Aadi Kothari
Andrew Su

CS 4375 Assignment 2 : NN

1)

1.1) Error (E_d) $= \frac{1}{2} \Sigma (t_k - o_k)^2$

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

$$Net_j = \sum_{i=1}^{n} x_{ji} w_{ji} \qquad \frac{\partial E_d}{\partial w_{ji}} = \frac{\partial E_d}{\partial net_j} \times \frac{\partial net_j}{\partial w_{ji}}$$

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

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

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

a) $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = O_j \qquad x = net_j$

$$\frac{\partial O_j}{\partial net_j} = \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2}$$

$$= \frac{4}{(e^x + e^{-x})^2} = 1 - (\tanh(x))^2 = 1 - (O_j)^2$$

Case 1: output unit:

$$\frac{\partial E_d}{\partial net_j} = -(t_j - o_j)(1 - o_j^2) = -\delta_j$$

$$\therefore \Delta w_{ji} = -\eta \frac{\partial E_d}{\partial w_{ji}} = \boxed{\eta (t_j - o_j)(1 - o_j^2)\ x_{ji}}$$

Case 2: j is hidden layer.

$$\frac{\partial E_d}{\partial net_j} = \Sigma - \delta_k\ w_{kj}\ (1 - o_j^2)$$

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

$$\therefore \Delta w_{ji} = \boxed{\eta\ (1 - o_j^2)\ \Sigma\ \delta_k\ w_{kj}\ \ x_{ji}}$$

b) $$Relu(x) = \max(0, x) = \begin{cases} 0 & x \le 0 \\ x & x > 0 \end{cases}$$

$$= o_j$$

$$\frac{\partial o_j}{\partial net_j} = \begin{cases} 0 & x \le 0 \\ 1 & x > 0 \end{cases}$$

Case 1: Output unit:

$$\frac{\partial E_d}{\partial net_j} = \begin{cases} 0 & x \le 0 \\ -(t_j - o_j) & x > 0 \end{cases}$$

$$\Delta w_{ji} = \begin{cases} 0 & x \le 0 \\ -\eta\ (t_j - o_j)\ x_{ji} & x > 0 \end{cases}$$

Case 2: j is hidden layer.

$$\frac{\partial E_d}{\partial net_j} = \begin{cases} 0 & x \le 0 \\ \sum_k -w_{kj}\,\delta_k & x > 0 \end{cases}$$

$$\therefore \Delta w_{ji} = \begin{cases} 0 & x \le 0 \\ \eta\, x_{ji} \sum_k -w_{kj}\,\delta_k & x > 0 \end{cases}$$

1.2 $w_\theta = w_0 + w_1(x_1 + x_1^2) + \cdots w_n(x_n + x_n^2)$

$$w_{j,new} = w_{j,old} - \eta \frac{\partial J(\theta)}{\partial w_j}$$

$$\frac{\partial J(\theta)}{\partial w_j} = \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^{m} \left(w_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^{m} \left((w_0 + w_1(x_1 + x_1^2) + \cdots w_n(x_n + x_n^2)) - y^{(i)}\right)^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(w_\theta(x)^{(i)} - y^{(i)}\right)\left(x_j + x_j^2\right)^{(i)}$$

$$\therefore w_{j,new} = w_{j,old} - \eta \frac{1}{m} \sum_{i=1}^{m} \left(w_\theta x^i - y^i\right)\left(x_j + x_j^2\right)^i$$

1.3 $o_3 = h(x_1 w_{31} + x_2 w_{32})$

$o_4 = h(x_1 w_{41} + x_2 w_{42})$

$$O_5 = h\left(w_{53}\, h(x_1 w_{31} + x_2 w_{32}) + w_{54}\, h(x_1 w_{41} + x_2 w_{42})\right)$$

b)

$$O_{\text{Hidden layer}} = h(W^{(1)} \cdot X)$$

$$O_{\text{output}} = h(W^{(2)} \cdot h(W^{(1)} \cdot X))$$

c) $h_s(x) = \dfrac{1}{1+e^{-x}}$ $\quad h_t(x) = \dfrac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

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

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

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

$$h_t(x) = \boxed{2\, h_s(2x) - 1}$$

$\therefore$ Since parameters differ only by linear transformation and constants, output functions are similar: Ex:- $O_5 = h_t(W^{(2)} h_t(W^{(1)}(x)))$

$$= h_t\left(W^{(2)}\left(2 \cdot h_s(2 \cdot W^{(1)}(x)) + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\right)$$

$$\therefore O_5 = 2\, h_s\left(2 W^{(2)}\left(2 \cdot h_s(2 \cdot W^{(1)}(x)) + \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)\right) + 1$$

1.4)

Q: 4.10 $$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - o_{kd})^2 + \gamma \sum_{i,j} w_{ji}^2$$

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} + \frac{\partial}{\partial w_{ji}} \left(\gamma \sum_{i,j} w_{ji}^2 \right)$$

$$\therefore \Delta w_{ji} = -\eta \left(\frac{\partial E_d}{\partial w_{ji}} + 2 \gamma w_{ji} \right)$$

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

Hidden layer $$\Delta w_{ji} = -\eta \left(\delta_j x_{ji} - 2 \gamma w_{ji} \right)$$

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