Softmax 函数

$$p_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}$$

前向传播

直接计算 e^{z_i} 可能导致数值溢出，特别是当 z_i 特别大的时候。因此，一般 Softmax 使用如下的计算公式:

$$p_k = \frac{e^{z_k - z_{max}}}{\sum_{j=1}^{K} e^{z_j - z_{max}}}$$

其中

$$z_{max} = \max_j z_j$$

反向传播公式推导

因为 p_i 的分母是对特征向量求和，因而 $\frac{\partial L}{\partial z_i}$ 需要用到链式法则:

$$\frac{\partial L}{\partial z_i} = \sum_{j=1}^{K} \frac{\partial L}{\partial p_j} \frac{\partial p_j}{\partial z_i}$$

设:

$$S = \sum_{j=1}^{K} e^{z_j - z_{max}}$$

先计算 $\frac{\partial p_j}{\partial z_i}$

$$\frac{\partial p_j}{\partial z_i} = \frac{\frac{\partial e^{z_j - z_{max}}}{\partial z_i} S - e^{z_j - z_{max}} e^{z_i - z_{max}}}{S^2}$$

即:

$$\frac{\partial p_j}{\partial z_i} = p_j \delta_{ij} - p_j p_i$$

则:

$$\frac{\partial L}{\partial z_i} = \sum_{j=1}^{K} \frac{\partial L}{\partial p_j} \left(p_j \delta_{ij} - p_j p_i \right)$$

$$\frac{\partial L}{\partial z_i} = \sum_{j=1}^{K} \frac{\partial L}{\partial p_j} p_j \delta_{ij} - p_i \sum_{j=1}^{K} \frac{\partial L}{\partial p_j} p_j = \frac{\partial L}{\partial p_i} p_i - p_i \sum_{j=1}^{K} \frac{\partial L}{\partial p_j} p_j$$

最终的公式如下：

$$\frac{\partial L}{\partial z_i} = p_i \left(\frac{\partial L}{\partial p_i} - \sum_{j=1}^{K} \frac{\partial L}{\partial p_j} p_j \right)$$