1. Theoretical Part

1.1. Gradient descent.

$$O = w_0 + w_1(x_1 + x_1^2) + \dots + W_n(x_n + x_n^2)$$

We have: $\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \frac{1}{2} \sum_d (t_d - o_d)^2$ with $E_d(\vec{w}) = \frac{1}{2} \sum_{k \in outputs} (t_k - o_k)^2$

$$= \frac{1}{2} \sum_d 2(t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d)$$

$$= \sum_d (t_d - o_d) \frac{\partial}{\partial w_i} (t_d - \vec{w} . \vec{x}_d)$$

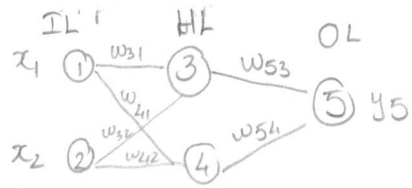
$$= \sum_d (t_d - o_d) - x_{id})$$

Now: $O = w_0 + w_1(x_1 + x_1^2) + \dots + w_n(x_n + x_n^2)$

Since: $\frac{\partial E}{\partial w_i} = \sum_d (t_d - o_d) - x_{id}$. (above)

then $\frac{\partial E}{\partial w_i} = \sum_d (t_d - o_d) - (x_{id} + x_{id}^2)$

1.2 Comparing Action function




a) Write down the output of the neural net $y_5$ in terms of weights, inputs, and general activation function $h(x)$.

Input layer:

- Neuron 1: Input is $x_1$
  Output is $f(x_1) = x_1$
- Neuron 2: Input is $x_2$
  Output is $f(x_2) = x_2$

Hidden layer

- Neuron 3: Input is $w_{31} \cdot x_1 + w_{32} x_2 = x_3$
  Output is $f(x_3) = x_3$
- Neuron 4: Input is $w_{41} . x_1 + x_{42} . x_2 = x_4$
  Output is $f(x_4) = x_4$

Output layer.

- Neuron 5: Input is $w_{53} . x_3 + w_{54} . x_4 = x_5$
  Output is $f(x_5) = x_5$

- Therefore:

$$y_5 = f(x_5) = x_5 = w_{53} . x_3 + w_{54} . x_4$$

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

b) $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\quad W^{(1)} = \begin{pmatrix} w_{3,1} & w_{3,2} \\ w_{4,1} & w_{4,2} \end{pmatrix}$ $\quad W^{(2)} = (w_{5,3} \quad w_{5,4})$

Write down the output of neutral net in vector format using above vector

◦) Input layer : $X \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Out put of the Input layer $f(x) = X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{2\times 1}$

◦) Hidden layer :

Input of Hidden layer: $z = W^{(1)} \cdot X = \begin{pmatrix} w_{3,1} & w_{3,2} \\ w_{4,1} & w_{4,2} \end{pmatrix}_{2\times 2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{2\times 1} = \begin{pmatrix} w_{3,1} \cdot x_1 + x_2 \cdot w_{3,2} \\ w_{4,1}\, x_1 + w_{4,2} \cdot x_2 \end{pmatrix}_{2\times}$

Out put of Hidden layer : $H(z) = z = \begin{pmatrix} w_{3,1} x_1 + w_{3,2} \cdot x_2 \\ w_{4,1} x_1 + w_{4,2} \cdot x_2 \end{pmatrix}_{2\times 1}$

◦) Output layer :

Input of Out put layer : $Out = W^{(2)} \cdot z = (w_{5,3} \quad w_{5,4})_{1\times 2} \begin{pmatrix} w_{3,1} x_1 + w_{3,2} x_2 \\ w_{4,1} x_1 + w_{4,2} x_2 \end{pmatrix}_{2\times 1}$

$= \left( w_{5,3}(w_{3,1} x_1 + w_{3,2} x_2) + w_{5,4}( w_{4,1} x_1 + w_{4,2} x_2) \right)_{1\times 1}$

Out put of Out put layer : $O(Out) = \left( w_{5,3}(w_{3,1} x_1 + w_{3,2} x_2) + w_{5,4}(w_{4,1} x_1 + w_{4,2} x_2) \right)_{1\times 1}$

Hence : Input layer : — Input matrix Size $2*1$ : $X$

Out put matrix size : $2*1$ : $f(x) = X$

Hidden layer — Input matriz Size : $2*1$ : $W^{(1)} X = z$

Output matriz Size : $2*1$ : $H(z) = z$

Out put layer — Input matriz Size : $1*1$ : $W^{(2)} \cdot z = Out$

Out put matriz Size : $1*1$ : $O(out) = Out$

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

Show that neural nets created using the above two activation function can generate the Same function.

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

Consider: $1 - h_s(x) = 1 - \sigma_{(x)} = 1 - \frac{1}{1+e^{-x}} = \frac{1}{1+e^{x}} = \sigma(-x)$

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

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

We have $\sigma(-2x) = \frac{1}{1+e^{2x}}$

then: $h_t(x) = 1 - 2 \cdot \sigma(-2x) = 1 - 2(1 - \sigma(2x)) = 2\sigma(2x) - 1$

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

$\Rightarrow$ tanh is rescale Sigmoid. Hence, they can generate Same output function