ISMET OKAN CELIK CWID: 11472265

HOMEWORK-2

Question-1:

$$E(w) = MSE(w) + \frac{\lambda}{2}\sum_{i=1}^{m} w_i^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h_w(x) - y\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{m} w_i^2$$

$$= \frac{1}{m}\left(w^Tx_1 - y_1 \quad w^Tx_2 - y_2 --- w^Tx_m - y_m\right)\begin{pmatrix} w^Tx_1 - y_1 \\ w^Tx_2 - y_2 \\ \vdots \\ w^Tx_m - y_m \end{pmatrix} + \frac{\lambda}{2} w^Tw$$

we can ipnore $\frac{1}{m}$ and $\frac{1}{2}$ because they are constant values.

$$= \left(\underbrace{w^T(x_1 \quad x_2 -- x_m)} - \underbrace{(y_1 \quad y_2 - y_m)}\right)\underbrace{\begin{pmatrix} w^Tx_1 - y_1 \\ w^Tx_2 - y_2 \\ \vdots \\ w^Tx_m - y_3 \end{pmatrix}} + \lambda w^Tw$$

$$= \left(w^TX^T - Y^T\right) \cdot \left(w^TX^T - Y^T\right)^T + \lambda w^T.w$$

$$= \left(w^TX^T - Y^T\right).\left(Xw - Y\right) + \lambda w^T.w$$

$$= w^TX^TXw - w^TX^TY - Y^TXw + Y^TY + \lambda w^Tw$$

$$E(w) = w^Tw\left(X^TX + \lambda\right) - 2w^TX^TY + Y^TY$$

If we take the partial derivatice of the equation in order to optimize it

$$\frac{\partial E(w)}{\partial w} = 2w\left(X^TX + \lambda\right) - 2X^Ty = 0$$

$$= 2\left(w\left(X^TX + \lambda\right) - X^Ty\right) = 0$$

$$= w\left(X^TX + \lambda\right) = X^TY$$

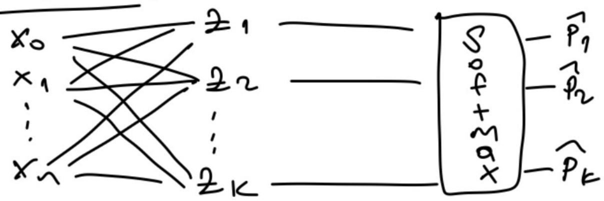
$$w = \frac{X^TY}{\left(X^TX + \lambda\right)} \rightarrow w = \left(X^TX + \lambda\right)^{-1}. X^TY$$

Propertiy of Identity Matrix I: $\boxed{I.\lambda = \lambda}$

Proof: $\boxed{w = \left(\lambda I + X^TX\right)^{-1}. X^TY}$

Question 2 :

1)



$\theta_1^T \cdot x_0 = z_1$ $\qquad$ $\theta_1^T \cdot x_1 = z_1$

$\theta_2^T \cdot x_0 = z_2$ $\qquad$ $\theta_2^T \cdot x_1 = z_2$

$\vdots$ $\qquad$ $\vdots$

$\theta_k^T \cdot x_0 = z_k$ $\qquad$ $\theta_k^T \cdot x_1 = z_k$

As we can see on above for x_0 value we need to estimate k number θ value, also it is same for x_1 value. we have k+1 number of x.

In order to learn Softmax Regression model we need to estimate k.(n+1) paramater

2) $\hat{p}_k = \sigma(s_k(x))_k = \dfrac{\exp(s_k(x))}{\sum_{j=1}^{K} \exp(s_j(x))}$, $s_k(x) = \theta_k^T \cdot x$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \cdot \log\left(\hat{p}_k^{(i)}\right)$$

$$= -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \cdot \log\left[\frac{\exp(\theta_k^T \cdot x^{(i)})}{\sum_{j=1}^{K} \exp(\theta_j^T \cdot x^{(i)})}\right]$$

$$= -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \left(\log\left(\exp(\theta_k^T \cdot x^{(i)})\right) - \log\left(\sum_{j=1}^{K} \exp(\theta_j^T \cdot x^{(i)})\right)\right)$$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \cdot \theta_k^T \cdot x^{(i)} + \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \cdot \log\left(\sum_{j=1}^{K} \exp(\theta_j^T \cdot x^{(i)})\right)$$

$$\frac{\partial J(\theta)}{\partial \theta} = -\frac{1}{m} \sum_{i=1}^{m} y_k^{(i)} \cdot x^{(i)} + \frac{1}{m} \sum_{i=1}^{m} 1 \cdot \frac{1}{\sum_{j=1}^{K} \exp(\theta_j^T \cdot x^{(i)})} \cdot \exp(\theta_k^T \cdot x^{(i)}) \cdot x^{(i)}$$

$$\nabla J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(\underbrace{\frac{\exp(\theta_k^T \cdot x^{(i)})}{\sum_{j=1}^{K} \exp(\theta_j^T \cdot x^{(i)})}} - y_k^{(i)}\right) \cdot x^{(i)}$$

$$\nabla J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{p}_k - y_k^{(i)}\right) \cdot x^{(i)}$$