$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{K=1}^{K} y_K^{(i)} \log(p_K^{(i)})$$

We know,

$$p_K = \delta(s_K(x))_K = \frac{\exp(s_K(x))}{\sum_{j=1}^{K} \exp(s_j(x))}$$

Where $s_K(x) = \theta_K^T . x$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{k} y_K^{(i)} \log \left[ \frac{\exp(\theta_K^T . x)}{\sum_{j=1}^{K} \exp(\theta_j^T x)} \right]$$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[ \sum_{K=1}^{K} y_K^{(i)} \log[\exp(\theta_K^T . x^i)] - \sum_{K=1}^{K} y_K^{(i)} \log \left[ \sum_{j=1}^{K} \exp(\theta_j^T x^i) \right] \right]$$

putting $y_K^{(i)} = 1$ as $y_K = \begin{cases} 1 & \text{.... if } i^{th} \text{ instance belongs to } K \\ 0 & \text{otherwise} \end{cases}$ lecture slide 3-3 pg 16

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[ \underbrace{\sum_{K=1}^{K} \log[\exp(\theta_K^T x^i)]}_{= \theta_K^T x} - \sum_{K=1}^{K} \log \left[ \sum_{j=1}^{K} \exp(\theta_j^T x^i) \right] \right]$$

$= \theta_K^T x$ ... since $\ln(e^x) = x$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[ \sum_{K=1}^{K} \theta_K^T x^i - \sum_{K=1}^{K} \log \left[ \sum_{j=1}^{K} \exp(\theta_j^T x^{(i)}) \right] \right]$$

$$\frac{\partial J(\theta)}{\partial \theta} = -\frac{1}{m} \sum_{i=1}^{m} \left[ x^{(i)} - \frac{1}{\sum_{j=1}^{K} \exp(\theta_j^T x)} \cdot \exp(\theta_K^T x^i)\, x^{(i)} \right]$$

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

$$= -\frac{1}{m} \sum_{i=1}^{m} \left[ 1 - \frac{\exp(s_K(x))}{\sum_{j=1}^{K} \exp(s_j(x))} \right] x^{(i)}$$ ...... as $\theta_K^T x = s_K(x)$, $\theta_j^T x = s_j(x)$

$$= -\frac{1}{m} \sum_{i=1}^{m} \left[ 1 - p_k^{(j)} \right] x^{(i)} \quad \ldots\ldots \text{ as } p_k = \frac{\exp(s_k(x))}{\sum_{j=1}^{K} \exp(s_j(x))}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left[ p_k^{(j)} - 1 \right] x^{(i)}$$