HW 2

Theoretical Part

1.1 Gradient Descent

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

$$w_i^{new} = w_i^{old} + \Delta w_i \quad \text{where} \quad \Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial\left(\frac{1}{2}\sum_{k\in outputs}(t_k - o_k)^2\right)}{\partial w_i} = \frac{1}{2}\sum_{k\in outputs}\frac{\partial}{\partial w_i}(t_k - o_k)^2$$

$$= \frac{1}{2}\sum_{k\in outputs} * 2 * (t_k - o_k)\cdot\frac{\partial}{\partial w_i}(t_k - o_k)$$

$$= \frac{1}{2}\sum_{k\in outputs} * 2 * (t_k - o_k)\cdot\frac{\partial}{\partial w_i}\left(t_k - \sum_i w_i(x_{ik} + x_{ik}^2)\right)$$

$$= \frac{1}{2}\sum_{k\in outputs} * 2 * (t_k - o_k)\cdot\left(-(x_{ik} + x_{ik}^2)\right)$$

$$= -\sum_{k\in outputs}(t_k - o_k)(x_{ik} + x_{ik}^2)$$

Therefore,

$$w_i^{new} = w_i^{old} + \Delta w_i = w_i^{old} - \eta\frac{\partial E}{\partial w_i} = w_i^{old} + \eta\sum_{k\in outputs}(t_k - o_k)(x_{ik} + x_{ik}^2)$$

1.2 Comparing Activation Function.

(a)

output 3: $h(x_1 w_{31} + x_2 w_{32})$

output 4: $h(x_1 w_{41} + x_2 w_{42})$

output y_5: $h(w_{53}\cdot h(x_1 w_{31} + x_2 w_{32}) + w_{54}\cdot h(x_1 w_{41} + x_2 w_{42}))$

(b)

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$W^{(1)} = \begin{pmatrix} w_{3,1} & w_{3,2} \\ w_{4,1} & w_{4,2} \end{pmatrix} \qquad W^{(2)} = (w_{5,3} \quad w_{5,4})$$

↓ After $W^{(1)} = W^{(1)}X$

After activation function $= h(W^{(1)}X)$

↓ After $W^{(2)} \Rightarrow W^{(2)}\cdot h(W^{(1)}X)$

output y_5: $h(W^{(2)}\cdot h(W^{(1)}X))$

(c)

$$hs(x) = \frac{1}{1+e^{-x}} \qquad ht(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} = \frac{\cancel{e^{x}}(1-e^{-2x})}{\cancel{e^{x}}(1+e^{-2x})} = \frac{1-e^{-2x}}{1+e^{-2x}}$$

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

Therefore $ht(x) = 2\cdot hs(2x) - 1$