Loss - Function — Forward

Softmax function

- to apply ~~ith~~ ~~one~~ softmax on i^{th} entry

$$q_i = \frac{e^{\hat{y}_i}}{\sum_j e^{\hat{y}_j}} \qquad \hat{y} : \text{network output}$$

substract by largest entry

- $$q_i = \frac{e^{\hat{y}_i - \max_j \hat{y}_j}}{\sum_K e^{\hat{y}_K - \max_j \hat{y}_j}} = \frac{e^{(-\max_j \hat{y}_j) \cdot \hat{y}_i}}{\sum_K e^{(-\max_j \hat{y}_j) \cdot \hat{y}_K}} = \frac{e^{-\max_j \hat{y}_j} \cdot e^{\hat{y}_i}}{e^{-\max_j \hat{y}_j} \cdot \sum_K e^{\hat{y}_K}} = \frac{e^{\hat{y}_i}}{\sum_K e^{\hat{y}_K}}$$

Loss Function

cross Entry Loss : measures the differences between two probability distributions.

The Loss for a single sample :

$$L(p,q) = -\sum_{K=1}^{K} p_K \log(q_K) = -\sum_{K=1}^{K} \log(q_K) \quad \text{where } p_K = 1$$

because we use one-hot Codierung, p_K is either 0 or 1

p_K : the K^{th} element of p

p : the true label with p_K

q_K : the K^{th} element of q

q : the network output with the softmax applied

K : the number of elements in p and q

Apply the loss function to multiple samples

$$\text{Loss} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{K=1}^{K} p_{n,K} \log(q_{n,K})$$

concret example :

$K=3$, $p = [0,1,0]$, network output : $[1.2, 2.3, 0.8]$

Step 1 : calculate q_i

$$q_K = \frac{e^{z_K}}{\sum_j e^{z_j}}$$

$$\sum_j e^{z_j} = e^{1.2} + e^{2.3} + e^{2.23} = 15.52$$

$$q = \left[\frac{e^{1.2}}{15.52}, \frac{e^{2.3}}{15.52}, \frac{e^{2.23}}{15.52}\right] = [0.214, 0.642, 0.144]$$

step: calculate cross entroy loss

$$L(p,q) = -p_1 \log(q_1) + [-p_2 \log(q_2)] + [-p_3 \log(q_3)]$$

$$= -p_2 \log(q_2) = -\log(0.642) = 0.443$$

* $p[0] = p_1 = 0$, $p[2] = p_3 = 0$

because $q_i = \dfrac{e^{\hat{y}_i - \max_j \hat{y}_j}}{\sum_K e^{\hat{y}_K - \max_j \hat{y}_j}}$

$L(p,q) = -\sum_{K=1}^{K} P_K \log(q_K) = -\sum_{K=1}^{K} \log(q_K)$, ~~ka~~ where $P_K = 1$

$= -\log(q_y)$, because we use one hot codieng.

~~$\Rightarrow L(p,q) =$~~

$$\Rightarrow L(p,q) = -\log\left(\frac{e^{\hat{y}_i - \max_j \hat{y}_j}}{\sum_K e^{\hat{y}_K - \max_j \hat{y}_j}}\right)$$

$$= -(\hat{y}_i - \max_j \hat{y}_j) + \log\left(\sum_{K=1}^{K} e^{\hat{y}_K - \max_j \hat{y}_j}\right)$$

$$\text{Loss} = -\frac{1}{N} \sum_{n=1}^{N} \log(q_{n,K})$$

$$= -\frac{1}{N} \sum_{n=1}^{N} \cdot \left(-(\hat{y}_i - \max_j \hat{y}_j) + \log\left(\sum_{K=1}^{K} e^{\hat{y}_K - \max_j \hat{y}_j}\right)\right)$$