ECE421 Homework 2

Name: Mark Qi

Student ID: 1006764645

Question 1,

$f: \mathbb{R}^d \rightarrow \mathbb{R}$

$$\nabla f(w) = \left[\frac{\partial f(w)}{\partial w_1}, \frac{\partial f(w)}{\partial w_2}, \ldots, \frac{\partial f(w)}{\partial w_d}\right]^T \in \mathbb{R}^{d \times 1}$$

$A \in \mathbb{R}^{d \times d}$, $w \in \mathbb{R}^{d \times 1}$

(a) $f(w) = w^T A v + w^T A^T v + v^T A w + v^T A^T w$, $v \in \mathbb{R}^d$

$$\nabla_w f(w) = \frac{\partial}{\partial w}(w^T A v) + \frac{\partial}{\partial w}(w^T A^T v) + \frac{\partial}{\partial w} v^T A w + \frac{\partial}{\partial w}(v^T A^T w)$$

$$= \frac{\partial}{\partial w}\left[\begin{bmatrix} w_1 & w_2 & \ldots & w_d \end{bmatrix} A v\right] + \frac{\partial}{\partial w}\left[\begin{bmatrix} w_1 & w_2 & \ldots & w_d \end{bmatrix} A^T v\right] + \frac{\partial}{\partial w}\left[v^T A \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}\right] + \frac{\partial}{\partial w}\left[v^T A^T \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}\right]$$

$$= Av + A^T v + v^T A + v^T A^T$$

(b) $f(w) = w^T A w$

A is not necessarily symmetric

$$\nabla_w f(w) = \frac{\partial}{\partial w}\left[w^T A w\right] = (A + A^T) w$$

$$\frac{\partial w^T A w}{\partial w_k} = \sum_{j=1}^{n} a_{kj} w_j + \sum_{i=1}^{n} a_{ik} w_i$$

$= (\text{Kth Row of } A)\, w + (\text{transpose of kth Column of } A)\, w$

(c) $f(w) = \log\left(\sum_{i=1}^{d} \exp(w_i)\right)$

$$\nabla_w f(w) = \frac{1}{\sum_{i=1}^{d} \exp(w_i)} \times \frac{\partial}{\partial w} \sum_{i=1}^{d} \exp(w_i)$$

$$= \frac{1}{\sum_{i=1}^{d} \exp(w_i)} \begin{bmatrix} e^{w_1} \\ e^{w_2} \\ \vdots \\ e^{w_d} \end{bmatrix}$$

(d) $f(w) = \sum_{i=1}^{d} \log(1 + \exp(w_i))$

$$\nabla_w f(w) = \frac{\partial}{\partial w} \sum_{i=1}^{d} \log(1 + \exp(w_i))$$

$$= \sum_{i=1}^{d} \frac{\partial}{\partial w} (\log(1 + \exp(w_i)))$$

$$= \sum_{i=1}^{d} \frac{1}{1 + \exp(w_i)} \times \exp(w_i)$$

$$= \sum_{i=1}^{d} \frac{e^{w_i}}{1 + e^{w_i}}$$

(e) $f(w) = \sqrt{1 + \|w\|_2^2}$

$$= \sqrt{1 + \sum_{i=1}^{d} w_i^2}$$

$$\nabla_w f(w) = \frac{\partial}{\partial w} \sqrt{1 + \sum_{i=1}^{d} w_i^2}$$

$$= \frac{1}{2}\left(1 + \sum_{i=1}^{d} w_i^2\right)^{-\frac{1}{2}} \times \frac{\partial}{\partial w}\left(1 + \sum_{i=1}^{d} w_i^2\right)$$

$$= \frac{1}{2}\left(1 + \sum_{i=1}^{d} w_i^2\right)^{-\frac{1}{2}} \begin{vmatrix} 2w_1 \\ 2w_2 \\ \vdots \\ 2w_d \end{vmatrix}$$

$$= \left(1 + \sum_{i=1}^{d} w_i^2\right)^{-\frac{1}{2}} w$$

Question 2,

$x_1 = -1, y_1 = -2$

$x_2 = 0, y_2 = 0$

$x_3 = 1, y_3 = 1$

$\hat{y} = w_0 + w_1 x$

$$E_{in}(w_0, w_1) = \frac{1}{3}\sum_{i=1}^{3}(y_i - \hat{y}_i)^2$$

Part A,

$\underline{w} = (w_0, w_1)^T \in \mathbb{R}^{2\times1}$

$$\underline{y} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathcal{X}\,\underline{w} \longrightarrow \mathbb{R}^{3\times2} \times \mathbb{R}^{2\times1} = \mathbb{R}^{3\times1}$$

$$\mathcal{X} = \begin{bmatrix} x_0^T \\ x_1^T \\ x_2^T \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Part B,

From Lecture, We learned $\underline{w^*}$ can be computed analytically.

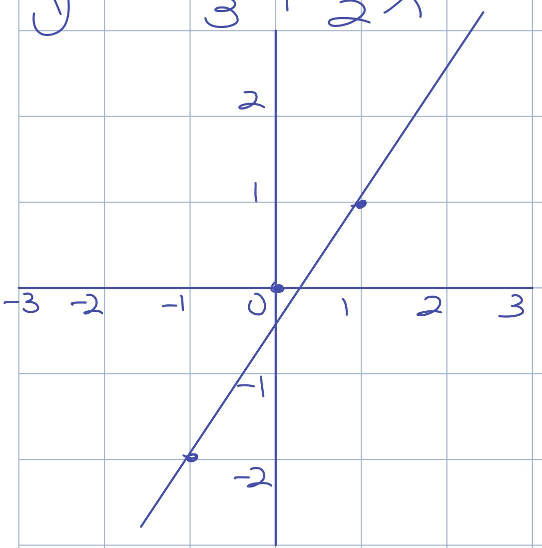
$$\underline{w^*} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T Y$$

$$\underline{w}^* = \left[\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{3}{2} \end{bmatrix}$$

$$y = -\frac{1}{3} + \frac{3}{2}x$$



Part C

$$\hat{y} = w_0 + w_1 x + w_2 x^2$$

$$\underline{w} \in \mathbb{R}^{3\times 1}$$

$$\mathcal{X} \in \mathbb{R}^{3\times 3}$$

$$\mathcal{X} = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

By matrix calculator,

$$\underline{w}^* = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \underline{y}$$

$$= \begin{bmatrix} 1 & 0 & -1 \\ 0 & \frac{1}{2} & 0 \\ -1 & 0 & \frac{3}{2} \end{bmatrix} \mathcal{X}^T \underline{y}$$

$$= \begin{bmatrix} 0 \\ \frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}$$

$$E_{in}(\underline{w}^*) = \frac{1}{3}(0 + 0 + 0) = 0$$

$E_{in}(\underline{w}^*)$ is zero since there is three degree of freedom and three datapoints