NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 24: Cross Entropy Loss

CONCEPTS COVERED

Concepts Covered:

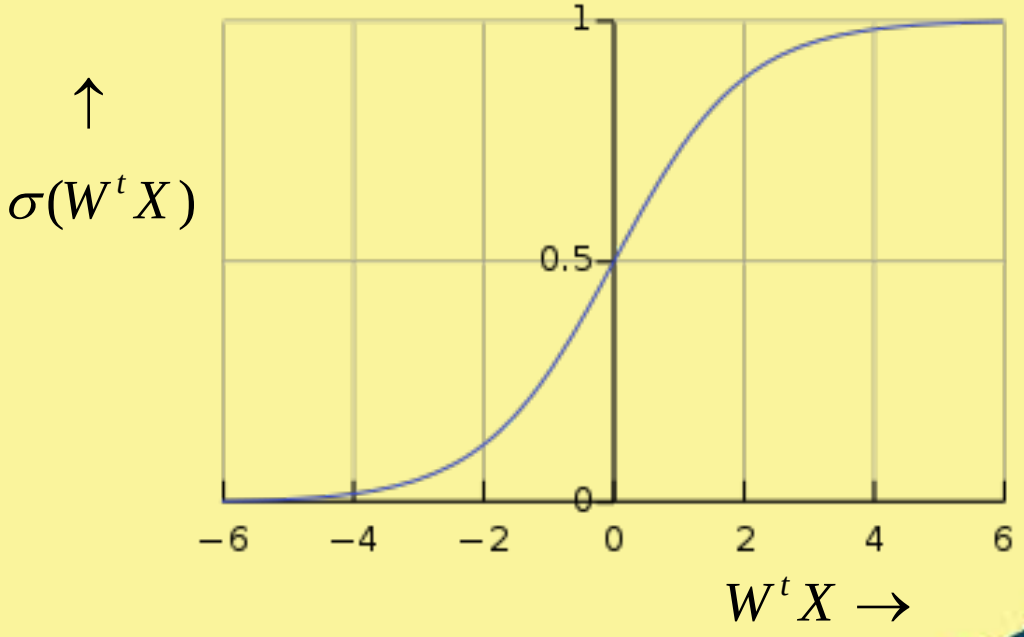
- Back Propagation Learning in MLP
  - Squared Error
- Cross Entropy Loss

Problem with Quadratic Loss Function

$$E = \frac{1}{2}\sum_{j=1}^{M_K}\left(O_j^K - t_j\right)^2$$

$$W_{ij}^K \leftarrow W_{ij}^K - \eta\delta_j^K O_i^{K-1}$$

$$\delta_j^K = O_j^K(1-O_j^K)(O_j^K - t_j)$$

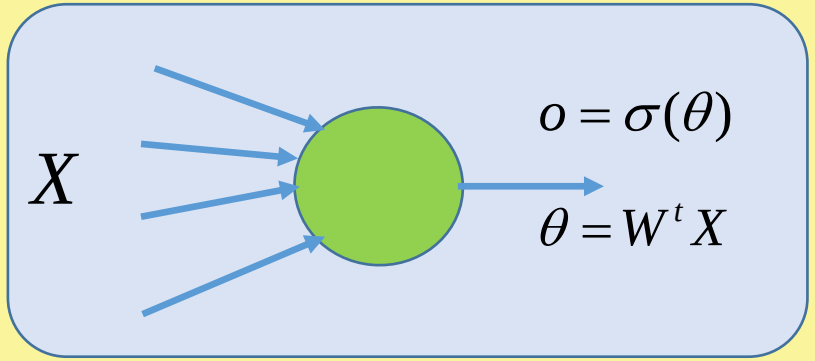


Cross Entropy Loss

Cross Entropy Loss- Two Class Problem

$o \Rightarrow$ likelihood that y *is* 1

$(1-o) \Rightarrow$ likelihood that y *is* 0



Likelihood that is to be maximized $\Rightarrow o^y (1-o)^{(1-y)}$

Loglikelihood $\Rightarrow y \log o + (1-y) \log(1-o)$

Cross Entropy Loss

$$\text{Minimize} \Rightarrow C = -\frac{1}{N}\sum_{\forall X}\left[y \log o + (1-y)\log(1-o)\right]$$

$$\frac{\partial C}{\partial W_i} = -\frac{1}{N}\sum_{\forall X}\left[\frac{y}{\sigma(\theta)} - \frac{(1-y)}{1-\sigma(\theta)}\right]\frac{\partial \sigma(\theta)}{\partial W_i}$$

$$= -\frac{1}{N}\sum_{\forall X}\left[\frac{y}{\sigma(\theta)} - \frac{(1-y)}{1-\sigma(\theta)}\right]\frac{\partial \sigma(\theta)}{\partial \theta}\cdot\frac{\partial \theta}{\partial W_i}$$

Cross Entropy Loss

$$\frac{\partial C}{\partial W_i} = -\frac{1}{N}\sum_{\forall X}\left[\frac{y}{\sigma(\theta)} - \frac{(1-y)}{1-\sigma(\theta)}\right]\frac{\partial \sigma(\theta)}{\partial \theta}\cdot\frac{\partial \theta}{\partial W_i}$$

$$= -\frac{1}{N}\sum_{\forall X}\left[\frac{y}{\sigma(\theta)} - \frac{(1-y)}{1-\sigma(\theta)}\right]\frac{\partial \sigma(\theta)}{\partial \theta}\cdot\frac{\partial \theta}{\partial W_i}$$

$$= -\frac{1}{N}\sum_{\forall X}\left[\frac{y-\sigma(\theta)}{\sigma(\theta)(1-\sigma(\theta))}\right]\sigma(\theta)(1-\sigma(\theta).x_i$$

$$= \frac{1}{N}\sum_{\forall X} x_i(\sigma(\theta)-y)$$

$$= \frac{1}{N}\sum_{\forall X} x_i(o-y)$$

Cross Entropy Loss- Multiclass Problem

$$C = -\frac{1}{N}\sum_{\forall X}\sum_{j}\left[y_j \log o_j^K + (1-y_i)\log(1-o_j^K)\right]$$

$$\frac{\partial C}{\partial W_{ij}^K} = \frac{1}{N}\sum_{\forall X} o_i^{K-1}(o_j^K - y_j)$$

$$W_{ij}^K \leftarrow W_{ij}^K - \eta\frac{1}{N}\sum_{\forall X} o_i^{K-1}(o_j^K - y_j)$$

NPTEL ONLINE CERTIFICATION COURSES

Thank you