ECE 421 Homework 3

Name: Mark Qi

Student ID: 1006764645

Question 1,

$D = \{(x_i, y_i)\}_{n=1}^{N}$

$x_n \in \mathbb{R}^2$

$y_n \in \{+1, -1\}$

Bias Term Set To Zero

w with same dimension as x

$$h(\underline{x}) = \frac{e^{\underline{w}^T \underline{x}}}{1+e^{\underline{w}^T \underline{x}}} = \frac{1}{1+e^{-\underline{w}^T \underline{x}}}$$

$$\underline{x_1} = [1 \quad 1]^T \qquad y_1 = 1$$

$$\underline{x_2} = [1 \quad 0]^T \qquad y_2 = -1$$

$$E_{in}(\underline{w}) = -\sum_{n=1}^{N} \log[P_{\underline{w}}(y_n | \underline{x_n})] + \lambda \| \underline{w} \|_2^2, \quad \lambda > 0$$

$$P_{\underline{w}}(y | \underline{x}) = \begin{cases} h(\underline{x}) & y = +1 \\ 1 - h(\underline{x}) & y = -1 \end{cases}$$

Part A,

$\lambda = 0$, Optimal $\underline{w}^* = [w_1 \quad w_2]^T$

$$E_{in}(\underline{w}^*) = -\log[P_{\underline{w}^*}(y_1 | \underline{x_1})] - \log[P_{\underline{w}^*}(y_2 | \underline{x_2})]$$

$$P_{\underline{w}^*}(y_1 | \underline{x_1}) = h(\underline{x_1}) = \frac{1}{1+e^{-\underline{w_*}^T x_1}} = \frac{1}{1+e^{-(w_1+w_2)}}$$

$$P_{\underline{w}^*}(y_2 | \underline{x_2}) = 1 - h(\underline{x_2}) = 1 - \frac{1}{1+e^{-\underline{w_*}^T x_2}}$$

$$= 1 - \frac{1}{1+e^{-w_1}} = \frac{1+e^{-w_1}}{1+e^{-w_1}} - \frac{1}{1+e^{-w_1}}$$

$$= \frac{e^{-w_1}}{1+e^{-w_1}} = \frac{1}{1+e^{w_1}}$$

$$E_{in}(\underline{w}^*) = -\log\left(\frac{1}{1+e^{-(w_1+w_2)}}\right) - \log\left(\frac{1}{1+e^{w_1}}\right)$$

$$= \log\left(1+e^{-(w_1+w_2)}\right) + \log\left(1+e^{w_1}\right)$$

To Minimize $E_{in}(\underline{w}^*)$, we minimize $e^{-(w_1+w_2)}$ and e^{w_1}

$e^{-\infty}$ is the minimum, thus

$w_1 + w_2 = \infty$

$w_1 = -\infty$

$w_2 = 2\infty$

$$\underline{w}^* = \begin{bmatrix} -\infty \\ 2\infty \end{bmatrix}$$

Part B,

$$||w||_2 << 1$$

$$\log(1+e^{-y_n w^T x_n}) \approx \log(2) - \frac{1}{2} y_n w^T x_n$$

$$\underline{E_{in}(w)} = -\log\left(\frac{1}{1+e^{-(w_1+w_2)}}\right) - \log\left(\frac{1}{1+e^{w_1}}\right)$$

$$= \log(1+e^{-(w_1+w_2)}) + \log(1+e^{w_1})$$

$$\approx \log(2) - \frac{1}{2}(w_1+w_2) + \log(2) + \frac{1}{2}w_1 + \lambda ||w||_2^2$$

$$= 2\log(2) - \frac{1}{2}(w_1+w_2) + \frac{1}{2}w_1 + \lambda(w_1^2+w_2^2)$$

$$\frac{\partial E_{in}(w)}{\partial w_1} = -\frac{1}{2} + \frac{1}{2} + 2\lambda w_1 = 0 \qquad \boxed{w_1^* = 0}$$

$$\frac{\partial E_{in}(w)}{\partial w_2} = -\frac{1}{2} + 2\lambda w_2 = 0$$

$$-1 + 4\lambda w_2 = 0$$

$$\boxed{w_2^* = \frac{1}{4\lambda}}$$

$$w^* = \begin{bmatrix} 0 \\ -\frac{1}{4\lambda} \end{bmatrix}$$

Question 2,

$\underline{x} = (x_1, x_2) \in \mathbb{R}^2$

$y \in \{1, 2, 3\}$

$D = \{(\underline{x_1}, y_1), (\underline{x_2}, y_2), (\underline{x_3}, y_3), (\underline{x_4}, y_4), (\underline{x_5}, y_5)\}$

$\underline{x_1} = |-1, 0|^T, \quad y_1 = 1$

$\underline{x_2} = |1, 0|^T, \quad y_2 = 2$

$\underline{x_3} = |1, 1|^T, \quad y_3 = 2$

$\underline{x_4} = |-1, 1|^T, \quad y_4 = 1$

$\underline{x_5} = |0, 3|^T, \quad y_5 = 3$

$\Omega = \{\underline{w(1)}, \underline{w(2)}, \underline{w(3)}\}$

$\underline{w(i)} \in \mathbb{R}^3, \quad i \in \{1, 2, 3\}$

$$\hat{P}_\Omega(i \mid \underline{x}) = \frac{e^{[w(i)^T \hat{x}]}}{\sum_{j=1}^{3} e^{[w(j)^T \hat{x}]}}, \quad i = 1, 2, 3$$

$\hat{x} = (x_0 = 1, x_1, x_2)^T \in \mathbb{R}^3$, Augmented Vector

$$E_{in}(\Omega) = \frac{1}{5} \sum_{n=1}^{5} e_n(\Omega)$$

$$e_n(\Omega) = -\log \hat{P}_\Omega(y_n \mid \underline{x_n})$$

$\underline{w(1)} = |1 \quad 0 \quad 0|^T$

$\underline{w(2)} = |0 \quad 1 \quad 0|^T$

$\underline{w(3)} = |0 \quad 0 \quad 1|^T$

$$\nabla_{\underline{w(1)}} \left\{ e_1(\underline{\Omega}) \right\} = \nabla_{\underline{w(1)}} \left\{ -\log \hat{P}_{\underline{\Omega}}(y_1 | \underline{x_1}) \right\}$$

$$y_1 = 1$$

$$\underline{x_1} = [-1, 0]^T$$

$$= \nabla_{\underline{w(1)}} \left\{ -\log \frac{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}} \right\}$$

$$= -\nabla_{\underline{w(1)}} \left\{ \log \frac{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}} \right\}$$

$$= -\frac{\cancel{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}}}{\cancel{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}} \cdot$$

$$\frac{\hat{\underline{x}}\, \cancel{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}} \sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]} - \cancel{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}\, \hat{\underline{x}}\, e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\left(\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}\right)^{\cancel{2}}}$$

$$= -\left(\hat{\underline{x}} - \frac{\hat{\underline{x}}\, e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}} \right) = -\hat{\underline{x}} \left(1 - \frac{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}} \right)$$

$$\boxed{= -\hat{\underline{x}}\, \frac{e^{[\underline{w}^T(2)\,\hat{\underline{x}}]} + e^{[\underline{w}^T(3)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}}} \quad \nabla_{\underline{w(1)}} \left\{ e_1(\underline{\Omega}) \right\}$$

$$\nabla_{\underline{w(2)}} \{ e_1(\underline{\Omega}) \} = \nabla_{\underline{w(2)}} \left\{ -\log \hat{P}_{\underline{\Omega}}(y_1 | \underline{x_1}) \right\}$$

$y_1 = 1$

$\underline{x_1} = [-1, 0]^T$

$$= \nabla_{\underline{w(2)}} \left\{ -\log \frac{e^{[\underline{w^T(1)\hat{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w^T(j)\hat{x}}]}} \right\}$$

$$= -\nabla_{\underline{w(2)}} \left\{ \log \frac{e^{[\underline{w^T(1)\hat{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w^T(j)\hat{x}}]}} \right\}$$

$$= \cancel{-} \frac{\cancel{\sum_{j=1}^{3} e^{[\underline{w^T(j)\hat{x}}]}}}{\cancel{e^{[\underline{w^T(1)\hat{x}}]}}} \cdot \frac{\cancel{-}\, \cancel{e^{[\underline{w^T(1)\hat{x}}]}}\, \hat{x}\, e^{[\underline{w^T(2)\hat{x}}]}}{\left(\sum_{j=1}^{3} e^{[\underline{w^T(j)\hat{x}}]}\right)^{\cancel{2}}}$$

$$\boxed{= \frac{\hat{x}\, e^{[\underline{w^T(2)\hat{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w^T(j)\hat{x}}]}}} \quad \nabla_{\underline{w(2)}} \{ e_1(\underline{\Omega}) \}$$

$$\nabla_{\underline{w}(3)} \{ e_1(\Omega) \} = \nabla_{\underline{w}(3)} \left\{ - \log \hat{P}_{\Omega}(y_1 | \underline{x}_1) \right\}$$

$$y_1 = 1$$

$$\underline{x}_1 = [-1, 0]^T$$

$$= \nabla_{\underline{w}(3)} \left\{ - \log \frac{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}} \right\}$$

$$= - \nabla_{\underline{w}(3)} \left\{ \log \frac{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}} \right\}$$

$$= \cancel{-} \frac{\cancel{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}}}{\cancel{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}} \cdot \frac{\cancel{-}\, \cancel{e^{[\underline{w}^T(1)\,\hat{\underline{x}}]}}\, \hat{\underline{x}}\, e^{[\underline{w}^T(3)\,\hat{\underline{x}}]}}{\left(\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}\right)^{\cancel{2}}}$$

$$\boxed{= \frac{\hat{\underline{x}}\, e^{[\underline{w}^T(3)\,\hat{\underline{x}}]}}{\sum_{j=1}^{3} e^{[\underline{w}^T(j)\,\hat{\underline{x}}]}}} \quad \nabla_{\underline{w}(3)} \{ e_1(\Omega) \}$$