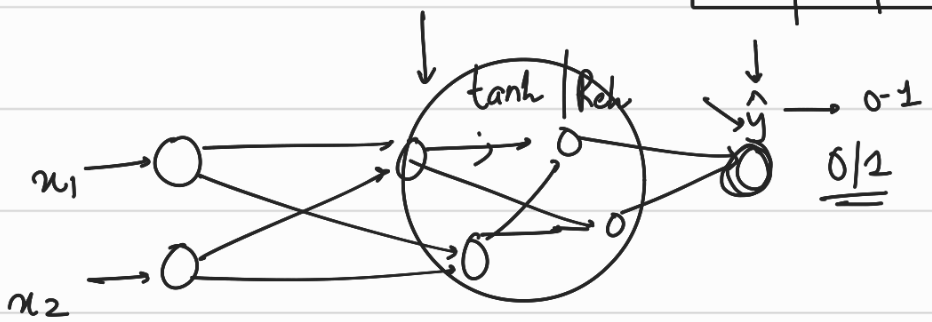
Neural Network train on a regression problem:
Initialise the NN
We forward propagate, calculate the loss (mse for regression)
given the loss, we backward propagate i.e. partial differentiate the loss wrt weights and biases
follow the gradient descent equation.
We perform this iteratively until loss saturates.

Classification (Binary Classification):

x_1	x_2	y
		0
		1

1)



Changes: 1) activation function at the O/L.
Sigmoid →
2) Loss → diff $(y, \hat{y})$

$0/1 \,,\ \hat{y} \to 0\text{-}1$

Log Loss

$$-\sum_{i=1}^{n} \frac{1}{n}\ \overset{1}{y_i \log \hat{y}_i} + \overset{0}{(1-y_i) \times \log(1-\hat{y}_i)}$$

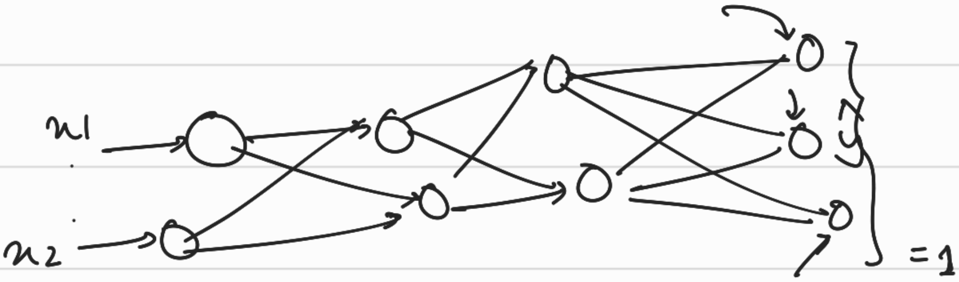
$y_i = 1$ $\hat{y}_i = 0.95$
$= 0.022$
$y_i = 1$ $\hat{y}_i = 0.2$
0.69

$y_i = 0$ $\hat{y}_i = 0.05$
$= 0.022$
$y_i = 0$ $\hat{y}_i = 0.8$
$= 0.69$

2) Multiclass Problems:

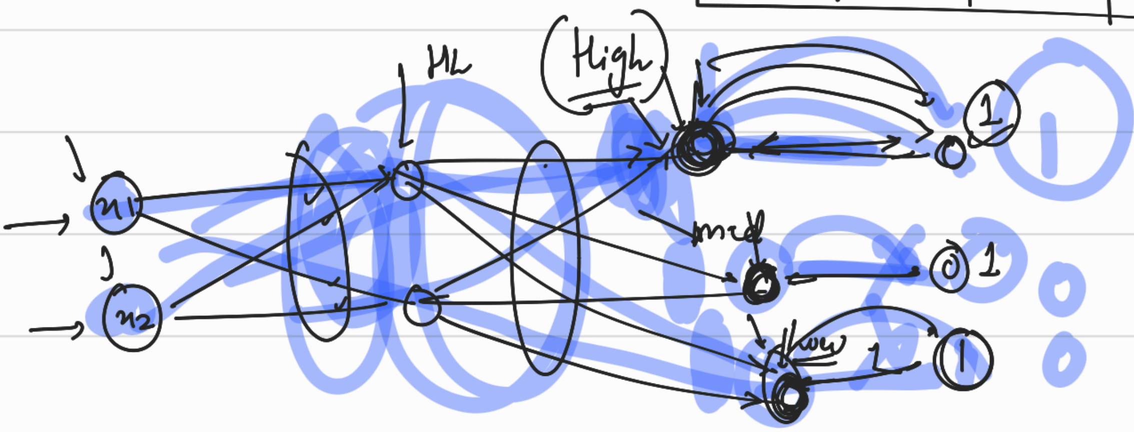
x_1	x_2	y_r
		High Med Low

X



one hot encoding

x_1	x_2	y		
		High	Med	Low
✓	✓	1	0	0
✓	✓	0	1	0
✓	✓	0	0	1
✓	✓	1	0	0



$$\begin{array}{cccc} & H & M & L \\ \hat{y} = & [0.7 & 0.2 & 0.1] \end{array}$$

$$[0.33 \quad 0.33 \quad 0.33]$$

Softmax

$$\begin{array}{ccc} H & M & L \\ [1 & 0 & 0] \end{array}$$

p_1

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \frac{e^{x_1}}{e^{x_1} + e^{x_2}}$$

Activation

$$\begin{bmatrix} \dfrac{e^{x_1}}{\sum_{i=1}^{3} e^{x_i}} \\[2ex] \dfrac{e^{x_2}}{\sum_{i=1}^{3} e^{x_i}} \\[2ex] \dfrac{e^{x_3}}{\sum_{i=1}^{3} e^{x_i}} \end{bmatrix} \qquad \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$\hat{p}_2 = \left[\frac{e^{x_2}}{e^{x_1} + \dots}\right]$$

$$i = 0 \qquad \hat{y}_i = \begin{array}{ccc} \hat{p}_1 & \hat{p}_2 & \hat{p}_3 \\ \end{array} \begin{bmatrix} 0.8 & 0.1 & 0.1 \end{bmatrix} \qquad y_i = \begin{array}{ccc} p_1 & p_2 & p_3 \end{array} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$\text{Log Loss} = -\left(\cancel{p_1 \times \log \hat{p}_1} + p_2 \times \log \hat{p}_2 + \cancel{p_3 \times \log \hat{p}_3}\right)$$

$$0.1$$

$$\sum_{j=1}^{K} p_j \times \log \hat{p}_j$$

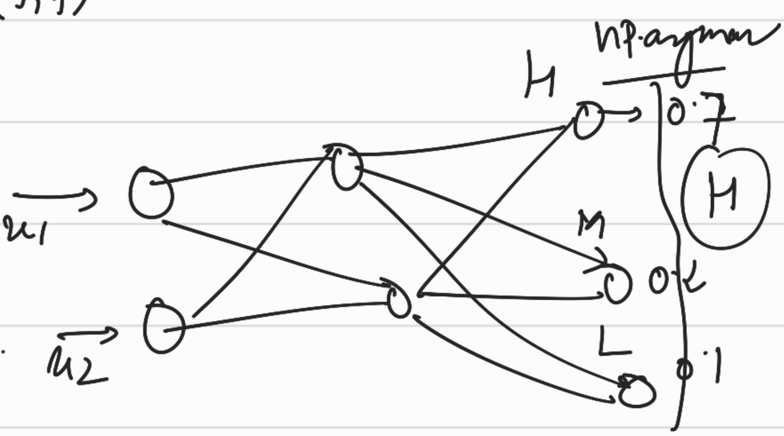
	x_1	x_2	y			$\hat{y}$			
			H	M	L	H	M	L	
1	✓	✓	1	0	0	0.2	0.3	0.5	$-\log 0.2$
2	✓	✓	0	1	0	0.1	0.6	0.1	$-\log 0.6$
3	.	-	-						-

$$\sum_{i=1} (\dots)$$

$$dL \qquad d(\text{diff}(y, \hat{y}))$$

dL

Training



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \frac{e^{x_1}}{e^{x_1}+e^{x_2}} = p_1$$

$$\frac{e^{x_2}}{e^{x_1}+e^{x_2}} = \cancel{p_2} \; 1-p_1$$

$p_1 =$

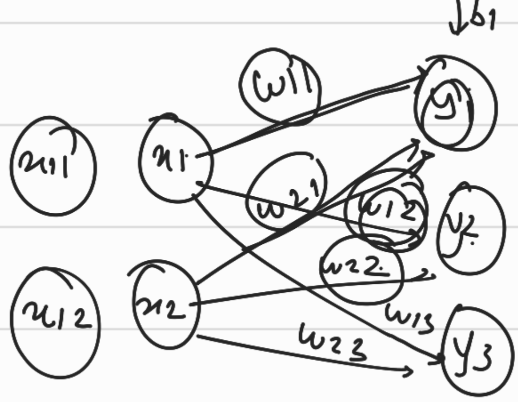
$$p_1 + p_2 = 1$$

$$\frac{e^{x_1}}{e^{x_1}+e^{x_2}} + \frac{e^{x_2}}{e^{x_1}+e^{x_2}} = 1$$

$$\frac{e^{x_1}}{e^{x_1}+e^{x_2}} = 1 - \frac{e^{x_2}}{e^{x_1}+e^{x_2}}$$

$$p_1 = \frac{1}{1 + \cancel{e^{-x_1}} \; \frac{1}{e^{x_1}}}$$

$$x = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \\ x_{300\,1} & x_{300\,2} \end{bmatrix}$$



x_1

$$w_{11} x_{11} + w_{21} x_{12} + b_1 = y_{11}$$

$$w_{12} x_{11} + w_{22} x_{12} + b_2 = y_{12}$$

$$w_{13} x_{11} + w_{23} x_{12} + b_3 = y_{13}$$

$$\begin{bmatrix} x_{11} & x_{12} \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} = y_0$$

$$= \begin{bmatrix} y_{11} & y_{12} & y_{13} \end{bmatrix}$$

labelencoded.

x_1	x_2	y
2	3	1

$$-\log\left[\, p[y] \,\right]$$

0.2

$$\begin{bmatrix} 0.6 & 0.2 & 0.2 \end{bmatrix}$$