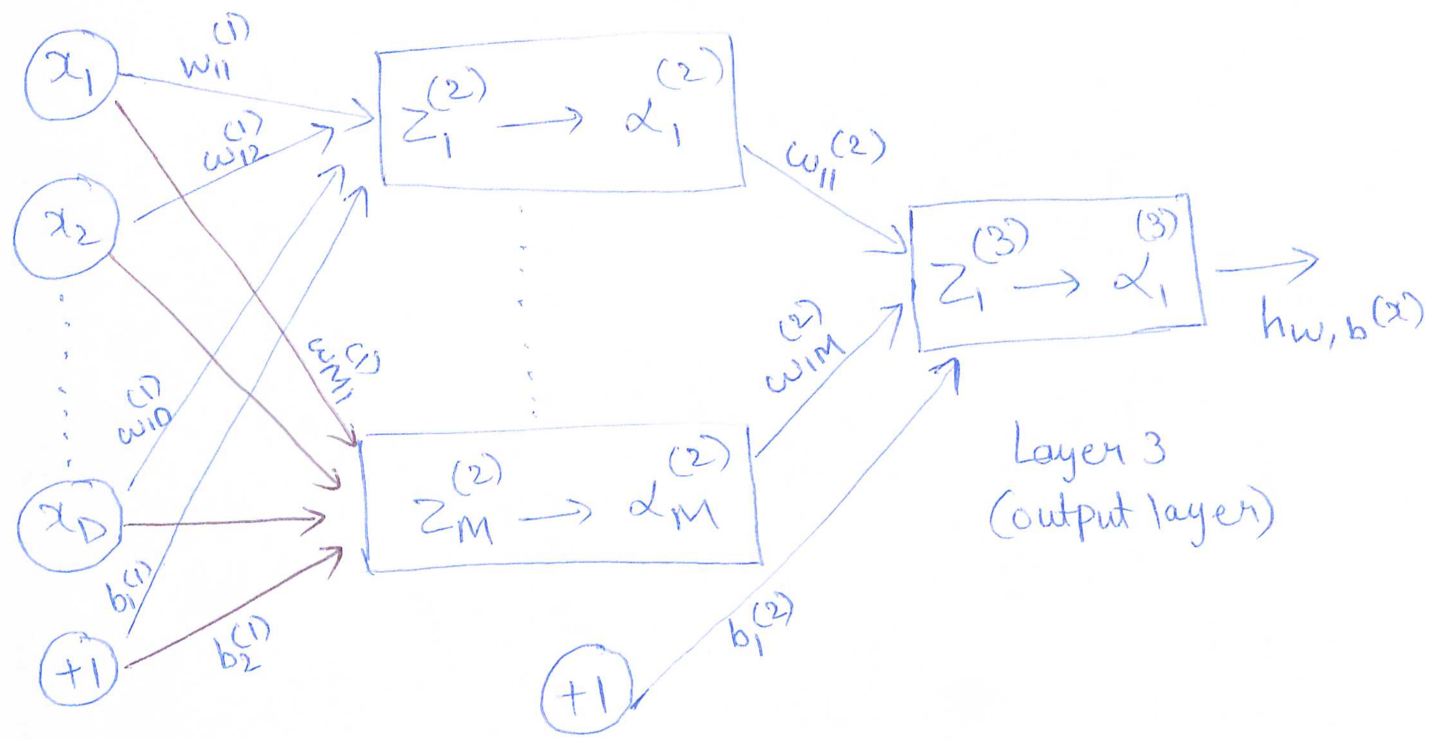
Question - 1:

(a) Schematic Representation:



Layer 1
(Input Layer)

Layer 2 (Hidden Layer)

$$f_\sigma = \text{Activation function } (\alpha) = \frac{1}{1+\exp(-\alpha)}$$

$$z_1^{(2)} = w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2 + w_{13}^{(1)} x_3 + \cdots + w_{1D}^{(1)} x_D + b_1^{(1)}$$

$$\vdots$$

$$z_M^{(2)} = w_{M1}^{(1)} x_1 + w_{M2}^{(1)} x_2 + w_{M3}^{(1)} x_3 + \cdots + w_{MD}^{(1)} x_D + b_M^{(1)}$$

$$\alpha_1^{(2)} = f_\sigma\left(z_1^{(2)}\right) \quad \text{or} \quad \alpha_M^{(2)} = f_\sigma\left(z_M^{(2)}\right)$$

$$\text{output} = h_{w,b}(x) = \alpha_1^{(3)} = f_\sigma\left(z_1^{(3)}\right)$$

In matrix Form,

$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & \cdots & w_{1D}^{(1)} \\ \vdots & & & \\ w_{M1}^{(1)} & w_{M2}^{(1)} & \cdots & w_{MD}^{(1)} \end{bmatrix}_{M \times D}$$

$$b^{(1)} = \begin{bmatrix} b_1^{(1)} & b_2^{(1)} & \cdots & b_M^{(1)} \end{bmatrix}_{1 \times M}$$

$$\alpha^{(2)} = f_\sigma(w^{(1)}x + b^{(1)})$$

$$h_{w,b}(x) = \alpha^{(3)} = f_\sigma(w^{(2)}\alpha^{(2)} + b^{(2)})$$

(b)

$$\text{output} = h_{w,b}(x) = \alpha_1^{(3)} = f_\sigma(w_{11}^{(2)}\alpha_1^{(2)} + \cdots + w_{1M}^{(2)}\alpha_M^{(2)} + b_1^{(2)})$$

$$\text{where } \alpha_M^{(2)} = f_\sigma(w_{M1}^{(1)}x_1 + w_{M2}^{(1)}x_2 + \cdots + w_{MD}^{(1)}x_D + b_M^{(1)})$$

In general,

$$h_{w,b}(x) = \alpha^{(3)} = f_\sigma(w^{(2)}\alpha^{(2)} + b^{(2)})$$

$$\text{where } \alpha^{(2)} = f_\sigma(w^{(1)}x + b^{(1)})$$

(c)

Number of Parameters:

Total weights between 1st, 2nd layer $= \underbrace{D+D+\cdots}_{M\text{ times}} + \underset{\text{biases}}{M}$

$= M(D+1)$

Total weights between 2nd, 3rd layer $= \underset{\text{weights}}{M} + \underset{\text{bias term}}{1}$

$$\boxed{\text{Total Parameters} = M(D+2) + 1}$$

(d) Consider two activation functions

$\text{Sigmoid}(\alpha) = f_{\sigma}(\alpha) = \dfrac{1}{1+e^{-\alpha}}$; $\tanh(\alpha) = f_{tan}(\alpha) = \dfrac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}}$

$\Rightarrow f_{tan}(x) = \dfrac{e^{2x} - 1}{e^{2x} + 1}$

$\Rightarrow f_{tan}(x) = 2\left[\dfrac{e^{2x}}{e^{2x}+1}\right] - 1$

$= 2 f_{\sigma}(2x) - 1$

$\Rightarrow \boxed{f_{tan}(x) = 2 f_{\sigma}(2x) - 1}$

Neural network - 1 with sigmoid Activation:

$$\text{output} = h_{w,b}(x) = \alpha_1^{(3)} = f_\sigma\left(w_{11}^{(2)} \alpha_1^{(2)} + \dots + w_{1M}^{(2)} \alpha_M^{(2)} + b_1^{(2)}\right)$$

$$(1) \rightarrow \text{ where } \alpha_M^{(2)} = f_\sigma\left(w_{M1}^{(1)} x_1 + w_{M2}^{(1)} x_2 + \dots + w_{MD}^{(1)} x_D + b_M^{(1)}\right)$$

Neural network - 2 with tanh Activation: $(w,b) \Leftrightarrow (v,c)$

$$\text{output} = h_{v,c}(x) = \alpha_1^{(3)} = f_{tan}\left(v_{11}^{(2)} \alpha_1^{(2)} + \dots + v_{1M}^{(2)} \alpha_M^{(2)} + c_1^{(2)}\right)$$

$$(2) \rightarrow \text{ where } \alpha_M^{(2)} = f_{tan}\left(v_{M1}^{(1)} x_1 + v_{M2}^{(1)} x_2 + \dots + v_{MD}^{(1)} x_D + c_M^{(1)}\right)$$

we know that, $f_{tan}(x) = 2 f_\sigma(2x) - 1$

Consider,
From (1), $\alpha_M^{(2)} = f_\sigma\left(w_{M1}^{(1)} x_1 + \dots + w_{MD}^{(1)} x_D + b_M^{(1)}\right)$

$$(3) \rightarrow \alpha_M^{(2)} = \frac{1}{2} f_{tan}\left(\frac{w_{M1}^{(1)} x_1 + w_{M2}^{(1)} x_2 + \dots + b_M^{(1)}}{2}\right) + \frac{1}{2}$$

comparing (2), (3)

$$\Rightarrow v_{M1}^{(1)} = k_1 w_{M1}^{(1)} + k_2$$ where k_1, k_2 are constants

$\vdots$

$$v_{MD}^{(1)} = k_1 w_{MD}^{(1)} + k_2$$

and $c_M^{(1)} = k_1 b_M^{(1)} + k_2$ so on.

where $V_{M1}, \ldots V_{MD}$ are weights in the network-2 and

$W_{M1}, \ldots W_{MD}$ are weights in the network-1

After substituting these in output equation, we can see that there exists an equivalent network with hidden unit activation functions given by hyperbolic tangent, which computes exactly the same function as that of with Sigmoid activation function.

END OF Question 1