Suppose the tuple $(x^{(i)}, y^{(i)})$ consists of the i^{th} training example and we define the loss for the i^{th} training example as follows:

$$L_i(\theta) = -\log\left[\frac{e^{a_{y^{(i)}}(x^{(i)})}}{\sum_j e^{a_j(x^{(i)})}}\right]$$

where,

$$a_j(x^{(i)}) = w_j^T x^{(i)}$$

and $a_{y^{(i)}}(x^{(i)})$ is the score corresponding to the correct class sample of $x^{(i)}$.

From Discussion 2, Problem 4 we know, for

$y^{(i)} = j$:

$$\frac{\partial L_i(\theta)}{\partial w_j} = \left[\frac{e^{w_{y^{(i)}}^T x^{(i)}}}{\sum_{p=1}^{K} e^{w_p^T x^{(i)}}} - 1 \right] x^{(i)}$$

$y^{(i)} \neq j$:

$$\frac{\partial L_i(\theta)}{\partial w_j} = \left[\frac{e^{w_j^T x^{(i)}}}{\sum_{p=1}^{K} e^{w_p^T x^{(i)}}} \right] x^{(i)}$$

Toy example:

Suppose we have 3 classes.

$$C = \{1, 2, 3\}$$

and suppose we have a training sample $(x^{(i)}, 2)$. Then

$$\frac{\partial L_i}{\partial w_2} = \left[\frac{e^{w_2^T x^{(i)}}}{\sum_{j=1}^{3} e^{w_j^T x^{(i)}}} - 1 \right] x^{(i)}$$

$$\frac{\partial L_i}{\partial w_1} = \left[\frac{e^{w_1^T x^{(i)}}}{\sum_{j=1}^{3} e^{w_j^T x^{(i)}}} \right] x^{(i)}$$

$$\frac{\partial L_i}{\partial w_3} = \left[\frac{e^{w_3^T x^{(i)}}}{\sum_{j=1}^{3} e^{w_j^T x^{(i)}}} \right] x^{(i)}$$