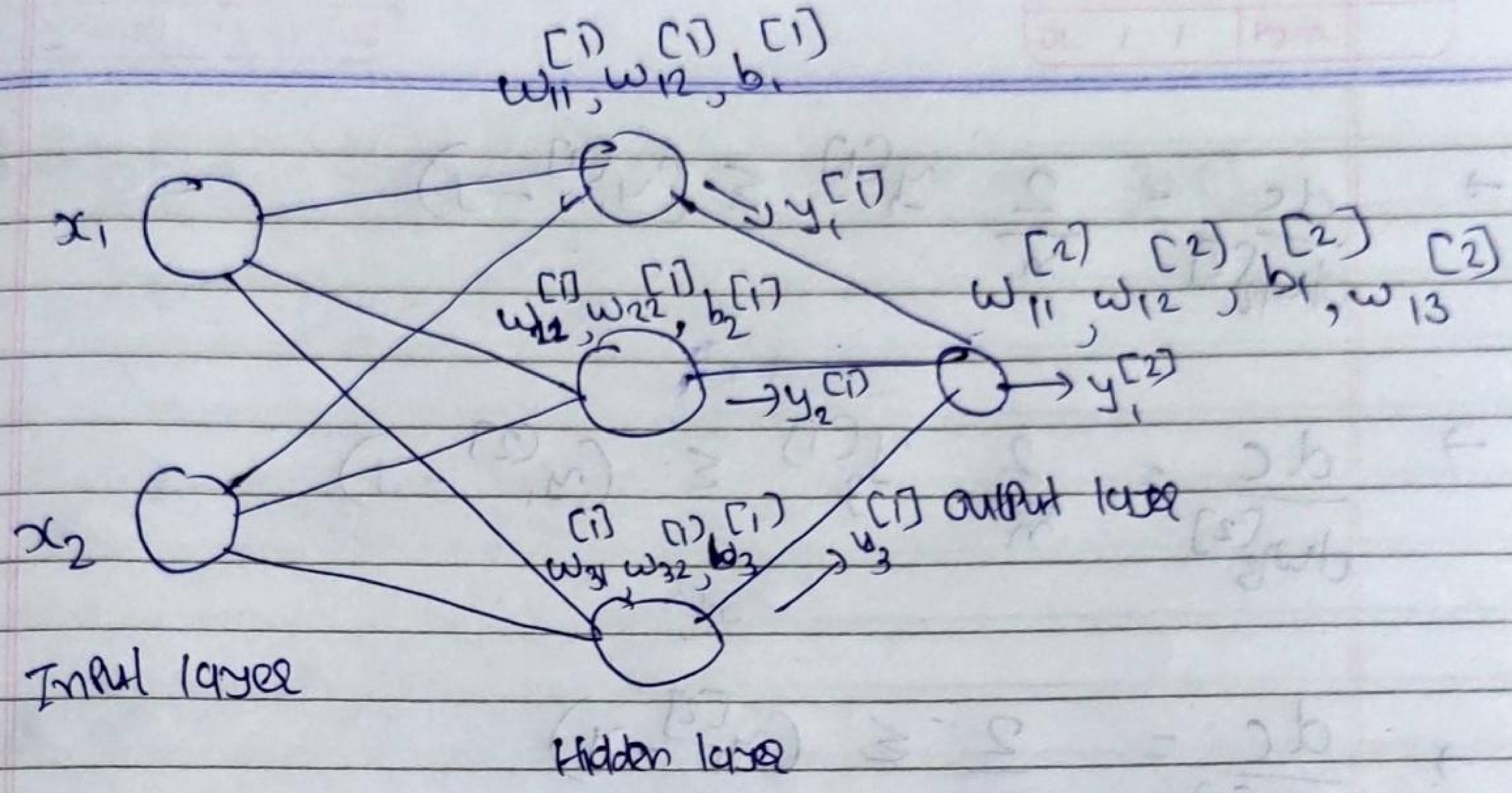

→ Forward Prop:

→ $y_1^{[1]} = w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2 + b_1^{[1]}$

→ $y_2^{[1]} = w_{21}^{[1]} x_1 + w_{22}^{[1]} x_2 + b_2^{[1]}$

→ $y_3^{[1]} = w_{31}^{[1]} x_1 + w_{32}^{[1]} x_2 + b_3^{[1]}$

→ $y_1^{[2]} = w_{11}^{[2]} y_1^{[1]} + w_{12}^{[2]} y_2^{[1]} + w_{13}^{[2]} y_3^{[1]} + b_1^{[2]}$

→ $MSE = \frac{1}{n} \sum_{i=1}^{n} (y_1^{[2]} - \underbrace{y}_{\text{Target variable}})^2$

→ Back Prop:

→ $\frac{dc}{dy_1^{[2]}} = \frac{2}{n} \sum_{i=1}^{n} (y_1^{[2]} - y)$ (Change in cost wrt $y_1^{[2]}$)

→ $\frac{dc}{dw_{11}^{[2]}} = \frac{dy_1^{[2]}}{dw_{11}^{[2]}} \times \frac{dc}{dy_1^{[2]}} = \frac{2}{n} y_1^{[1]} \sum_{i=1}^{n} (y_1^{[2]} - y)$

$\rightarrow \dfrac{dc}{dw_{12}^{[2]}} = \dfrac{2}{n} y_2^{[1]} \sum (y_1^{[2]} - y)$

$\rightarrow \dfrac{dc}{dw_{13}^{[2]}} = \dfrac{2}{n} y_3^{[1]} \sum (y_1^{[2]} - y)$

$\rightarrow \dfrac{dc}{db_1^{[2]}} = \dfrac{2}{n} \sum (y_1^{[2]} - y)$

$\rightarrow \dfrac{dc}{dy_1^{[1]}} = \dfrac{dy_1^{[2]}}{dy_1^{[1]}} * \dfrac{dc}{dy_1^{[2]}}$

$= w_{11}^{[2]} * \dfrac{2}{n} \sum_{i=1}^{n} (y_{11}^{[2]} - y)$

$\rightarrow$ Similarly for $\dfrac{dc}{dy_2^{[1]}}, \dfrac{dc}{dy_3^{[1]}}$

$\rightarrow \dfrac{dc}{dw_{11}^{[1]}} = \dfrac{dy_1}{dw_{11}^{[1]}} \times \dfrac{dc}{dy_1^{[1]}}$

$= x_1 \, w_{11}^{[2]} * \dfrac{2}{n} \sum (y_1^{[2]} - y)$

$\rightarrow \dfrac{dc}{db^{[1]}} = w_1^{[2]} * \dfrac{2}{n} \sum_{1}^{3} (y^{[2]} - y)$

$\rightarrow$ Similarly, we do chain derivation for other variables.

→ <u>Updating</u> <u>variables</u>: (α = learning rate)

→ $w_{11}^{[2]} = w_{11}^{[2]} - \alpha \dfrac{dC}{dw_{11}^{[2]}}$

→ $w_{12}^{[2]} = w_{12}^{[2]} - \alpha \dfrac{dc}{dw_{12}^{[2]}}$

→ $b_1^{[2]} = b_1^{[2]} - \alpha \dfrac{dc}{db_1^{[2]}}$

→ So to fast the operations for number of examples we do vectorisation for example:

$$\begin{bmatrix} | & | & | & \\ x_1 & x_2 & x_3 & \cdots \\ | & | & | & \end{bmatrix}$$ → This is one input to NN.

first data pts 2^{nd} data pts