Khoa Diep ktd170030
Farrel Raja fjr190000
Maviya Yaseen mxy200016

1.1a)

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

<u>Denotation for Hidden Layer:</u>

- Input to the activation function for neuron: z_j^n
- Output to the activation function for neuron: a_j^n
- Where neuron *j* is in hidden layer *h*

<u>Denotion for Output Layer:</u>

- Neuron k in the output layers
- Input to the output layer: z_k^0
- Output to the output layer: a_k^0

<u>Forward Pass for Hidden Layer:</u>

$$z_j^n = \sum_{i=1}^{n} w_{ij}^n * x_i + b_j^n$$

$$a_j^n = \tanh(z_j^n)$$

<u>Forward Pass for Output Layer:</u>

$$z_k^0 = \sum_{j=1}^{m} w_{ij}^0 * a_j^h + b_k^0$$

$$a_k^0 = \tanh(z_k^0)$$

<u>Backward Pass for Hidden Layer:</u>

- Error:

$$\delta_j^n = (\sum_{k=1}^{L} w_{jk}^0 * \delta_k^0) * (1 - (a_j^n)^2)$$

- Weights:

$$\Delta w_{ij}^n = \eta \delta_j^n * x_i$$

$$\Delta b_j^n = \eta \delta_j^n$$

Backward Pass for Output Layer:

- Error:

$$\delta_k^0 = (a_k^0 - y_k) * (1 - (a_k^0)^2)$$

- Weights:

$$\Delta w_{jk}^0 = \eta \delta_k^0 * a_j^n$$

$$\Delta b_k^0 = \eta \delta_k^0$$

1.1b) (Using same denotation from part a)

$$\text{ReLU}(x) = \max(0, x)$$

Forward Pass for Hidden Layer:

$$z_j^n = \sum_{i=1}^{n} w_{ij}^n * x_i + b_j^n$$

$$a_j^n = ReLu(z_j^n)$$

Forward Pass for Output Layer:

$$z_k^0 = \sum_{j=1}^{m} w_{jk}^0 * a_i^m + b_k^0$$

$$a_k^0 = ReLu(z_k^0)$$

Backwards Pass for Hidden Layer:

- Error:

$$\delta_j^n = \begin{cases} 1 & z_j^n > 0 \\ 0 & else \end{cases}$$

- Weight:

$$\Delta w_{ij}^n = \eta \delta_j^n * x_i$$

$$\Delta b_j^n = \eta \delta_j^n$$

Backwards Pass for Output Layer:

- Error:

$$\delta_k^0 = \begin{cases} 1 & z_k^0 > 0 \\ 0 & else \end{cases}$$

- Weights:

$$\Delta w_{jk}^0 = \eta \delta_k^0 * a_j^n$$

$$\Delta b_k^0 = \eta \delta_k^0$$

1.2)

<u>Error:</u>

Error E is defined as:

$$E = \frac{1}{2}(O - y)^2$$

- Where y = desired weight

<u>W0</u>

$$\frac{\delta E}{\delta O} = \frac{\delta \frac{1}{2}(O - y)^2}{\delta O} = (O - y)$$

$$\frac{\delta O}{\delta w_0} = \frac{\delta w_0 + w_1(x_1 + x_1^2) + ... + w_n(x_n + x_n^2)}{w_0} = 1$$

$$\frac{\delta E}{\delta w_0} \rightarrow \frac{\delta E}{\delta O} * \frac{\delta O}{\delta w_0} = \frac{(O - y)}{1} = (O - y)$$

$$w_0 = w_0 - \eta \frac{\delta E}{\delta w_0} = w_0 - \eta(O - y)$$

<u>Wi</u>

$$\frac{\delta O}{\delta w_i} = \frac{\delta w_0 + w_1(x_1 + x_1^2) + ... + w_i(x_i + x_i^2) + ... + w_n(x_n + x_n^2)}{\delta w_i}$$

$$\frac{\delta O}{\delta w_i} = \frac{\delta w_i(x_i + x_i^2)}{\delta w_i} = (x_i + x_i^2)$$

$$\frac{\delta E}{\delta w_i} \rightarrow \frac{\delta E}{\delta O} * \frac{\delta O}{\delta w_i} = (O - y)(x_i + x_i^2)$$

$$w_i \rightarrow w_i - \eta \frac{\delta E}{\delta w_i} = w_i - \eta(O - y)(x_i + x_i^2)$$

1.3a)

$$\text{input layer function: } f(x) = x$$

$$\text{hidden and output layer function: } h(x)$$

$$z_3 = x_1 w_{31} + x_2 w_{32}$$

$$h_1 = h(z_3) = h(x_1 w_{31} + x_2 w_{32})$$

$$z_4 = x_1 w_{41} + x_2 w_{42}$$

$$h_2 = h(z_4) = h(x_1 w_{41} + x_2 w_{42})$$

$$y_5 = h_1 * w_{53} + h_2 * w_{54}$$

1.3b)

$$z = x * w^1 \rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{32} \end{pmatrix}$$

$$z = \begin{pmatrix} x_1(w_{31} + w_{41} \\ x_2(w_{32} + w_{42} \end{pmatrix}$$

Applying the activation function $h(x)$ to z element-wise:

$$H = h(z) = \begin{pmatrix} h(x_1(w_{31} + w_{41}) \\ h(x_2(w_{32} + w_{42}) \end{pmatrix}$$

$$\text{Output} = y_5 = H * w^{(2)}$$

1.3c)

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

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

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

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

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

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

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

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

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

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

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

Using sigmoid and tanh notation:

$$\tanh(x) = 1 - 2 * \mathrm{sigmoid}(-2x)$$

We just derived the relationship between tanh and sigmoid are simple transformations and can the output functions could be the same if with these transformation:

Below is a desmos graph showing that these functions are the same.
Note: Blue graph and Green graph overlaps