NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas

Department : E & ECE, IIT Kharagpur

Topic

Lecture 26: Back propagation Learning – Examples II

CONCEPTS COVERED

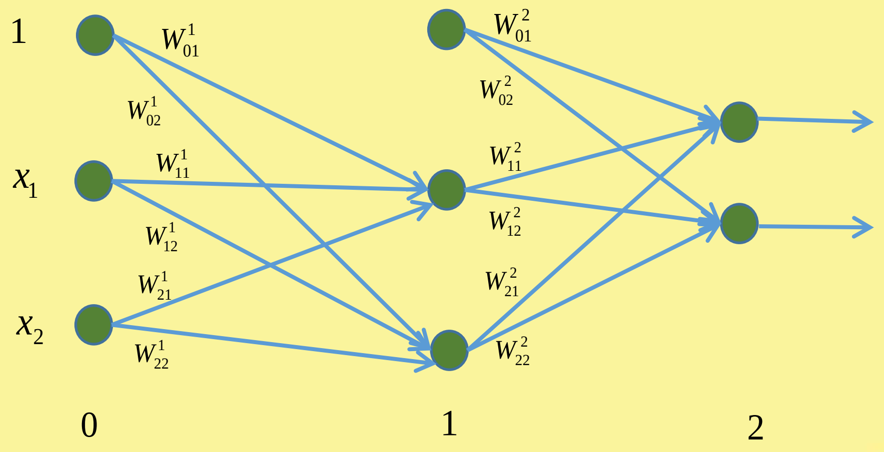
Concepts Covered:

- ❑ Back Propagation Learning in MLP
- ❑ Different Loss Functions
- ❑ Back Propagation Learning - Example
- ❑ Back Propagation – Node Level

Back Propagation Learning
an
Example

Back Propagation Learning:- Output Layer

$$E = \frac{1}{2}\sum_{j=1}^{2}\left(x_j^2 - t_j\right)^2 \qquad x_j^2 = \frac{1}{1+e^{-\theta_j^2}} \qquad \theta_j^2 = \sum_{i=0}^{2} W_{ij}^2 x_i^1$$



$$\frac{\partial E}{\partial W_{ij}^2} = \frac{\partial E}{\partial x_j^2}\cdot\frac{\partial x_j^2}{\partial \theta_j^2}\cdot\frac{\partial \theta_j^2}{\partial W_{ij}^2} = (x_j^2 - t_j)x_j^2(1 - x_j^2)x_i^1$$

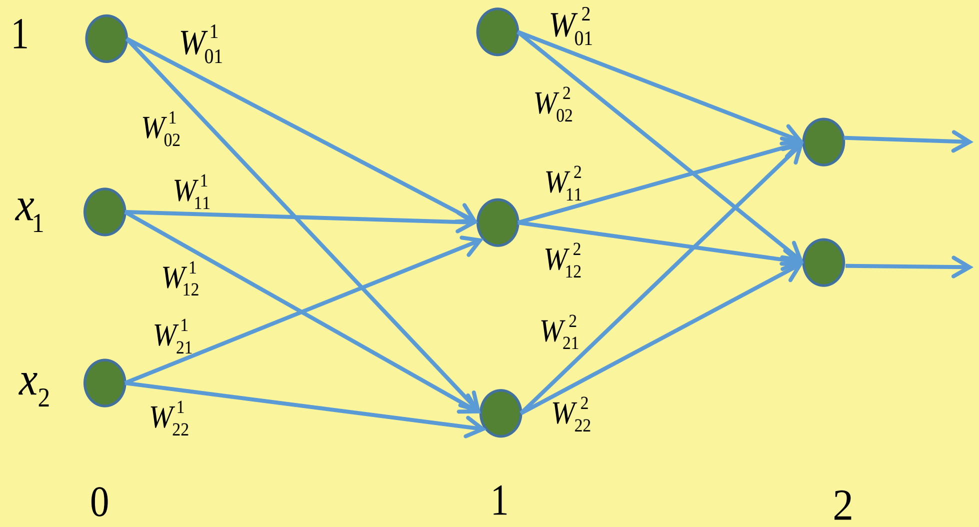
We set $\boxed{\delta_j^2 = x_j^2(1 - x_j^2)(x_j^2 - t_j)} \Rightarrow \frac{\partial E}{\partial W_{ij}^2} = \delta_j^2 x_i^1$

$$\boxed{W_{ij}^2 \leftarrow W_{ij}^2 - \eta\frac{\partial E}{\partial W_{ij}^2}}$$

Feed Forward Pass

$$W^1 \qquad x_i^0 \qquad \theta_j^1 = \sum W_{ij}^1 x_i^0 \qquad x_j^1 = \frac{1}{1+e^{-\theta_j^1}}$$

$$\begin{bmatrix} 0.5 & 1.5 & 0.8 \\ 0.8 & 0.2 & -1.6 \end{bmatrix} \begin{bmatrix} 1 \\ 0.7 \\ 1.2 \end{bmatrix} = \begin{bmatrix} 2.51 \\ -9.8 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.92 \\ 0.27 \end{bmatrix}$$



$$W^2 \qquad x_i^1 \qquad \theta_j^2 = \sum_i W_{ij}^2 x_i^1 \qquad x_j^2 = \frac{1}{1+e^{-\theta_j^2}}$$

$$\begin{bmatrix} 0.9 & -1.7 & 1.6 \\ 1.2 & 2.1 & -1.-0.26 \end{bmatrix} \begin{bmatrix} 1 \\ 0.92 \\ 0.27 \end{bmatrix} = \begin{bmatrix} -0.232 \\ 3.057 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.44 \\ 0.95 \end{bmatrix}$$

Back Propagation Learning:- Output Layer

$$\boxed{\delta_j^2 = x_j^2(1-x_j^2)(x_j^2-t_j)}$$

$$\delta_1^2 = x_1^2(1-x_1^2)(x_1^2-t_1)$$
$$= 0.44*(1-0.44)*(0.44-1)$$
$$= -0.138$$
$$\Rightarrow \frac{\partial E}{\partial W_{11}^2} = \delta_1^2 x_1^1 = -0.126$$

$$\delta_2^2 = x_2^2(1-x_2^2)(x_2^2-t_2)$$
$$= 0.95*(1-0.95)*(0.95-0)$$
$$= 0.045$$
$$\Rightarrow \frac{\partial E}{\partial W_{12}^2} = \delta_2^2 x_1^1 = 0.04$$

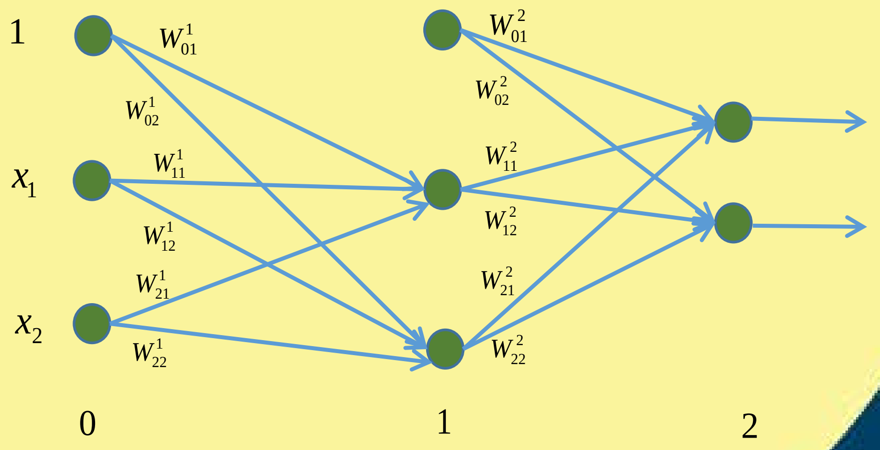
$$\boxed{W_{11}^2 \leftarrow W_{11}^2 + \eta * 0.126}$$

$$\boxed{W_{12}^2 \leftarrow W_{12}^2 - \eta * 0.04}$$

Back Propagation Learning:- Output Layer

$$\frac{\partial E}{\partial W_{21}^{2}} = \delta_1^2 x_2^1 = -0.037 \qquad \frac{\partial E}{\partial W_{22}^{2}} = \delta_2^2 x_2^1 = 0.012$$

$$\frac{\partial E}{\partial W_{01}^{2}} = \delta_1^2 x_0^1 = -1.38 \qquad \frac{\partial E}{\partial W_{02}^{2}} = \delta_2^2 x_0^1 = 0.045$$



Back Propagation Learning:- Hidden Layer

We set $\boxed{\delta_i^k = O_i^k(1-O_i^k)\sum_{j=1}^{M_{k+1}} \partial_j^{k+1} W_{ij}^{k+1}} \Rightarrow \delta_i^1 = x_i^1(1-x_i^1)\sum_{j=1}^{2} \partial_j^2 W_{ij}^2 \Rightarrow \boxed{\frac{\partial E}{\partial W_{ij}^k} = \delta_i^k x_i^{k-1}}$

$$\delta_1^1 = x_1^1(1-x_1^1)\left[\delta_1^2 * W_{11}^2 + \delta_2^2 W_{12}^2\right]$$
$$= 0.92*(1-0.92)[(-0.137)*(-1.7)+0.045*2.1]$$
$$= 0.024$$

$$\delta_2^1 = x_2^1(1-x_2^1)\left[\delta_1^2 * W_{21}^2 + \delta_2^2 W_{22}^2\right]$$
$$= 0.27*(1-0.27)[(-0.137)*0.8+0.045*(-0.2)]$$
$$= -0.02$$

Back Propagation Learning:- Hidden Layer

$$\frac{\partial E}{\partial W_{11}^{1}} = \delta_1^1 * x_1^0 = 0.024 * 0.7 = 0.017$$

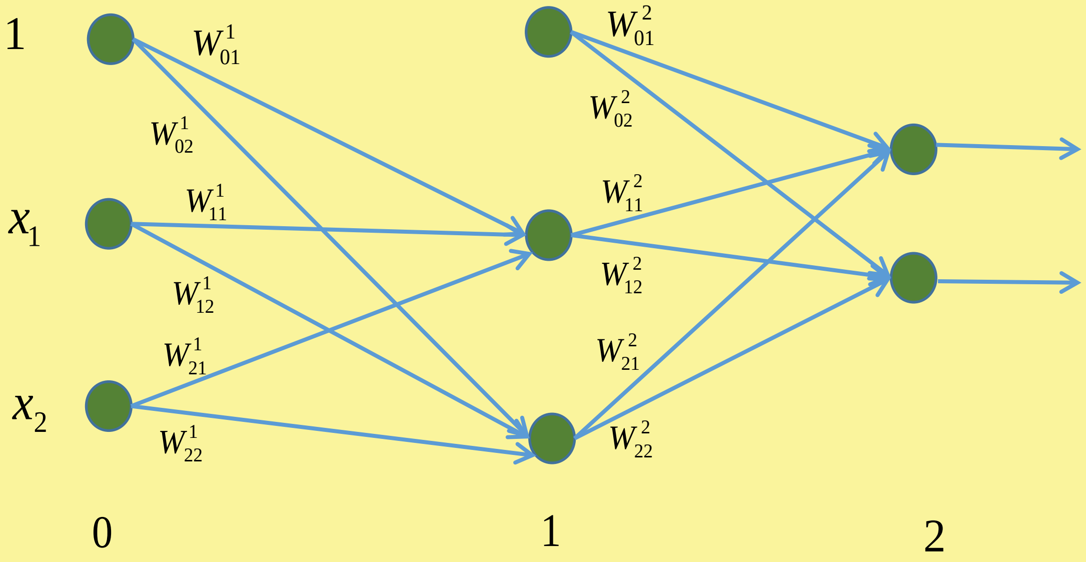
$$\frac{\partial E}{\partial W_{21}^{1}} = \delta_2^1 * x_1^0 = -0.02 * 0.7 = -0.014$$

$$\frac{\partial E}{\partial W_{21}^{1}} = \delta_1^1 * x_2^0 = 0.024 * 1.2 = 0.0288$$

$$\frac{\partial E}{\partial W_{22}^{1}} = \delta_2^1 * x_2^0 = -0.02 * 1.2 = -0.024$$

$$\frac{\partial E}{\partial W_{01}^{1}} = \delta_1^1 * x_0^0 = 0.024 * 1 = 0.024$$

$$\frac{\partial E}{\partial W_{02}^{1}} = \delta_2^1 * x_0^0 = -0.02 * 1 = -0.02$$



$$W_{ij}^1 \leftarrow W_{ij}^1 - \eta \frac{\partial E}{\partial W_{ij}^1}$$

NPTEL ONLINE CERTIFICATION COURSES

Thank you