Advanced Data Analysis
Homework Week 6

Aswin Vijay

June 5, 2023

Homework 6

Delta in the last layer of for the j^th unit in multi-class classification is,

$$\delta_j^{(L)} = \frac{\partial J_n}{\partial u_j^{(L)}}$$

$$J_n(w) = -\sum_{k=1}^{K} y_{nk} \log f_k(x_n; w)$$

$$f_k = z_k^{(L)} = \frac{exp(u_k^{(L)})}{\sum_{j=1}^{K} exp(u_j^{(L)})}$$

For K output clsses and n^{th} sample. Then J_n becomes,

$$J_n(w) = -\sum_{k=1}^{K} y_{nk} \log \frac{exp(u_k^{(L)})}{\sum_{j=1}^{K} exp(u_j^{(L)})}$$

$$J_n(w) = -\sum_{k=1}^{K} y_{nk} \log f_k$$

$$\frac{\partial J_n}{\partial u_j^{(L)}} = \frac{\partial J_n}{\partial f_k} \cdot \frac{\partial f_k}{\partial u_j^{(L)}}$$

$$\frac{\partial J_n}{\partial u_j^{(L)}} = -\sum_{k=1}^{K} \frac{y_{nk}}{f_k} \cdot \frac{\partial f_k}{\partial u_j^{(L)}}$$

We need to calculate $\frac{\partial f_k}{\partial u_j^{(L)}} = \frac{\partial z_k^{(L)}}{\partial u_j^{(L)}}$. Let us compute the following derivative,

$$\frac{\partial}{\partial u_j} \frac{exp(u_k)}{\sum_{i=1}^{K} exp(u_i)}$$

Case 1: when $k = j$

$$\frac{\partial}{\partial u_j} \frac{exp(u_j)}{\sum_{i=1}^{K} exp(u_i)} = \frac{\sum_{i=1}^{K} exp(u_i)^2 - exp(u_j)^2}{\sum_{i=1}^{K} exp(u_i)^2} = f_j(1 - f_j)$$

Case 2: when $k \neq j$

$$\frac{\partial}{\partial u_j} \frac{exp(u_k)}{\sum_{i=1}^{K} exp(u_i)} = \frac{-exp(u_j)exp(u_k)}{\sum_{i=1}^{K} exp(u_i)^2} = f_k f_j$$

Using the above we have,

$$\frac{\partial J_n}{\partial u_j^{(L)}} = -\sum_{k=1}^{K} \frac{y_{nk}}{f_k} \cdot \frac{\partial f_k}{\partial u_j^{(L)}} = -\sum_{k=1}^{K} \cdot y_n k(1 - f_j)$$

Since $y_{nk} = 1$ only when $k = j$, we can further simlify as,

$$\frac{\partial J_n}{\partial u_j^{(L)}} = f_j - y_{nj}$$

Thus proven.