Gradient Derivation

$$J = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log\left(\frac{\exp fx}{\sum_{c=1}^{K}\exp fx}\right) + \lambda\sum_{j=1}^{d} w_{kj}^2$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\left[\log(\exp fx) - \log\left(\sum_{c=1}^{K}\exp fx\right)\right] + \lambda\sum_{j=1}^{d} w_{kj}^2$$

$$= -\frac{1}{N}\sum_{k=1}^{K} y_k\left[fx - \log\left(\sum_{c=1}^{K}\exp fx\right)\right] + \lambda\sum_{j=1}^{d} w_{kj}^2$$

$$\frac{\partial J}{\partial w_k} = -\frac{1}{N}\sum_{k=1}^{K}\frac{\partial}{\partial w_k}\left[y_k\left[fx - \log\left(\sum_{c=1}^{K}\exp fx\right)\right] + \lambda w_{kj}^2\right]$$

$$= -\frac{1}{N}\sum_{k=1}^{K}\frac{\partial}{\partial w_k}(y_k)\left[fx - \log\left(\sum_{c=1}^{K}\exp fx\right)\right] + \frac{\partial}{\partial w_k}\left(fx - \log\left(\sum_{c=1}^{K}\exp fx\right)\right)\cdot y_x$$

$$+\frac{\partial}{\partial w_k}(\lambda w_k^2)$$

$$fx = (w_k^T x)$$

$$\frac{\partial}{\partial w_k} = (w_k^T x) = x \qquad \frac{\partial}{\partial w_k}\left(\log\left(\sum_{c=1}^{K}\exp(w_c^T x)\right)\right)$$

$$\frac{\partial}{\partial w_k}(\lambda w_k^2) = 2\lambda w_k \qquad = \frac{1}{\sum_{c=1}^{K}\exp(w_c^T x)}\cdot\sum_{j=1}^{K}\frac{\partial}{\partial w_k}\exp(w_j^T x)$$

$$= \exp(w_j^T x)$$

Therefore,

$$\frac{\partial J}{\partial w_k} = -\frac{1}{N}\sum_{k=1}^{K} y_k\left(x - \frac{\exp(w_k^T x)}{\exp(w_c^T x)}\right) + 2\lambda w_k$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[p(c_k|x_i) - y_{ik}\right]x_i$$