Q.1 Given,

$X = (x^{(1)}, x^{(2)} \ldots, x^{(m)})^T \rightarrow$ i/p matrix

$x^{(i)} = (1, x_1, x_2 \cdots, x_n) \rightarrow i^{th}$ sample

$y = (y^{(1)}, y^{(2)} \ldots, y^{(n)}) \rightarrow$ features

$h_w(x) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n \rightarrow$ hypothesis

$y^{(j)}$ is measurement of $h_w(x)$ for j^{th} train sample

$E(w) = \Sigma_{i=1}^{m} (w^T \cdot x^{(i)} - y^{(i)})^2 + \lambda \Sigma_{i=1}^{m} w_i^2$

P.T closed form sol^n is $w = (\lambda I + X^T \cdot X)^{-1} \cdot X^T \cdot y$

input feature matrix $X \Rightarrow$ ~~n x m~~ m x n

$E(w) = (Xw - y)^T (Xw - y) + \lambda w^T w$

$= (Xw)^T (Xw) - (Xw)^T (y) - (Xw) y^T + y^T y + \lambda w^T w$

$= \cancel{X^T w^T X w}$

$\frac{dE}{dw} = 0 \Rightarrow (X^T X + \lambda I) w = X^T y$

multiplying with $(X^T X + \lambda I)^{-1}$ on both sides

$w = (X^T X + \lambda I)^{-1} X^T y$

Hence proved the closed-form solution

Q-2 a) total number of parameters

$$= n \times K$$

b) $S_K(x) = \theta_K^T \cdot x$, $\hat{p}_K = \dfrac{e^{(S_K(x))}}{\sum_{i=1}^{K} e^{(S_K(x))}}$

$$J(\theta) = \frac{-1}{M} \sum_{i=1}^{M} \sum_{i=1}^{K} y_K^{(i)} \log\left(\hat{p}_K^{(i)}\right)$$

$$= \frac{1}{M} \sum_{i=1}^{M} \sum_{K=1}^{K} y(K)^{(i)} \log\left(\frac{\exp\left(S_K(x^{(i)})\right)}{\sum_{j=1}^{K} \exp\left(S_j(x^{(i)})\right)} \right)$$

$$= \frac{-1}{M} \sum_{i=1}^{M} \left(\sum_{K=1}^{K} y_{(K)}^{(i)} \log\left(\exp\left(S_K(x^{(i)})\right)\right) - \sum_{K=1}^{K} y_{(K)}^{(i)} \log\left(\sum_{j=1}^{K} \exp S_j(x^{(i)})\right)\right)$$

$$\nabla_{\theta_K} J(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left(\hat{p}_K^{(i)} - y_{(K)}^{(i)}\right) x^{(i)}$$

$$= \frac{-1}{M} \sum_{i=1}^{M} x^{(i)} - \frac{1}{\sum_{j=1}^{K} \exp\left(\theta_j^T x(i)\right)} \cdot \exp\left(\theta^T_K x^{(i)}\right) x^i$$

$$= \frac{1}{M} \sum_{i=1}^{M} \left(1 - \hat{P}_K(i)\right) x^{(i)}$$

$$= \frac{1}{M} \sum_{i=1}^{M} \left(\hat{P}_K(i) - y_K^{(i)}\right) x^{(i)} = \nabla_{\theta_K} J(\theta)$$