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

Rearranging the terms,

$$J = \frac{-1}{N} \sum_{i=1}^{N} \sum_{K=1}^{K} y_{iK} \left(\log \exp f_K - \log \sum_{C=1}^{K} \exp f_c\right) + \lambda \sum_{j=1}^{d} w_{Kj}^2$$

Since $\log e^a = a$

$$J = \frac{-1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} f_k + \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \log \sum_{c=1}^{k} e^{f_c} + \lambda \sum_{w=1}^{d} W_{Kj}^2$$

For gradient we need to differentiate the function J wrt to w_R

$$\frac{dJ}{dw_K} = \frac{d}{dw_R}\left[\frac{-1}{N} \sum_{i=1}^{N} \sum_{K=1}^{K} y_{iK} f_K\right] + \frac{d}{dw_R}\left[\frac{1}{N} \sum_{i=1}^{N} \sum_{K=1}^{K} \log \sum_{C=1}^{K} e^{f_c}\right]$$

$$+ \quad 2 w_R$$

$$= \frac{-1}{N} \sum_{i=1}^{N} y_{iK} f'(K) + \frac{1}{N} \sum_{i=1}^{N} \frac{e^{f_K}}{\sum_{c=1}^{K} e^{f_c}} \cdot f'(K) + 2w_K$$

Taking common terms out.

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{e^{f_k}}{\sum_{c=1}^{k} e^{f_c}} - y_{ik}\right) \cdot f'(k) + 2w_k$$

According to the defination of softmax

$$\frac{e^{f_k}}{\sum_{c=1}^{k} e^{f_c}} = P(C_i|x)$$

The above equation becomes

$$\frac{1}{N}\sum_{i=1}^{N}\left[P(C_i|x) - y_{ik}\right] f'(k) + 2w_k$$