Group 7

November 6, 2024

Derivative

Calculations:

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{a=1}^{K} e^{z_a}}. \tag{1}$$

$$L(z) = -\log(\text{softmax}(z)_j) = -\log\left(\frac{e^{z_j}}{\sum_{a=1}^{K} e^{z_a}}\right). \tag{2}$$

$$L(z) = -z_j + \log\left(\sum_{a=1}^{K} e^{z_a}\right). \tag{3}$$

$$\frac{\partial L}{\partial z_i}. \tag{4}$$

$$\frac{\partial}{\partial z_i}(-z_j) = -\delta_{i,j}, \tag{5}$$

$$\frac{\partial}{\partial z_i}\log\left(\sum_{a=1}^{K} e^{z_a}\right) = \frac{1}{\sum_{a=1}^{K} e^{z_a}} \cdot \frac{\partial}{\partial z_i}\left(\sum_{a=1}^{K} e^{z_a}\right). \tag{6}$$

$$\frac{\partial}{\partial z_i}\log\left(\sum_{a=1}^{K} e^{z_a}\right) = \frac{e^{z_i}}{\sum_{a=1}^{K} e^{z_a}} = \text{softmax}(z)_i. \tag{7}$$

$$\frac{\partial L}{\partial z_i} = -\delta_{i,j} + \text{softmax}(z)_i. \tag{8}$$

$$\frac{\partial L}{\partial z_i} = -\delta_{i,j} + \frac{e^{z_i}}{\sum_{a=1}^{K} e^{z_a}} = -\delta_{i,j} + \text{softmax}(z_i) \tag{9}$$