601}{3} \approx 0.534
\end{aligned}
$$

$$
\begin{aligned}
\mu_1^{new} &= \frac{1}{1.399}(0.881 \times 1 + 0.500 \times 2 + 0.018 \times 4) = \frac{1.953}{1.399} \approx 1.396 \\
\mu_2^{new} &= \frac{1}{1.601}(0.119 \times 1 + 0.500 \times 2 + 0.982 \times 4) = \frac{5.047}{1.601} \approx 3.152
\end{aligned}
$$

$$
\begin{aligned}
(\sigma_1^2)^{new} &= \frac{1}{1.399}\left(0.881(1-1.396)^2 + 0.500(2-1.396)^2 + 0.018(4-1.396)^2\right) \\
&= \frac{1}{1.399}\left(0.881(0.1568) + 0.500(0.3648) + 0.018(6.7808)\right) \\
&= \frac{1}{1.399}\left(0.1381 + 0.1824 + 0.1221\right) = \frac{0.4426}{1.399} \approx 0.316
\end{aligned}
$$

$$
\begin{aligned}
(\sigma_2^2)^{new} &= \frac{1}{1.601}\left(0.119(1-3.152)^2 + 0.500(2-3.152)^2 + 0.982(4-3.152)^2\right) \\
&= \frac{1}{1.601}\left(0.119(4.6311) + 0.500(1.3271) + 0.982(0.7191)\right) \\
&= \frac{1}{1.601}\left(0.5511 + 0.6636 + 0.7062\right) = \frac{1.9209}{1.601} \approx 1.200
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) &= \left(\frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1-0)^2}{2\cdot 1}}\right) \times \left(\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{(2-0)^2}{2\cdot 2}}\right) \\
&= \frac{1}{2\pi\sqrt{2}} \exp\left(-0.5 - 1\right) = \frac{1}{2\pi\sqrt{2}} e^{-1.5} \\
&\approx 0.1125 \times 0.2231 \approx 0.0251
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) &= \left(\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{(1-2)^2}{2\cdot 2}}\right) \times \left(\frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(2-2)^2}{2\cdot 1}}\right) \\
&= \frac{1}{2\pi\sqrt{2}} \exp\left(-0.25 - 0\right) = \frac{1}{2\pi\sqrt{2}} e^{-0.25} \\
&\approx 0.1125 \times 0.7788 \approx 0.0876
\end{aligned}
$$

$$
\begin{aligned}
w_1 &= \pi_1 \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = 0.3 \times 0.0251 = 0.00753 \\
w_2 &= \pi_2 \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = 0.7 \times 0.0876 = 0.06132
\end{aligned}
$$

$$
\begin{aligned}
\gamma_1 &= \frac{0.00753}{0.06885} \approx 0.109 \\
\gamma_2 &= \frac{0.06132}{0.06885} \approx 0.891
\end{aligned}
$$

Markov Chains and Hidden Markov Models (HMM) Basics

$$
\begin{aligned}
P(Q) &= \pi_1 \cdot a_{11} \cdot a_{12} \cdot a_{21} \\
&= 0.6 \cdot 0.8 \cdot 0.2 \cdot 0.3 \\
&= 0.0288
\end{aligned}
$$

$$
\begin{aligned}
P(O, Q \mid \lambda) &= \pi_1 \cdot b_1(v_1) \cdot a_{11} \cdot b_1(v_2) \cdot a_{12} \cdot b_2(v_2) \\
&= 0.5 \cdot 0.8 \cdot 0.7 \cdot 0.2 \cdot 0.3 \cdot 0.7 \\
&= 0.4 \cdot 0.14 \cdot 0.21 \\
&= 0.056 \cdot 0.21 \\
&= 0.01176
\end{aligned}
$$

HMM Algorithms: Forward and Viterbi

$$
\begin{aligned}
\alpha_1(H) &= \pi_H b_H(1) = 0.8 \times 0.7 = 0.56 \\
\alpha_1(C) &= \pi_C b_C(1) = 0.2 \times 0.1 = 0.02
\end{aligned}
$$

$$
\begin{aligned}
\alpha_2(H) &= (\alpha_1(H) a_{HH} + \alpha_1(C) a_{CH}) b_H(2) \\
&= (0.56 \times 0.2 + 0.02 \times 0.6) \times 0.2 \\
&= (0.112 + 0.012) \times 0.2 = 0.124 \times 0.2 = 0.0248
\end{aligned}
$$

$$
\begin{aligned}
\alpha_2(C) &= (\alpha_1(H) a_{HC} + \alpha_1(C) a_{CC}) b_C(2) \\
&= (0.56 \times 0.8 + 0.02 \times 0.4) \times 0.4 \\
&= (0.448 + 0.008) \times 0.4 = 0.456 \times 0.4 = 0.1824
\end{aligned}
$$

$$
\begin{aligned}
\alpha_3(H) &= (\alpha_2(H) a_{HH} + \alpha_2(C) a_{CH}) b_H(3) \\
&= (0.0248 \times 0.2 + 0.1824 \times 0.6) \times 0.1 \\
&= (0.00496 + 0.10944) \times 0.1 = 0.1144 \times 0.1 = 0.01144
\end{aligned}
$$

$$
\begin{aligned}
\alpha_3(C) &= (\alpha_2(H) a_{HC} + \alpha_2(C) a_{CC}) b_C(3) \\
&= (0.0248 \times 0.8 + 0.1824 \times 0.4) \times 0.5 \\
&= (0.01984 + 0.07296) \times 0.5 = 0.0928 \times 0.5 = 0.0464
\end{aligned}
$$

$$
\begin{aligned}
v_1(j) &= \pi_j b_j(o_1) \\
bp_1(j) &= 0
\end{aligned}
$$

$$
\begin{aligned}
v_t(j) &= \max_{1 \le i \le N} (v_{t-1}(i) a_{ij}) \cdot b_j(o_t) \\
bp_t(j) &= \arg\max_{1 \le i \le N} (v_{t-1}(i) a_{ij})
\end{aligned}
$$

$$
\begin{aligned}
P^* &= \max_{1 \le j \le N} v_T(j) \\
q_T^* &= \arg\max_{1 \le j \le N} v_T(j)
\end{aligned}
$$

$$v_1(H) = \pi_H b_H(1) = 0.8 \times 0.7 = 0.56, \quad bp_1(H) = 0$$
$$v_1(C) = \pi_C b_C(1) = 0.2 \times 0.1 = 0.02, \quad bp_1(C) = 0$$

$$\begin{aligned}
v_2(H) &= \max(v_1(H)a_{HH}, v_1(C)a_{CH}) \cdot b_H(2) \\
&= \max(0.56 \times 0.2, 0.02 \times 0.6) \times 0.2 \\
&= \max(0.112, 0.012) \times 0.2 = 0.112 \times 0.2 = 0.0224 \\
bp_2(H) &= H \quad \text{(since 0.112 came from } H\text{)}
\end{aligned}$$

$$\begin{aligned}
v_2(C) &= \max(v_1(H)a_{HC}, v_1(C)a_{CC}) \cdot b_C(2) \\
&= \max(0.56 \times 0.8, 0.02 \times 0.4) \times 0.4 \\
&= \max(0.448, 0.008) \times 0.4 = 0.448 \times 0.4 = 0.1792 \\
bp_2(C) &= H \quad \text{(since 0.448 came from } H\text{)}
\end{aligned}$$

$$\begin{aligned}
v_3(H) &= \max(v_2(H)a_{HH}, v_2(C)a_{CH}) \cdot b_H(3) \\
&= \max(0.0224 \times 0.2, 0.1792 \times 0.6) \times 0.1 \\
&= \max(0.00448, 0.10752) \times 0.1 = 0.10752 \times 0.1 = 0.010752 \\
bp_3(H) &= C \quad \text{(since 0.10752 came from } C\text{)}
\end{aligned}$$

$$\begin{aligned}
v_3(C) &= \max(v_2(H)a_{HC}, v_2(C)a_{CC}) \cdot b_C(3) \\
&= \max(0.0224 \times 0.8, 0.1792 \times 0.4) \times 0.5 \\
&= \max(0.01792, 0.07168) \times 0.5 = 0.07168 \times 0.5 = 0.03584 \\
bp_3(C) &= C \quad \text{(since 0.07168 came from } C\text{)}
\end{aligned}$$

$$\begin{aligned}
P^* &= \max(v_3(H), v_3(C)) = \max(0.010752, 0.03584) = 0.03584 \\
q_3^* &= C \quad \text{(since maximum probability is at } C\text{)}
\end{aligned}$$

$$\begin{aligned}
q_3^* &= C \\
q_2^* &= bp_3(q_3^*) = bp_3(C) = C \\
q_1^* &= bp_2(q_2^*) = bp_2(C) = H
\end{aligned}$$

CHAPTER 55

Statistical Computingprof Dootika

Discrete Inverse Transform Method

$$\begin{aligned}
F(x_1) &= p_1 \\
F(x_2) &= p_1 + p_2 = F(x_1) + p_2 \\
F(x_3) &= p_1 + p_2 + p_3 = F(x_2) + p_3 \\
&\vdots \\
F(x_k) &= \sum_{i=1}^{k} p_i
\end{aligned}$$

$$\begin{aligned}
F(1) &= 0.2 \\
F(2) &= 0.2 + 0.15 = 0.35 \\
F(3) &= 0.35 + 0.4 = 0.75 \\
F(4) &= 0.75 + 0.25 = 1.0
\end{aligned}$$

$$\begin{aligned}
F(0) &= 0.7 \\
F(1) &= 0.7 + 0.3 = 1.0
\end{aligned}$$

$$\begin{aligned}
U &\leq 1 - (1-p)^k \\
(1-p)^k &\leq 1 - U \\
k \ln(1-p) &\leq \ln(1-U) \\
k &\geq \frac{\ln(1-U)}{\ln(1-p)}
\end{aligned}$$

$$\begin{aligned}
X &= \left\lceil \frac{\ln(1-0.85)}{\ln(1-0.4)} \right\rceil \\
X &= \left\lceil \frac{\ln(0.15)}{\ln(0.6)} \right\rceil \\
X &= \left\lceil \frac{-1.8971}{-0.5108} \right\rceil \\
X &= \lceil 3.714 \rceil = 4
\end{aligned}$$

Box-Muller Method and Transformations

$$\begin{aligned}
Z_1 &= \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \\
Z_2 &= \sqrt{-2\ln(U_1)}\sin(2\pi U_2)
\end{aligned}$$

$$R = \sqrt{-2\ln(U_1)} = \sqrt{-2\ln(0.5)} = \sqrt{-2(-0.6931)} = \sqrt{1.3863} \approx 1.1774$$
$$\theta = 2\pi U_2 = 2\pi(0.75) = 1.5\pi$$

$$Z_1 = R\cos(\theta) = 1.1774 \times \cos(1.5\pi) = 1.1774 \times 0 = 0$$
$$Z_2 = R\sin(\theta) = 1.1774 \times \sin(1.5\pi) = 1.1774 \times (-1) = -1.1774$$

$$Z_1 = \sqrt{-2\ln(0.2)}\cos(2\pi \times 0.1) = \sqrt{3.2189}\cos(0.2\pi) \approx 1.7941 \times 0.8090 \approx 1.4514$$
$$Z_2 = \sqrt{-2\ln(0.2)}\sin(2\pi \times 0.1) = 1.7941\sin(0.2\pi) \approx 1.7941 \times 0.5878 \approx 1.0546$$

$$X_1 = \mu + \sigma Z_1 = 10 + 2(1.4514) = 10 + 2.9028 = 12.9028$$
$$X_2 = \mu + \sigma Z_2 = 10 + 2(1.0546) = 10 + 2.1092 = 12.1092$$

$$Z_1 = V_1 M = 0.5 \times 1.6651 = 0.8326$$
$$Z_2 = V_2 M = -0.5 \times 1.6651 = -0.8326$$

$$X_1 = \mu_1 + \sigma_1 Z_1$$
$$X_2 = \mu_2 + \sigma_2\left(\rho Z_1 + \sqrt{1-\rho^2}Z_2\right)$$

$$\sigma_1^2 = 4 \implies \sigma_1 = 2$$
$$\sigma_2^2 = 9 \implies \sigma_2 = 3$$
$$\rho\sigma_1\sigma_2 = 3 \implies \rho(2)(3) = 3 \implies 6\rho = 3 \implies \rho = 0.5$$

$$X_2 = \mu_2 + \sigma_2\left(\rho Z_1 + \sqrt{1-\rho^2}Z_2\right)$$
$$X_2 = 10 + 3\left(0.5(1.2) + \sqrt{1-0.5^2}(-0.5)\right)$$
$$X_2 = 10 + 3\left(0.6 + \sqrt{0.75}(-0.5)\right)$$
$$X_2 = 10 + 3\,(0.6 + 0.8660(-0.5))$$
$$X_2 = 10 + 3\,(0.6 - 0.4330)$$
$$X_2 = 10 + 3(0.167) = 10 + 0.501 = 10.501$$

Ratio-of-Uniforms Method

$$a = \sup_x \sqrt{h(x)}$$
$$b_- = \inf_x x\sqrt{h(x)}$$
$$b_+ = \sup_x x\sqrt{h(x)}$$

$$U \sim \text{Uniform}(0, a)$$
$$V \sim \text{Uniform}(b_-, b_+)$$

$$a = \sup_x \sqrt{h(x)}$$
$$b_- = \inf_x (x-\mu)\sqrt{h(x)}$$
$$b_+ = \sup_x (x-\mu)\sqrt{h(x)}$$

Multidimensional and Miscellaneous Sampling

$$\begin{aligned}
\mathbf{x}^{(1)} &= \boldsymbol{\mu} + \mathbf{L}\mathbf{z} \\
&= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0.8 & 0.6 \end{bmatrix} \begin{bmatrix} 0.5 \\ -1.2 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.4 - 0.72 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0.5 \\ -0.32 \end{bmatrix} = \begin{bmatrix} 1.5 \\ -1.32 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathbf{x}^{(2)} &= \boldsymbol{\mu} + \mathbf{L}\mathbf{z} \\
&= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0.8 & 0.6 \end{bmatrix} \begin{bmatrix} -0.3 \\ 1.0 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} -0.3 \\ -0.24 + 0.6 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} -0.3 \\ 0.36 \end{bmatrix} = \begin{bmatrix} 0.7 \\ -0.64 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{a} - \boldsymbol{\mu}_2) \\
&= 0 + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{bmatrix} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right) \\
&= \begin{bmatrix} 0.6 & 0.2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
&= (0.6)(-1) + (0.2)(1) = -0.6 + 0.2 = -0.4
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\
&= 2 - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= 2 - \begin{bmatrix} 0.6 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= 2 - 0.6 = 1.4
\end{aligned}$$

Advanced Optimization: MM and EM Algorithms

$$\begin{aligned}
g(\theta \mid \theta^{(t)}) &\geq f(\theta) \quad \text{for all } \theta \\
g(\theta^{(t)} \mid \theta^{(t)}) &= f(\theta^{(t)})
\end{aligned}$$

$$\begin{aligned}
P(X_1 \mid A) &= (\theta_A^{(0)})^8(1 - \theta_A^{(0)})^2 = (0.6)^8(0.4)^2 = 0.0168 \times 0.16 = 0.002687 \\
P(X_1 \mid B) &= (\theta_B^{(0)})^8(1 - \theta_B^{(0)})^2 = (0.5)^8(0.5)^2 = 0.003906 \times 0.25 = 0.000976
\end{aligned}$$

$$\begin{aligned}
P(X_2 \mid A) &= (0.6)^5(0.4)^5 = 0.07776 \times 0.01024 = 0.000796 \\
P(X_2 \mid B) &= (0.5)^5(0.5)^5 = 0.03125 \times 0.03125 = 0.000976
\end{aligned}$$

$$\begin{aligned}
\text{Expected Heads}_A &= \gamma_{1,A}(8) + \gamma_{2,A}(5) = 0.733(8) + 0.449(5) = 5.864 + 2.245 = 8.109 \\
\text{Expected Tails}_A &= \gamma_{1,A}(2) + \gamma_{2,A}(5) = 0.733(2) + 0.449(5) = 1.466 + 2.245 = 3.711
\end{aligned}$$

$$\text{Expected Heads}_B = \gamma_{1,B}(8) + \gamma_{2,B}(5) = 0.267(8) + 0.551(5) = 2.136 + 2.755 = 4.891$$
$$\text{Expected Tails}_B = \gamma_{1,B}(2) + \gamma_{2,B}(5) = 0.267(2) + 0.551(5) = 0.534 + 2.755 = 3.289$$

$$\begin{aligned}
\mathcal{N}(1.0 \mid 1.0, 1.0) &= 0.3989 \exp(0) = 0.3989 \\
\mathcal{N}(1.0 \mid 4.0, 1.0) &= 0.3989 \exp(-4.5) = 0.0044 \\
\gamma_{1,1} &= \frac{0.5(0.3989)}{0.5(0.3989) + 0.5(0.0044)} = \frac{0.1995}{0.2017} = 0.989 \\
\gamma_{1,2} &= 1 - 0.989 = 0.011
\end{aligned}$$

$$\begin{aligned}
\mathcal{N}(1.5 \mid 1.0, 1.0) &= 0.3989 \exp(-0.125) = 0.3521 \\
\mathcal{N}(1.5 \mid 4.0, 1.0) &= 0.3989 \exp(-3.125) = 0.0175 \\
\gamma_{2,1} &= \frac{0.5(0.3521)}{0.5(0.3521) + 0.5(0.0175)} = \frac{0.1761}{0.1848} = 0.953 \\
\gamma_{2,2} &= 1 - 0.953 = 0.047
\end{aligned}$$

$$\begin{aligned}
\mathcal{N}(4.5 \mid 1.0, 1.0) &= 0.3989 \exp(-6.125) = 0.0009 \\
\mathcal{N}(4.5 \mid 4.0, 1.0) &= 0.3989 \exp(-0.125) = 0.3521 \\
\gamma_{3,1} &= \frac{0.5(0.0009)}{0.5(0.0009) + 0.5(0.3521)} = \frac{0.0005}{0.1765} = 0.003 \\
\gamma_{3,2} &= 1 - 0.003 = 0.997
\end{aligned}$$

$$\begin{aligned}
N_1 &= 0.989 + 0.953 + 0.003 = 1.945 \\
N_2 &= 0.011 + 0.047 + 0.997 = 1.055
\end{aligned}$$

$$\begin{aligned}
\pi_1^{(new)} &= 1.945/3 = 0.648 \\
\pi_2^{(new)} &= 1.055/3 = 0.352
\end{aligned}$$

$$\begin{aligned}
\mu_1^{(new)} &= \frac{1}{1.945}[0.989(1.0) + 0.953(1.5) + 0.003(4.5)] = \frac{2.432}{1.945} = 1.250 \\
\mu_2^{(new)} &= \frac{1}{1.055}[0.011(1.0) + 0.047(1.5) + 0.997(4.5)] = \frac{4.568}{1.055} = 4.330
\end{aligned}$$

$$\begin{aligned}
\sigma_1^{2,(new)} &= \frac{1}{1.945}[0.989(1.0 - 1.250)^2 + 0.953(1.5 - 1.250)^2 + 0.003(4.5 - 1.250)^2] \\
&= \frac{0.0618 + 0.0596 + 0.0317}{1.945} = \frac{0.1531}{1.945} = 0.0787 \\
\sigma_2^{2,(new)} &= \frac{1}{1.055}[0.011(1.0 - 4.330)^2 + 0.047(1.5 - 4.330)^2 + 0.997(4.5 - 4.330)^2] \\
&= \frac{0.1220 + 0.3760 + 0.0288}{1.055} = \frac{0.5268}{1.055} = 0.4993
\end{aligned}$$

$$\theta^{(t+1)} = \arg\min_{\theta} g(\theta \mid \theta^{(t)})$$

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$$

$$\phi(\mathbb{E}[X]) \geq \mathbb{E}[\phi(X)]$$

$$g(\theta \mid \theta^{(t)}) = f(\theta^{(t)}) + \nabla f(\theta^{(t)})^T(\theta - \theta^{(t)}) + \frac{L}{2}\|\theta - \theta^{(t)}\|^2$$

$$\sqrt{z} \leq \sqrt{z_t} + \frac{1}{2\sqrt{z_t}}(z - z_t)$$

$$g(\theta \mid \theta^{(t)}) = \sum_{i=1}^{n} \left(\sqrt{(y_i - \theta^{(t)})^2 + \epsilon} + \frac{(y_i - \theta)^2 - (y_i - \theta^{(t)})^2}{2\sqrt{(y_i - \theta^{(t)})^2 + \epsilon}} \right)$$

$$\tilde{g}(\theta \mid \theta^{(t)}) = \sum_{i=1}^{n} \frac{(y_i - \theta)^2}{2\sqrt{(y_i - \theta^{(t)})^2 + \epsilon}}$$

$$\frac{d\tilde{g}}{d\theta} = \sum_{i=1}^{n} \frac{-2(y_i - \theta)}{2\sqrt{(y_i - \theta^{(t)})^2 + \epsilon}} = 0$$

$$\sum_{i=1}^{n} \frac{y_i}{\sqrt{(y_i - \theta^{(t)})^2 + \epsilon}} - \theta \sum_{i=1}^{n} \frac{1}{\sqrt{(y_i - \theta^{(t)})^2 + \epsilon}} = 0$$

$$\theta^{(t+1)} = \frac{\sum_{i=1}^{n} w_i^{(t)} y_i}{\sum_{i=1}^{n} w_i^{(t)}} \quad \text{where} \quad w_i^{(t)} = \frac{1}{\sqrt{(y_i - \theta^{(t)})^2 + \epsilon}}$$

$$Q(\theta \mid \theta^{(t)}) = \mathbb{E}_{Z|X,\theta^{(t)}}[\ln P(X, Z \mid \theta)]$$

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta \mid \theta^{(t)})$$

$$\gamma_{1,A} = \frac{\pi_A P(X_1 \mid A)}{\pi_A P(X_1 \mid A) + \pi_B P(X_1 \mid B)} = \frac{0.5(0.002687)}{0.5(0.002687) + 0.5(0.000976)} = 0.733$$

$$\gamma_{1,B} = 1 - 0.733 = 0.267$$

$$\gamma_{2,A} = \frac{0.5(0.000796)}{0.5(0.000796) + 0.5(0.000976)} = 0.449$$

$$\gamma_{2,B} = 1 - 0.449 = 0.551$$

$$\theta_A^{(1)} = \frac{8.109}{8.109 + 3.711} = 0.686$$

$$\theta_B^{(1)} = \frac{4.891}{4.891 + 3.289} = 0.598$$

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}$$

$$N_k = \sum_{i=1}^{N} \gamma_{ik}$$

$$\pi_k^{(t+1)} = \frac{N_k}{N}$$

$$\mu_k^{(t+1)} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_{ik} x_i$$

$$\Sigma_k^{(t+1)} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_{ik} (x_i - \mu_k^{(t+1)})(x_i - \mu_k^{(t+1)})^T$$

CHAPTER 56

Statistics For Data Science 2

Foundations of Probability

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
P(A \cup B) &= 0.5 + 0.4 - 0.2 \\
P(A \cup B) &= 0.7
\end{aligned}
$$

$$
\begin{aligned}
P((A \cup B)^c) &= 1 - P(A \cup B) \\
P((A \cup B)^c) &= 1 - 0.7 \\
P((A \cup B)^c) &= 0.3
\end{aligned}
$$

$$
\begin{aligned}
P(M \cup P) &= 1 - P((M \cup P)^c) \\
P(M \cup P) &= 1 - 0.10 = 0.90
\end{aligned}
$$

$$
\begin{aligned}
P(M \cap P) &= P(M) + P(P) - P(M \cup P) \\
P(M \cap P) &= 0.60 + 0.50 - 0.90 \\
P(M \cap P) &= 1.10 - 0.90 = 0.20
\end{aligned}
$$

$$
\begin{aligned}
P(A \cup B \cup C) = \quad & P(A) + P(B) + P(C) \\
& - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
& + P(A \cap B \cap C)
\end{aligned}
$$

$$
\begin{aligned}
P(S \cup T \cup L) = \quad & 0.50 + 0.40 + 0.30 \\
& - 0.20 - 0.15 - 0.10 \\
& + 0.05
\end{aligned}
$$

$$
\begin{aligned}
P(S \cup T \cup L) &= 1.20 - 0.45 + 0.05 \\
&= 0.75 + 0.05 \\
&= 0.80
\end{aligned}
$$

$$
\begin{aligned}
P(\text{None}) &= 1 - P(S \cup T \cup L) \\
P(\text{None}) &= 1 - 0.80 = 0.20
\end{aligned}
$$

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
P(A \cup B) &= 0.3 + 0.6 - 0.1 = 0.8
\end{aligned}
$$

$$
\begin{aligned}
P(A^c \cap B^c) &= P((A \cup B)^c) \\
&= 1 - P(A \cup B) \\
&= 1 - 0.8 = 0.2
\end{aligned}
$$

$$
\begin{aligned}
P(A^c \cup B^c) &= P((A \cap B)^c) \\
&= 1 - P(A \cap B) \\
&= 1 - 0.1 = 0.9
\end{aligned}
$$

Conditional Probability and Bayes' Theorem

$$
\begin{aligned}
P(D) &= P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2) + P(D \mid B_3)P(B_3) \\
&= (0.02)(0.50) + (0.03)(0.30) + (0.05)(0.20) \\
&= 0.010 + 0.009 + 0.010 \\
&= 0.029
\end{aligned}
$$

$$
\begin{aligned}
P(T^+) &= P(T^+ \mid D)P(D) + P(T^+ \mid D^c)P(D^c) \\
&= (0.95)(0.01) + (0.10)(0.99) \\
&= 0.0095 + 0.0990 \\
&= 0.1085
\end{aligned}
$$

Multiple Discrete Random Variables

$$
\begin{aligned}
E[XY] &= (1)(1)(0.2) + (1)(2)(0.1) + (2)(1)(0.3) + (2)(2)(0.4) \\
&= 0.2 + 0.2 + 0.6 + 1.6 = 2.6
\end{aligned}
$$

$$
\begin{aligned}
E[X+Y] &= (1+1)(0.2) + (1+2)(0.1) + (2+1)(0.3) + (2+2)(0.4) \\
&= 2(0.2) + 3(0.1) + 3(0.3) + 4(0.4) \\
&= 0.4 + 0.3 + 0.9 + 1.6 = 3.2
\end{aligned}
$$

Expectations Variance and Moments

$$
\begin{aligned}
E[X] &= (0 \times 0.1) + (1 \times 0.3) + (2 \times 0.4) + (3 \times 0.2) \\
E[X] &= 0 + 0.3 + 0.8 + 0.6 \\
E[X] &= 1.7
\end{aligned}
$$

$$
\begin{aligned}
E[X] &= \int_0^1 2x^2 \, dx \\
&= \left[\frac{2x^3}{3} \right]_0^1 \\
&= \frac{2(1)^3}{3} - 0 \\
&= \frac{2}{3}
\end{aligned}
$$

$$\begin{aligned} E[3X-2Y+5] &= 3(4)-2(10)+5 \\ &= 12-20+5 \\ &= -3 \end{aligned}$$

$$\begin{aligned} E[X] &= (-1)(0.2)+(0)(0.5)+(1)(0.3) \\ &= -0.2+0+0.3 = 0.1 \end{aligned}$$

$$\begin{aligned} E[X^2] &= (-1)^2(0.2)+(0)^2(0.5)+(1)^2(0.3) \\ &= 1(0.2)+0+1(0.3) = 0.5 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2]-(E[X])^2 \\ &= 0.5-(0.1)^2 \\ &= 0.5-0.01 = 0.49 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(3X-2Y) &= (3)^2\mathrm{Var}(X)+(-2)^2\mathrm{Var}(Y) \\ &= 9\mathrm{Var}(X)+4\mathrm{Var}(Y) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(3X-2Y) &= 9(4)+4(9) \\ &= 36+36 = 72 \end{aligned}$$

$$\begin{aligned} M_X'(t) &= \frac{d}{dt}\left(e^{2t+8t^2}\right) \\ &= e^{2t+8t^2}\cdot(2+16t) \end{aligned}$$

$$\begin{aligned} E[X] &= M_X'(0) \\ &= e^0\cdot(2+0) \\ &= 2 \end{aligned}$$

$$\begin{aligned} M_X''(t) &= \frac{d}{dt}\left[(2+16t)e^{2t+8t^2}\right] \\ &= 16\cdot e^{2t+8t^2}+(2+16t)\cdot e^{2t+8t^2}(2+16t) \\ &= e^{2t+8t^2}\left[16+(2+16t)^2\right] \end{aligned}$$

$$\begin{aligned} E[X^2] &= M_X''(0) \\ &= e^0\left[16+(2+0)^2\right] \\ &= 1\cdot(16+4) = 20 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2]-(E[X])^2 \\ &= 20-(2)^2 \\ &= 20-4 = 16 \end{aligned}$$

Functions and Expectations of Continuous Random Variables

$$F_Y(y) = P(Y \le y) = P(-2\ln(X) \le y)$$

$$P\left(\ln(X) \ge -\frac{y}{2}\right)$$

$$P\left(X \ge e^{-y/2}\right)$$

$$P\left(X \ge e^{-y/2}\right) = 1 - P\left(X \le e^{-y/2}\right) = 1 - F_X(e^{-y/2})$$

$$F_Y(y) = 1 - e^{-y/2}$$

$$f_Y(y) = \frac{d}{dy}\left(1 - e^{-y/2}\right) = -\left(-\frac{1}{2}e^{-y/2}\right) = \frac{1}{2}e^{-y/2}$$

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|$$

$$\frac{dx}{dy} = \frac{d}{dy}(\sqrt{y}) = \frac{1}{2\sqrt{y}}$$

$$f_Y(y) = f_X(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| = 2(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) = 1$$

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx$$

$$\int x^2\lambda e^{-\lambda x}\,dx = -x^2 e^{-\lambda x} - \int -e^{-\lambda x}(2x)\,dx = -x^2e^{-\lambda x} + \int 2xe^{-\lambda x}\,dx$$

$$\int 2xe^{-\lambda x}\,dx = -\frac{2x}{\lambda}e^{-\lambda x} - \int -\frac{2}{\lambda}e^{-\lambda x}\,dx = -\frac{2x}{\lambda}e^{-\lambda x} - \frac{2}{\lambda^2}e^{-\lambda x}$$

$$E[X^2] = \left[-x^2e^{-\lambda x} - \frac{2x}{\lambda}e^{-\lambda x} - \frac{2}{\lambda^2}e^{-\lambda x}\right]_0^\infty$$

$$E[X^2] = 0 - \left(0 - 0 - \frac{2}{\lambda^2}\right) = \frac{2}{\lambda^2}$$

$$E[XY] = \int_0^1\int_0^1 (xy)(x+y)\,dx\,dy = \int_0^1\int_0^1 (x^2y + xy^2)\,dx\,dy$$

$$\int_0^1 (x^2y + xy^2)\,dx = \left[\frac{x^3}{3}y + \frac{x^2}{2}y^2\right]_{x=0}^{x=1} = \left(\frac{1}{3}y + \frac{1}{2}y^2\right) - 0 = \frac{y}{3} + \frac{y^2}{2}$$

$$E[XY] = \int_0^1\left(\frac{y}{3} + \frac{y^2}{2}\right)\,dy = \left[\frac{y^2}{6} + \frac{y^3}{6}\right]_0^1 = \left(\frac{1}{6} + \frac{1}{6}\right) - 0 = \frac{1}{3}$$

$$J = \det\begin{bmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{bmatrix} = \left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial y}{\partial v}\right) - \left(\frac{\partial x}{\partial v}\right)\left(\frac{\partial y}{\partial u}\right)$$

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))|J|$$

$$J = \det\begin{bmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{bmatrix} = \det\begin{bmatrix}1/2 & 1/2\\ 1/2 & -1/2\end{bmatrix}$$

$$J = \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$$

$$x^2 + y^2 = \left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2 = \frac{u^2 + 2uv + v^2}{4} + \frac{u^2 - 2uv + v^2}{4} = \frac{2u^2 + 2v^2}{4} = \frac{u^2 + v^2}{2}$$

$$f_{U,V}(u,v) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}|J| = \frac{1}{2\pi}e^{-(u^2+v^2)/4}\left(\frac{1}{2}\right) = \frac{1}{4\pi}e^{-u^2/4}e^{-v^2/4}$$

Bayesian Estimation and Confidence Intervals

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

$$\mu_n = \tau_n^2 \left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{x}}{\sigma^2} \right)$$

CHAPTER 57

Statistics For Data Science Ii

Independence and Functions of Discrete Random Variables

$$
\begin{aligned}
P_X(0) &= 0.1 + 0.2 + 0.1 = 0.4 \\
P_X(1) &= 0.15 + 0.3 + 0.15 = 0.6
\end{aligned}
$$

$$
\begin{aligned}
P_Y(0) &= 0.1 + 0.15 = 0.25 \\
P_Y(1) &= 0.2 + 0.3 = 0.50 \\
P_Y(2) &= 0.1 + 0.15 = 0.25
\end{aligned}
$$

$$
\begin{aligned}
P_X(0)P_Y(0) &= (0.4)(0.25) = 0.10 = P_{X,Y}(0,0) \\
P_X(0)P_Y(1) &= (0.4)(0.50) = 0.20 = P_{X,Y}(0,1) \\
P_X(0)P_Y(2) &= (0.4)(0.25) = 0.10 = P_{X,Y}(0,2) \\
P_X(1)P_Y(0) &= (0.6)(0.25) = 0.15 = P_{X,Y}(1,0) \\
P_X(1)P_Y(1) &= (0.6)(0.50) = 0.30 = P_{X,Y}(1,1) \\
P_X(1)P_Y(2) &= (0.6)(0.25) = 0.15 = P_{X,Y}(1,2)
\end{aligned}
$$

$$
\begin{aligned}
(-2)^2 &= 4 \\
(-1)^2 &= 1 \\
(0)^2 &= 0 \\
(1)^2 &= 1 \\
(2)^2 &= 4
\end{aligned}
$$

$$
\begin{aligned}
(1,1) &\implies \max(1,1) = 1 \\
(1,2) &\implies \max(1,2) = 2 \\
(2,1) &\implies \max(2,1) = 2 \\
(2,2) &\implies \max(2,2) = 2
\end{aligned}
$$

Continuous Random Variables and Key Distributions

$$
\begin{aligned}
\int_1^2 \frac{3}{8}x^2\,dx &= \left[\frac{3}{8}\cdot\frac{x^3}{3}\right]_1^2 \\
&= \left[\frac{x^3}{8}\right]_1^2 \\
&= \frac{2^3}{8} - \frac{1^3}{8} \\
&= \frac{8}{8} - \frac{1}{8} = \frac{7}{8}
\end{aligned}
$$

$$\begin{aligned} E[X] &= \int_0^1 x(2x)\,dx = \int_0^1 2x^2\,dx \\ &= \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3} \end{aligned}$$

$$\begin{aligned} E[X^2] &= \int_0^1 x^2(2x)\,dx = \int_0^1 2x^3\,dx \\ &= \left[\frac{2x^4}{4}\right]_0^1 = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} \\ &= \frac{9-8}{18} = \frac{1}{18} \end{aligned}$$

$$\begin{aligned} P(X > 5) &= e^{-0.2\times 5} \\ &= e^{-1} \approx 0.3679 \end{aligned}$$

$$\begin{aligned} P(60 \leq X \leq 85) &= 0.9332 - 0.1587 \\ &= 0.7745 \end{aligned}$$

Point Estimation and Properties of Estimators

$$\begin{aligned} \mathrm{MSE}(\hat{\sigma}^2) &= \mathrm{Var}(\hat{\sigma}^2) + [\mathrm{Bias}(\hat{\sigma}^2)]^2 \\ &= \frac{2(n-1)\sigma^4}{n^2} + \left(-\frac{\sigma^2}{n}\right)^2 \\ &= \frac{2n-2}{n^2}\sigma^4 + \frac{1}{n^2}\sigma^4 \\ &= \frac{2n-1}{n^2}\sigma^4 \end{aligned}$$

Fundamentals of Hypothesis Testing and Z-Tests

$$\frac{\bar{X} - 50}{12/\sqrt{36}} > 1.645$$

$$\frac{\bar{X} - 50}{2} > 1.645$$

$$\bar{X} - 50 > 3.29 \implies \bar{X} > 53.29$$

$$\text{Power} = P(\text{Reject } H_0 \mid \mu = \mu_a)$$

$$Z = \frac{53.29 - 55}{2} = \frac{-1.71}{2} = -0.855$$

$$P(Z > -0.855) = 1 - P(Z < -0.855) \approx 1 - 0.1963 = 0.8037$$

$$H_0 : \mu = 500$$

$$H_a : \mu < 500$$

$$Z = \frac{498 - 500}{5/\sqrt{40}} = \frac{-2}{5/6.325} = \frac{-2}{0.7906} \approx -2.53$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$H_0 : \mu_A - \mu_B = 0$$

$$H_a : \mu_A - \mu_B \neq 0$$

$$Z = \frac{(12.5 - 11.8) - 0}{\sqrt{\frac{1.2^2}{50} + \frac{1.5^2}{45}}} = \frac{0.7}{\sqrt{\frac{1.44}{50} + \frac{2.25}{45}}}$$

$$Z = \frac{0.7}{\sqrt{0.0288 + 0.05}} = \frac{0.7}{\sqrt{0.0788}} \approx \frac{0.7}{0.2807} \approx 2.49$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$H_0 : p = 0.60$$

$$H_a : p > 0.60$$

$$Z = \frac{0.64 - 0.60}{\sqrt{\frac{0.60 \times 0.40}{400}}} = \frac{0.04}{\sqrt{\frac{0.24}{400}}} = \frac{0.04}{\sqrt{0.0006}} \approx \frac{0.04}{0.0245} \approx 1.63$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$H_0 : p_B - p_A = 0$$

$$H_a : p_B - p_A > 0$$

$$\hat{p}_A = \frac{45}{500} = 0.09$$

$$\hat{p}_B = \frac{72}{600} = 0.12$$

$$Z = \frac{0.12 - 0.09}{\sqrt{0.1064(1 - 0.1064)\left(\frac{1}{600} + \frac{1}{500}\right)}}$$

$$Z = \frac{0.03}{\sqrt{0.1064(0.8936)\,(0.001667 + 0.002)}}$$

$$Z = \frac{0.03}{\sqrt{0.0951 \times 0.003667}} = \frac{0.03}{\sqrt{0.0003487}} \approx \frac{0.03}{0.01867} \approx 1.61$$

Likelihood Ratio Tests and Goodness of Fit

$$\begin{aligned}
\chi^2 &= \frac{(15-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(27-20)^2}{20} + \frac{(18-20)^2}{20} \\
&= \frac{25}{20} + \frac{25}{20} + \frac{4}{20} + \frac{9}{20} + \frac{49}{20} + \frac{4}{20} \\
&= \frac{116}{20} = 5.8
\end{aligned}$$

CHAPTER 58

Strategies For Professional Growth Spg Iit Madras Bs Degree

Mid-Course Correction and Feedback Loops

$$(82 - 85)^2 = 9$$
$$(88 - 85)^2 = 9$$
$$(85 - 85)^2 = 0$$
$$(90 - 85)^2 = 25$$
$$(80 - 85)^2 = 25$$

CHAPTER 59

System Commands

No prominent formulae extracted for this course.

CHAPTER 60

Tools In Data Science

Correlation Analysis and Outlier Detection

$$r_{XY\cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{1 - r_{XZ}^2}\sqrt{1 - r_{YZ}^2}}$$

$$D_i^2 = (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$