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

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

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

$$\frac{\partial}{\partial w}=\frac{\partial((Xw-Y)^T(Xw-Y)+\lambda w^Tw)}{\partial w}=0$$

$$2X^T(Xw-Y)+2\lambda Iw=0$$

$$X^TXw-X^TY+\lambda Iw=0$$

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

$$w=(X^TX+\lambda X^T)^{-1}X^TY$$

2.

(1) $k \cdot n$ parameters

$$\theta = \begin{bmatrix} \theta_1^1 & \cdots & \theta_1^n \\ \cdots & \vdots & \cdots \\ \theta_k^1 & \cdots & \theta_k^n \end{bmatrix}$$

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

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m} \frac{\partial}{\partial \theta_j}\left[\sum_{i=1}^{m} \sum_{j=1}^{K} 1\{y^i = j\} \log \frac{\exp(s_k(x))}{\sum_{l=1}^{k} \exp(s_j(x))} \right]$$

$$= -\frac{1}{m} \frac{\partial}{\partial \theta_j}\left[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y^i = j\} \left(\theta_j^T x^{(i)} - \log \sum_{l=1}^{k} \exp(s_k(x))\right) \right]$$

$$= -\frac{1}{m}\left[\sum_{i=1}^{m} 1\{y^{(i)} = j\} \left(x^{(i)} - \sum_{j=1}^{k} \frac{\exp(s_k(x))}{\sum_{l=1}^{k} \exp(s_j(x))}\right) \right]$$

$$= -\frac{1}{m}\left[\sum_{i=1}^{m} x^{(i)} 1\{y^{(i)} = j\} - \sum_{j=1}^{K} 1\{y^{(i)} = j\} \sum_{j=1}^{k} \frac{\exp(s_k(x))}{\sum_{l=1}^{k} \exp(s_j(x))} \right]$$

$$= -\frac{1}{m}\left[\sum_{i=1}^{m} x^{(i)} \left(1\{y^{(i)} = j\} - \sum_{j=1}^{k} \frac{\exp(s_k(x))}{\sum_{l=1}^{k} \exp(s_j(x))}\right) \right]$$

$$= -\frac{1}{m}\left[\sum_{i=1}^{m} x^{(i)} \left(1\{y^{(i)} = j\} - P(y^{(i)} = j \mid x^{(i)}; \theta)\right) \right]$$