$$J(w_1, w_2, b) = \frac{1}{4} \sum (x^{(i)T} w + b - y^{(i)})^2$$ **Q1**

$$= \frac{1}{4}\left(\left[\left([0\ 0]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b\right) - 0\right]^2 + \left[\left([1\ 1]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b\right) - 0\right]^2 + \left[\left([1\ 0]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b\right) - 1\right]^2 + \right.$$

$$\left. \left[\left([0\ 1]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b\right) - 1\right]^2 \right)$$

$$J = \frac{1}{4}\left([b]^2 + [w_1 + w_2 + b]^2 + [w_1 + b - 1]^2 + [w_2 + b - 1]^2 \right)$$

$$\frac{\partial J}{\partial w_1} = \frac{1}{4}\left(2 \cdot 1 [w_1 + w_2 + b] + 2 \cdot 1 [w_1 + b - 1]\right)$$

$$\frac{\partial J}{\partial w_1} = \frac{1}{2}(w_1 + w_2 + b + w_1 + b - 1) = \frac{1}{2}(2w_1 + w_2 + 2b - 1) = 0$$

$$\frac{\partial J}{\partial w_2} = \frac{1}{4}\left([w_1 + w_2 + b]^2 + [w_2 + b - 1]^2 \right)$$

$$\frac{\partial J}{\partial w_2} = \frac{1}{4}\left(2 \cdot 1 [w_1 + w_2 + b] + 2 \cdot 1 [w_2 + b - 1]\right)$$

$$\frac{\partial J}{\partial w_2} = \frac{1}{2}(w_1 + w_2 + b + w_2 + b - 1) = \frac{1}{2}(w_1 + 2w_2 + 2b - 1) = 0$$

$$\frac{\partial J}{\partial b} = \frac{1}{4}\left([b]^2 + [w_1 + w_2 + b]^2 + [w_1 + b - 1]^2 + [w_2 + b - 1]^2 \right)$$

$$\frac{\partial J}{\partial b} = \frac{1}{4}\left(2 \cdot 1 [b] + 2 \cdot 1 [w_1 + w_2 + b] + 2 \cdot 1 [w_1 + b - 1] + 2 \cdot 1 [w_2 + b - 1]\right)$$

$$\frac{\partial J}{\partial b} = \frac{1}{2}(b + w_1 + w_2 + b + w_1 + b - 1 + w_2 + b - 1)$$

$$\frac{\partial J}{\partial b} = \frac{1}{2}(2w_1 + 2w_2 + 4b - 2) = 0$$

$$\frac{1}{2}(2w_1 + w_2 + 2b - 1) = \frac{1}{2}(w_1 + 2w_2 + 2b - 1)$$

$$2w_1 + w_2 + 2b - 1 = w_1 + 2w_2 + 2b - 1$$
$$-w_1 \ -w_2 \qquad\qquad -w_1 \ -w_2$$

$$w_1 + 2b - 1 = w_2 + 2b - 1$$
$$-2b + 1 \qquad\qquad -2b + 1$$

$$w_1 = w_2$$

$\frac{1}{2}(2w_1 + 2w_2 + 4b - 2) = 0$

$\frac{1}{2}(4w_1 + 4b - 2) = 0$ because $w_1 = w_2$

$2w_1 + 2b - 1 = 0$

$w_1 = \frac{1-2b}{2}$

$\frac{1}{2}(2w_2 + w_1 + 2b - 1) = 0$ because $w_1 = w_2$

$\frac{1}{2}(3w_1 + 2b - 1) = 0$

$\frac{1}{2}(3(\frac{1-2b}{2}) + 2b - 1) = 0$

$\frac{1}{2}(\frac{3-6b}{2} + 2b - 1)$

$\frac{3}{4} - \frac{6b}{4} + b - \frac{1}{2} = 0$

$-\frac{6b}{4} + b + \frac{1}{4} = 0$

$-\frac{3b}{2} + b = -\frac{1}{4}$

$-\frac{1}{2}b = -\frac{1}{4}$

$-2 \qquad -2$

$b = \frac{1}{2}$

$2w_1 + 2(\frac{1}{2}) - 1 = 0$

$w_1 = w_2 = 0$

$$\alpha w^T w = \alpha w^T I w = \alpha w^T S w$$ Q3

$$S = \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix}$$

$$\alpha \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\alpha \begin{bmatrix} w_1 S_1 + w_2 S_3 & w_1 S_2 + w_2 S_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$w_1(w_1 S_1 + w_2 S_3) + w_2(w_1 S_2 + w_2 S_4)$$

$$\alpha [w_1^2 S_1 + w_1 w_2 S_3 + w_1 w_2 S_2 + w_2^2 S_4]$$

$$\alpha [w_1^2 S_1 + w_1 w_2 (S_3 + S_2) + w_2^2 S_4]$$

$$w_1^2 - 2 w_1 w_2 + w_2^2 \qquad (w_1 - w_2)^2$$

$$S_1 = S_4 = 1$$

$$S_2 = S_3 = -1$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$ conditional probability

Q4

Student state based off observable state

$$P(x_t \mid y_1, \dots, y_t) = \frac{P(x_t, y_1, \dots, y_t)}{P(y_1, \dots, y_t)}$$

$$\Downarrow$$

$$P(x_t \mid y_1, \dots, y_t) \propto P(x_t, y_1, \dots, y_t)$$

$$P(A \mid B) = \frac{P(A, B)}{P(B)} \quad \propto \quad P(A \mid B) = P(A, B)$$

$P(B)$ ← not function of A

$P(a \mid b, c) \propto P(b \mid a, c)\, P(a, c) \propto$ bayes rule:

$$P(\overset{c}{x_t} \mid \underbrace{y_1, \dots,}_{y_1, \dots, y_{t-1}] a} \overset{b}{y_t}) \propto P(y_t \mid x_t, y_1, \dots, y_{t-1})\, P(x_t \mid y_1, \dots, y_{t-1})$$

based off total probability (example):

$$P(q) = \sum_r P(q, r)$$

Using total probability:

$$P(y_t \mid x_t) \sum_{x_{t-1}} P(x_t, x_{t-1} \mid y_1, \dots, y_{t-1})$$

Using conditional probability:

$$P(y_t \mid x_t) \sum_{x_{t-1}} P(x_t \mid x_{t-1}, y_1, \dots, y_{t-1})\, P(x_{t-1} \mid y_1, \dots, y_{t-1})$$

→

Using Markov Property and Conditional independence:

$$P(y_t \mid x_t) \sum_{x_{t-1}} \underbrace{P(x_t \mid x_{t-1})}_{\text{Markov}} P(x_{t-1} \mid y_1, \dots, y_{t-1})$$

Q5

$$P(y \mid x) = \mathcal{N}(\mu = x^T w, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - x^T w)^2}{2\sigma^2}\right)$$

$$P(D \mid \mu = x^T w, \sigma^2) = \log \prod_{i=0}^{n} P(y^{(i)} \mid x^{(i)}, w, \sigma^2)$$

$$= \sum_{i=1}^{n} \log P(y^{(i)} \mid x^{(i)}, w, \sigma^2)$$

$$= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x^T w - y)^2}{2\sigma^2}\right)\right)$$

$$= \sum_{i=1}^{n} - (x^T w - y)^2 - \log\left(\sqrt{2\pi\sigma^2}\right)$$

$$= -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x^T w - y)^2 - \underbrace{\log(\sqrt{2\pi\sigma^2})}_{\log(\sqrt{2\pi}) + \log(\sigma)}$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left((x^T w - y)^2 - \log(\sigma)\right) + C$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((x^T w - y)^2 - \log(\sigma)\right) + C$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} (x^T w - y)^2 - n\log(\sigma) + C$$

$$\nabla_w \log P(D \mid w, \sigma^2) = \nabla_w\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} (x^T w - y)^2 - n\log(\sigma) + C\right)$$

$$= \nabla_w\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n} (x^T w - y)^2 + n\log(\sigma) - C\right)$$ ← factor -1

$$0 = \frac{1}{\sigma^2}\sum_{i=1}^{n} x(x^T w - y)$$ ← multiply σ^2 to both sides

$$0 = \sum_{i=1}^{n} x x^T w - xy$$

$$\sum_{i=1}^{n} xy = \sum_{i=1}^{n} x x^T w$$

$$w = \left(\sum_{i=1}^{n} x x^T\right)^{-1}\left(\sum_{i=1}^{n} xy\right)$$

→

$$= \nabla_\sigma \left(\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x^T w - y)^2 + n \log(\sigma) - C \right) \text{ factor } -1$$

$$= -\frac{1}{\sigma^3} \sum_{i=1}^{n} (x^T w - y)^2 + \frac{n}{\sigma}$$

$$-\frac{n}{\sigma} = -\frac{1}{\sigma^3} \sum_{i=1}^{n} (x^T w - y)^2$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x^T w - y)^2$$