Theoretical Part - Rhed Santiago, Jeremy Pacheco

1.1 Gradient Descent

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

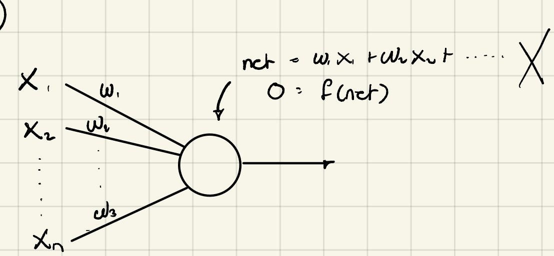
Inputs: $x_1, x_2, x_3 \dots x_n$ Output: O

Weights: $w_1, w_2, w_3 \dots w_n$

Bias: w_0

Learning Rate: η

net $= w_1 x_1 + w_2 x$



$\rightarrow w_i^{new} = w_i^{old} - \alpha\left(\frac{do}{dw_i}\right)$

$\rightarrow w_i^{new} = w_i^{old} - \eta\left(\frac{do}{dw_i}\right)$

$\rightarrow \left(\frac{do}{dw_i}\right) = x_i + x_i^2$

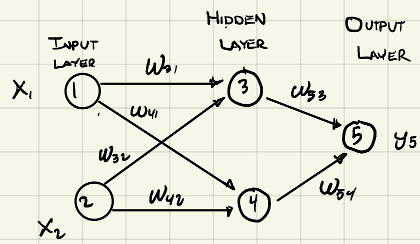
$\therefore$ ANS// $w_i^{new} = w_i^{old} - \eta(x_i + x_i^2)$

1.2 Comparing Activation Function

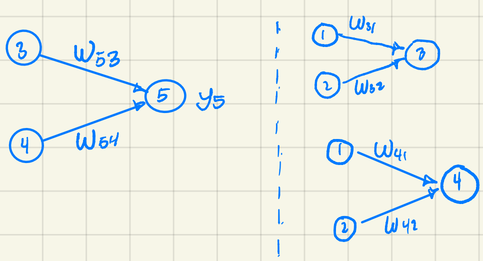
Neural Net with:

f(x) activation : 2 Input Layer Neurons

h(x) activation : 1 Hidden Layer



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

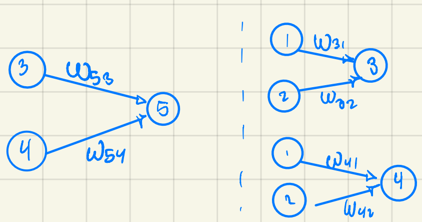


$y_5 = h(w_{53} x_3 + w_{54} x_4)$

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

b.) Suppose vector notation, where: $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $W^{(1)} = \begin{pmatrix} w_{3,1} & w_{3,2} \\ w_{4,1} & w_{4,2} \end{pmatrix}$, $W^{(2)} = (w_{5,3} \; w_{5,4})$

Output of the Neural net in vector format using above vectors.



$y_5 = h(w_{53}\, h(w_{31} x_1 + w_{32} x_2) + w_{54}\, h(w_{41} x_1 + w_{42} x_2))$

$y_5 = h(w_{53}\, h(w^{(1)}_{1 \times n} \cdot X) + w_{54}\, h(w^{(1)}_{2 \times n} \cdot X))$

where $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ $= h(w_{53}\, h(\underbrace{[w_{31} \; w_{32}] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\alpha_1}) + w_{54}\, h(\underbrace{[w_{41} \; w_{42}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\alpha_2}))$

$= h(w^{(2)} \cdot \alpha)$

Ans// $h\left([w_{53} \; w_{54}] \begin{bmatrix} h([w_{31} \; w_{32}] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) \\ h([w_{41} \; w_{42}] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) \end{bmatrix}\right)$

or $h\left(w^{(2)} \cdot \begin{bmatrix} h(w^{(1)}_{1,n} \cdot X) \\ h(w^{(1)}_{2,n} \cdot X) \end{bmatrix}\right)$

c.) Suppose two choices for activation function $h(x)$, as shown below

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

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

*Hint: Compute relationship between $h_s(x)$ and $h_t(x)$ and then show that the output functions are the same, with parameters differing only by linear transformations and constants

- Show that neural nets using the above two activation functions can generate the same function

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

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

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

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

$$\text{Sig} + \text{Sig}(e^x) = e^x \Rightarrow \text{Sig} = e^x - \text{Sig}(e^x)$$

$$\text{Sig} = e^x(1-\text{Sig}) \Rightarrow \left(\frac{\text{Sig}}{1-\text{Sig}}\right) = e^x$$

$$\text{Tanh} = \frac{\left(\frac{\text{Sig}}{1-\text{Sig}}\right)^2 - 1}{\left(\frac{\text{Sig}}{1-\text{Sig}}\right)^2 + 1} \cdot \frac{(1-\text{Sig})^2}{(1-\text{Sig})^2} = \frac{\cancel{(1-\text{Sig})^2}\left(\frac{\text{Sig}^2}{\cancel{(1-\text{Sig})^2}}\right) - (1-\text{Sig})^2 \cdot \cancel{1}}{\cancel{(1-\text{Sig})^2}\left(\frac{\text{Sig}^2}{\cancel{(1-\text{Sig})^2}}\right) - (1-\text{Sig})^2 \cdot \cancel{1}}$$

$$= \frac{\text{Sig}^2 - (1-\text{Sig})^2}{\text{Sig}^2 + (1-\text{Sig})^2}$$

$$= \frac{\text{Sig}^2 - (1^2 - 2(1)(\text{Sig}) + \text{Sig}^2)}{\text{Sig}^2 + (1^2 - 2(1)(\text{Sig}) + \text{Sig}^2)}$$

$$= \frac{\cancel{\text{Sig}^2} - 1^2 + 2\,\text{sig} - \cancel{\text{Sig}^2}}{\text{Sig}^2 + 1^2 - 2\,\text{Sig} + \text{Sig}^2}$$

$$\text{Tanh} = \frac{2\,\text{sig} - 1}{2\,\text{Sig}^2 - 2\,\text{sig} + 1} \Rightarrow h_t(x) = \frac{2h_s(x) - 1}{2(h_s(x))^2 - 2h_s(x) + 1}$$

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

$$y_5 = h(w_{53}h(w_{31}x_1 + w_{32}x_2) + w_{54}h(w_{41}x_1 + w_{42}x_2))$$

with $h_t(x)$ activation //

$$y_5 = h(w_{53}h(w_{31}(2h_s(2x_1)-1) + w_{32}(2h_s(2x_2)-1)) + w_{54}h(w_{41}(2h_s(2x_1)-1) + w_{42}(2h_s(2x_2)-1)))$$

$$y_5 = h\left(w_{53}h\left(w_{31}\left(2\left(\frac{1}{1+e^{-2x_1}}\right)-1\right) + w_{32}\left(2\left(\frac{1}{1+e^{-2x_2}}\right)-1\right)\right) + w_{54}h\left(w_{41}\left(2\left(\frac{1}{1+e^{-2x_1}}\right)-1\right) + w_{42}\left(2\left(\frac{1}{1+e^{-2x_2}}\right)-1\right)\right)\right)$$

with $h_s(x)$ activation //

$$y_5 = h(w_{53}h(w_{31}(h_s(x_1)) + w_{32}(h_s(x_2))) + w_{54}h(w_{41}(h_s(x_1)) + w_{42}(h_s(x_2))))$$

$$y_5 = h\left(w_{53}h\left(w_{31}\left(\frac{1}{1+e^{-x_1}}\right) + w_{32}\left(\frac{1}{1+e^{-x_2}}\right)\right) + w_{54}h\left(w_{41}\left(\frac{1}{1+e^{-x_1}}\right) + w_{42}\left(\frac{1}{1+e^{-x_2}}\right)\right)\right)$$