Subject :

Year : Month : Data : ( )

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}} = P(y_i|x)$$

فرض استقلال:

$$L(\theta) = \prod_{i=1}^{N} P(y^{(i)}|x^{(i)})$$

برای سادگی عبارت

$$\log L(\theta) = \sum_{i=1}^{N} \log P(y^{(i)}|x^{(i)}) = \sum_{i=1}^{N} \log P\left(\frac{e^{z_y}}{\sum_{j=1}^{c} e^{z_j}}\right)$$

به جای به حداکثر رساندن عبارت بالا منفی این عبارت را مینیمم می‌کنیم

$$\xi = Loss = -\log P(y|x)$$

در نهایت به صورت وان هات شده (one hot) داریم:

$$\xi = \text{Cross Entropy loss} = -\sum_{i=1}^{c} y_i \log(\hat{y}_i)$$

مشتق تابع softmax:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}}$$

① $i=j \rightarrow \frac{\partial y_i}{\partial z_i} = \frac{e^{z_i}\Sigma_c - e^{z_i}e^{z_i}}{\Sigma_c^2} \xrightarrow{\text{در نهایت}} y_i(1-y_i)$ شبیه لاجستیک فانکشن شد!

② $i \neq j \rightarrow \frac{\partial y_i}{\partial z_j} = \frac{0 - e^{z_i}e^{z_j}}{\Sigma_c^2} = -y_i y_j$

$$\frac{\partial \xi}{\partial z_i} = -\sum_{j=1}^{c} y_j \frac{\partial \log(\hat{y}_j)}{\partial z_i} \xrightarrow{①\&②} \frac{\partial \xi}{\partial z_i} = -\left(\frac{y_i}{\hat{y}_i}\hat{y}_i(1-\hat{y}_i) + \sum_{j\neq i}\frac{y_j}{\hat{y}_j}(-\hat{y}_i\hat{y}_j)\right)$$

$$= -y_i(1-\hat{y}_i) + \sum_{i\neq j} y_j\hat{y}_i = -y_i + \hat{y}_i\sum_{j=1}^{c} y_j = \hat{y}_i - y_i$$

$$w_{i+1} \leftarrow w_i - \eta \underbrace{\frac{\partial Loss}{\partial w_i}}_{(\hat{y}_i - y_i)x}$$

اگر $y_i > \hat{y}_i$ ، w_{i+1} کاهش و برعکس