CS 541 - Deep Learning.

Abhiroop Ajith.

Youness Bani.

Problem 1:-

$$J(\omega_1, \omega_2, b) = \frac{1}{4} \sum (x^{i^T} \omega + b - y^i)^2$$

$$= \frac{1}{4} \Big[(\omega_1 \times 0 + \omega_2 \times 0 + b - 0)^2 + (\omega_1 \times 1 + \omega_2 0 + b - 1)^2$$

$$+ (\omega_1 \times 0 + \omega_2 \times 1 + b - 1)^2 + (\omega_1 \times 1 + \omega_2 \times 2 + b - 0)^2 \Big]$$

Taking Partial :-

$$\frac{\delta J}{\delta \omega_1} = \frac{1}{2} (\omega_1 + b - 1 + \omega_1 + \omega_2 + b)$$

$$= \frac{1}{2} (2\omega_1 + \omega_2 + 2b - 1) = 0 \quad \text{--- } (1)$$

$$\frac{\delta J}{\delta \omega_2} = \frac{1}{2} (\omega_2 + b - 1 + \omega_1 + \omega_2 + b)$$

$$= \frac{1}{2} (\omega_1 + 2\omega_2 + 2b - 1) = 0 \quad \text{--- } (2)$$

$$\frac{\delta J}{\delta b} = \frac{1}{2} (b + \omega_1 + b - 1 + \omega_2 + b - 1 + \omega_1 + \omega_2 + b)$$

$$= \frac{1}{2} (2\omega_1 + 2\omega_2 + 4b - 2) = 0 \quad \text{--- } (3)$$

From 1 & 2 we get :-

$$2w_1 + w_2 + \cancel{2b} - \cancel{1} = w_1 + 2w_2 + \cancel{2b} - \cancel{1}$$

$$\Rightarrow 2w_1 - w_1 = 2w_2 - w_2$$

$$\Rightarrow \boxed{w_1 = w_2} \quad \text{— } 4$$

Substituting 4 in 3

$$2w_1 + 2w_2 + 4b - 2 = 0$$

$$2w_1 + 2w_1 + 4b - 2 = 0$$

$$\boxed{w_1 = \frac{1-2b}{2}} \quad \text{— } 5$$

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

Substituting 5 $\Rightarrow$ $3\left(\frac{1-2b}{2}\right) + 2b - 1 = 0$

$$\Rightarrow (1-2b)\left(\frac{3}{2} - 1\right) = 0$$

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

In equation 3 $\Rightarrow$ $2w_1 + 2w_2 + 4b - 2 = 0$

$$2w_1 + 2\left(\tfrac{1}{2}\right) - 1 = 0$$

$$w_1 = 0 \,//$$

From 4 we can conclude that $\boxed{w_1 = w_2 = 0}$

Problem 3:-

a) $$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\sigma(-x) = \frac{1}{1+e^{x}}$$

$$\sigma(x) + \sigma(-x) = \frac{1}{1+e^{-x}} + \frac{1}{1+e^{x}} = \frac{1+e^{x}+1+e^{-x}}{(1+e^{-x})(1+e^{x})}$$

$$= \frac{1+e^{x}+1+e^{-x}}{1+e^{x}+1+e^{-x}} = 1.$$

Hence Proved//

b) $$\frac{\partial}{\partial x}\sigma(x) = \frac{\partial}{\partial x}\frac{1}{1+e^{-x}} = \frac{\partial}{\partial x}(1+e^{-x})^{-1}$$

Applying the reciprocal rule:-

$$\sigma'(x) = -(1+e^{-x})^{-2}\frac{\partial}{\partial x}(1+e^{-x}) = -(1+e^{-x})^{-2}(-e^{-x})$$

$$= (1+e^{-x})^{-2}e^{-x} = \frac{e^{-x}}{(1+e^{-x})(1+e^{-x})}$$

$$= \frac{1}{1+e^{-x}}\left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right)$$

$$= \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

$$\boxed{\sigma'(x) = \sigma(x)(1-\sigma(x))}$$

Problem 4:-

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

Here $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

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

Substituting

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

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

$$\alpha \left[(w_1 S_1 + w_2 S_3) w_1 + w_2 (w_1 S_2 + w_2 S_4) \right]$$

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

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

To make it the least asymmetric the difference between $(w_1 - w_2)^2$ should be the smallest

$$\Rightarrow w_1^2 - 2w_1w_2 + w_2^2$$

By comparing,

$$s_2 + s_3 = -2$$

Here $\boxed{s_1 = s_4 = 1}$ $\boxed{s_2 = s_3 = -1}$

$$S = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Problem 5:-

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

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

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

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

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

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

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

Differentiating wrt w and σ and setting derivative to 0 to find the MLE

$$\nabla_w \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y^{(i)} - x^{(i)T} w)^2 - n \log \sigma + c \right) = 0$$

$$\sum_{i=1}^{n} x^{(i)} (y^{(i)} - x^{(i)T} w) = 0$$

$$\sum_{i=1}^{n} x^{(i)} y^{(i)} - \sum_{i=1}^{n} x^{(i)} x^{(i)T} \omega = 0$$

$$\Rightarrow \quad \omega = \left(\sum_{i=1}^{n} x^{(i)} x^{(i)T} \right)^{-1} \left(\sum_{i=1}^{n} x^{(i)} y^{(i)} \right)$$

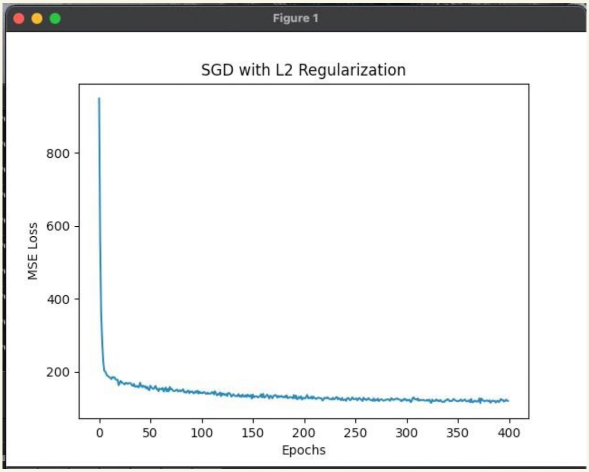
$$\nabla_\sigma \log P(D \mid \omega, \sigma^2) = \nabla_\sigma \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y^{(i)} - x^{(i)T} \omega \right)^2 - n \log \sigma + c \right) = 0$$

$$\frac{1}{\sigma^3} \sum_{i=1}^{n} \left(y^{(i)} - x^{(i)T} \omega \right)^2 - \frac{n}{\sigma} = 0$$

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

Hence Proved.

Problem 2:-



Tuned Hyperparameters:-

Learning Rate :- 0.005

Epochs :- 400

Batchsize :- 2048

Alpha :- 0.002.