

SOFTMAX

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{k'}^K e^{z_{k'}}}, \text{ where } z_k = w_k^T \cdot x$$

$$= \sum w_{ki} \cdot x_i$$

$\frac{\partial \hat{y}_k}{\partial z_k}$ depends on $k=k'$

$$\Sigma = \sum_{k'}^K e^{z_{k'}}$$

(1) if $k=k'$

$$\frac{\partial \hat{y}_k}{\partial z_k} = \frac{e^{z_k} \Sigma - e^{z_k} e^{z_k}}{\Sigma^2} = \frac{e^{z_k} (\Sigma - e^{z_k})}{\Sigma^2}$$

$$= \hat{y}_k^n (1 - \hat{y}_k^n)$$

(2) if $k \neq k'$

$$\frac{\partial \hat{y}_k^n}{\partial z_k} = \frac{0 - e^{z_k} e^{z_{k'}}}{\Sigma^2} = -\hat{y}_k^n \hat{y}_{k'}^n$$

Cross Entropy Loss

$$C^n(w) = - \sum_{k=1}^K y_k^n \ln(\hat{y}_k^n)$$

$$\frac{\partial C^n}{\partial z_k} = - \sum y_k^n \frac{1}{\hat{y}_k^n} \frac{\partial \hat{y}_k^n}{\partial z_k}$$