Assignment 2 NN's

1.1

a) <u>tanh activation function</u>:

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

$$\tanh'(x) = 1 - \tanh^2(x)$$

$$L = \frac{1}{2}(y - \hat{y})$$

$$\frac{dL}{d\hat{y}} = \hat{y} - y$$ (slope of loss function w.r.t. predicted val $\hat{y}$)

so...

$$\delta_o = (\hat{y} - y) \cdot \tanh'(net_o)$$
$$\delta_o = (\hat{y} - y) \cdot (1 - \tanh^2(net_o))$$

$$\delta_h = \delta_o \cdot w_{ho} \cdot \tanh'(net_h)$$
$$\delta_h = \delta_o \cdot w_{ho} \cdot (1 - \tanh^2(net_h))$$ ← for each hidden layer

h: hidden layer(s)
o: output layer

b) <u>ReLU activation function</u>:

$$\mathrm{ReLu}(x) = \max(0, x) \qquad \mathrm{ReLu}'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

$$L = \frac{1}{2}(y - \hat{y}) \qquad \frac{dL}{d\hat{y}} = \hat{y} - y$$

so...

$$\delta_o = (\hat{y} - y) \cdot \mathrm{ReLU}'(net_o)$$
$$\delta_h = \delta_o \cdot w_{ho} \cdot \mathrm{ReLU}'(net_h)$$ ← for each hidden layer

1.2 $o = w_0 + w_1(x_1 + x_1^2) + \ldots + w_n(x_n + x_n^2)$

Here's the loss function: $L = \frac{1}{2}(t - o)^2$

or $L = \frac{1}{2}(y - \hat{y})^2$

We want to minimize L...

$$w_{i(new)} = w_{i(old)} - \eta \cdot \frac{dL}{w_i}$$

$$\frac{dL}{dw_i} = \frac{dL}{do} \cdot \frac{do}{dw_i} \qquad \frac{dL}{do} = \frac{d}{do}\left(\frac{1}{2}(t-o)^2\right) = -(t-o)$$

$$\frac{do}{dw_i} = \begin{cases} 1 & \text{if } i = 0 \\ x_1 + x_1^2 & \text{if } i = 1 \\ x_2 + x_2^2 & \text{if } i = 2 \end{cases}$$

$$\frac{dL}{dw_i} = -(t-o) \cdot \frac{do}{dw_i}$$

so...

$$\boxed{w_{i(new)} = w_{i(old)} - \eta \cdot (t-o) \cdot (x_i + x_i^2)}$$

t = target value

o = perdicted output from model

1.3.

a. $y_5 = h(net_5) = h(\omega_{53} \cdot x_3 + \omega_{54} \cdot x_4)$

$= \boxed{h(\omega_{53} \cdot h(\omega_{31} \cdot x_1 + \omega_{32} \cdot x_2) + \omega_{54} \cdot h(\omega_{41} \cdot x_1 + \omega_{42} \cdot x_2))}$

b. $y_5 = h(W^{(2)} \cdot h(W[0]^{(1)} \cdot X) + h(W[1]^{(1)} \cdot X))$

$\boxed{y_5 = h(W^{(2)} \cdot h(W^{(1)} \cdot X))}$

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

We can use identity: $\tanh(x) = 2\sigma(2x) - 1$ ← * This is computed on next page *

$h_t(x) = 2h_s(2x) - 1$ ← we get this now

we can consider tanh linear activation function:

$y_t = h_t(wx+b)$

$$\boxed{\begin{aligned} y_t &= 2 \cdot h_s(2 \cdot (wx+b_1)) - 1 \\ y_s &= h_s(wx + b_2) \end{aligned}}$$

← convert to sigmoid terms

b_1 and b_2 are both bias terms, but may be different in each function.

* Derivation of $\tanh(x) = 2\sigma(2x) - 1$ on next page *

Derivation of $h_s(X)$ and $h_t(X)$ relationship

$$\sigma(X) = h_s(X) = \frac{1}{1+e^{-x}} \qquad \tanh(x) = h_t(X) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$1 - \sigma(X) = (-X)$ ← we know this from looking at graph

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

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

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

$$= 1 - 2 + 2\sigma(2x) = 2\sigma(2x) - 1$$

$$\boxed{\tanh(X) = 2\sigma(2x) - 1}$$