CS 6375
ASSIGNMENT 3
Neural Network

Names of students in your group:
1. Akhila Kancharana (axk180025)

CS 6375.502

Number of free late days used: 0

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

12/10/18

Assignment - Neural Network

1.1 Back propagation:

(a) $\tanh(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

$$\frac{d}{dx}(\tanh(x)) = \frac{(e^x+e^{-x})^2 - (e^x-e^{-x})^2}{(e^x+e^{-x})^2}$$

$$= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2$$

$$\boxed{\frac{d}{dx}(\tanh(x)) = 1 - \tanh^2(x)} \quad — \ 1$$

$$E_d = \frac{1}{2} \sum_{k \in o/p} (t_k - o_k)^2$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} * \frac{\partial net_j}{\partial w_{ji}}$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \times x_{ji} \quad — \ 2$$

case I: when j is an output unit

$$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j} \times \frac{\partial o_j}{\partial net_j}$$

$$\frac{\partial E_d}{\partial o_j} = \frac{\partial}{\partial o_j}\left[\frac{1}{2}\sum_{k \in o/p} (t_k - o_k)^2\right]$$

$$\frac{\partial E_d}{\partial o_j} = \frac{2(t_j - o_j)(-1)}{2} = (o_j - t_j) \quad — \ 3$$

using ①

$$\frac{\partial o_j}{\partial net_j} = 1 - o_j^2$$

so.

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} x_{ji}$$

$$= -(t_j - o_j)(1 - o_j^2) x_{ji}$$

$$\Delta w_{ji} = \eta (o_j - t_j)(1 - o_j^2) x_{ji}$$

$$\boxed{\Delta w_{ji} = \eta \delta_j x_{ji}}$$

case 2 : j is a hidden unit

$$\frac{\partial E_d}{\partial net_j} = \sum_{k \in} \frac{\partial E_d}{\partial net_k} \cdot \frac{\partial net_k}{\partial net_j}$$

$$\frac{\partial E_d}{\partial net_j} = \sum_k - \delta_k \frac{\partial net_k}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j}$$

$$\sum_k - \delta_k w_{kj} (1 - o_j^2)$$

$$= (o_j^2 - 1) \sum_k \delta_k w_{kj}$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$

$$= -\eta \left[(o_j^2 - 1) \sum \delta_k w_{kj} \right] x_{ji}$$

$$= \eta (1 - o_j^2) \sum \delta_k w_{kj} (x_{ji})$$

$$= \eta \delta_j x_{ji}$$

Scanned by CamScanner

$\therefore$ when j is output layer

$$\delta_j = (t_j - o_j)(1 - o_j^2)^2$$

when j is hidden

$$\delta_j = (1 - o_j^2) \sum \delta_k w_{kj}$$

weight update rule

$$w_{ij}^{new} = w_{ij}^{old} + \Delta w_{ij}$$

$$\Delta w_{ij} = \eta \delta_j x_{ij}$$

b) $ReLu(x) = max(0, x)$

$$\frac{d}{dx} ReLu(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E_d = \frac{1}{2} \sum_k (t_k - o_k)^2$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\delta w_{ji}}$$

$$\frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \cdot x_{ji}$$

case I : when j is output unit

$$\frac{\partial E_d}{\partial net_j} = \frac{\partial E_d}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j}$$

$$\frac{\partial E_d}{\partial o_j} = \frac{\partial}{\partial o_j} \left[\frac{1}{2} \sum_k (t_k - o_k) \right]$$

$$= (o_k - t_k)$$

Scanned by CamScanner

$$\begin{cases} \dfrac{\partial o_j}{\partial net_j} = 1 & \text{for } net_j > 0 \\ \dfrac{\partial o_j}{\partial net_j} = 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial E_d}{\partial net_j} = -(t_j - o_j) \quad \text{for } net_j > 0$$

$$= 0 \quad \text{for } net_j \leq 0$$

$$\frac{\partial E_d}{\partial w_{ji}} = (o_j - t_j) x_{ji} \quad \text{for } net_j > 0$$

$$= 0 \quad \text{for } net_j \leq 0$$

$$\cancel{\frac{\partial E_d}{\partial w_{ji}} = (t_j - o_j) x_{ji} \text{ for net}}$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$

$$\Delta w_{ji} = \eta (t_j - o_j) x_{ji} \quad \text{for } net_j > 0$$

$$= 0 \quad \text{for } net_j \leq 0$$

$$\delta_j = (t_j - o_j)$$

case II : when j is hidden unit

$$\frac{\partial E_d}{\partial net_j} = \sum_k \frac{\partial E_d}{\partial net_k} \cdot \frac{\partial net_k}{\partial net_j}$$

$$\frac{\partial E_d}{\partial net_j} = \sum_k -\delta_k \frac{\partial net_k}{\partial net_j}$$

$$= \sum_k -\delta_k \frac{\partial net_k}{\partial o_j} \frac{\partial o_j}{\partial net_j}$$

$$\boxed{\frac{\partial E_d}{\partial net_j} = \Sigma - \delta_k w_{kj}}$$ for $net_j > 0$

$$\frac{\partial E_d}{\partial net_j} = 0 \text{ for } net_j \leq 0$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$

$$= -\eta \left[\Sigma - \delta_k w_{kj}\right] x_{ji} \text{ for } net_j > 0$$

$$\Delta w_{ji} = 0 \text{ for } net_j \leq 0$$

when j is o/p unit

$$\delta_j = (t_j - o_j) \quad \text{for } net_j > 0$$

$$0 \quad \text{for } net_j \leq 0$$

when j is hidden unit

$$\delta_j = \Sigma\, \delta_k w_{kj} \quad \text{for } net_j > 0$$

$$0 \quad \text{for } net_j \leq 0$$

$$w_{ij}^{new} = w_{ij}^{old} + \eta \Delta w_{ij}$$

$$\Delta w_{ij} = \eta\, \delta_j\, x_{ij}$$

2. $o = w_0 + w_1(x_1 + x_1^2) + \cdots + w_n(x_n + x_n^2)$

activation function $f(x) = x$

learning rate $= \eta$

bias weight $= w_0$

$$E_d = \frac{1}{2} \sum_{k \in o/p} (t_k - o_k)^2$$

$$\Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}}$$

$$\frac{\partial net_j}{\partial w_{ji}} = (x_j + x_j^2)$$

$$= \frac{\partial E_d}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ji}}$$

$$= \frac{\partial E_d}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j}$$

$\because$ Identity activation function is used,

$$\frac{\partial o_j}{\partial net_j} = 1$$

So, $$\frac{\partial E_d}{\partial o_j} = \frac{1}{2}(2)(t_j - o_j)(-1)$$

$$\boxed{\frac{\partial E_d}{\partial o_j} = (o_j - t_j)}$$

$$\Delta w_{ji} = -\eta\left((o_j - t_j)(x_j + x_j^2)\right)$$

$$= (t_j - o_j)(x_j + x_j^2)$$

$\Rightarrow w_{ij}^{new} = w_{ij}^{old} + \Delta w_{ij}$

3.

(a) node net output

1 $net_1 = x_1$ $O_1 = x_1$

2 $net_2 = x_2$ $O_2 = x_2$

3 $net_3 = x_1 w_{31} + x_2 w_{32}$ $O_3 = h(net_3)$

4 $net_4 = x_1 w_{41} + w_{42} x_2$ $O_4 = h(net_4)$

5 $net_5 = w_{53} h(net_3) + w_{54} h(net_4)$ $O_5 = net_5$

$$Y_5 = (net_5)$$
$$= w_{53} h(net_3) + w_{54} h(net_4)$$
$$= w_{53} h(x_1 w_{31} + x_2 w_{32}) + w_{54} h(x_1 w_{41} + x_2 w_{42})$$

(b) $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $w^1 = \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix}$

$w^2 = (w_{53} \quad w_{54})$

$$\begin{pmatrix} net_3 \\ net_4 \end{pmatrix} = w^1 X$$
$$= \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} w_{31} x_1 & w_{32} x_2 \\ w_{41} x_1 & w_{42} x_2 \end{bmatrix}$$

$$\begin{bmatrix} O_3 \\ O_4 \end{bmatrix} = \begin{bmatrix} h(w_{31} x_1 + w_{32} x_2) \\ h(w_{41} x_1 + w_{42} x_2) \end{bmatrix}$$

$$Y_5 = net_5 = w^2 \begin{bmatrix} O_3 \\ O_4 \end{bmatrix} = w^2 h(w^1 X)$$
$$= (w_{53} \quad w_{54}) \begin{bmatrix} h(w_{31} x_1 + w_{32} x_2) \\ h(w_{41} x_1 + w_{42} x_2) \end{bmatrix}$$

$$= w_{53} h(w_{31}x_1 + w_{32}x_2) + w_{54} h(w_{41}x_1 + w_{42}x_2)$$

$$Y_5 = w^2 h(w^1 x)$$

(c) sigmoid = $h_s(x) = \frac{1}{1+e^{-x}}$

$$\tanh = h_t(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

simplifying tanh

$$h_t(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

$$= \frac{1}{1+e^{-2x}} - \frac{e^{-2x}}{1+e^{-2x}} \quad \text{— (1)}$$

using property of sigmoid function

$$1 - h_s(x) = 1 - \frac{1}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}}$$

$$(1) \Rightarrow h_t(x) = h_s(2x) - [1 - h_s(2x)]$$

$$h_t(x) = 2h_s(2x) - 1$$

By observation, tan h function is a scaled version of sigmoid, hence both will generate same function.

u) considering the activation function to be sigmoid

$$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in o/p} (t_{kd} - o_{kd})^2 + \gamma \sum_{ij} w_{ji}^2$$

weight update rule:

$$w_{ji} = w_{ji} + \Delta w_{ji} \; ; \quad \Delta w_{ji} = -\eta \frac{\partial E(w)}{\partial w_{ji}}$$

$$\frac{\partial E(w)}{\partial w_{ji}} = \left[\frac{1}{2} \sum_{d \in D} \sum_{k} (t_{kd} - o_{kd})^2 + \gamma \sum_{ij} w_{ji}^2 \right]$$

$$= \frac{1}{2} \cdot 2(-1)(t_j - o_j)(o_j)(1 - o_j) x_{ji} + 2\gamma w_{ji}$$

$$\Delta w_{ji} = -\eta \left[(o_j - t_j) o_j (1 - o_j) x_{ji} + 2\gamma w_{ji} \right]$$

$$w_{ji} = w_{ji} + \eta \delta_j x_{ji} - 2\gamma\eta w_{ji}$$

$$w_{ji} = w_{ji} (1 - 2\gamma\eta) + \eta \delta_j x_{ji}$$

$$\delta_j = (t_j - o_j) o_j (1 - o_j)$$

also for hidden layer,

$$\boxed{w_{ji} = w_{ji} (1 - 2\eta\gamma) + \eta \delta_j x_{ji}}$$

$$\delta_j = o_j (1 - o_j) \sum_{k} \delta_k w_{kj}$$