

Neural Network train on a regression problem:

Initialise the NN

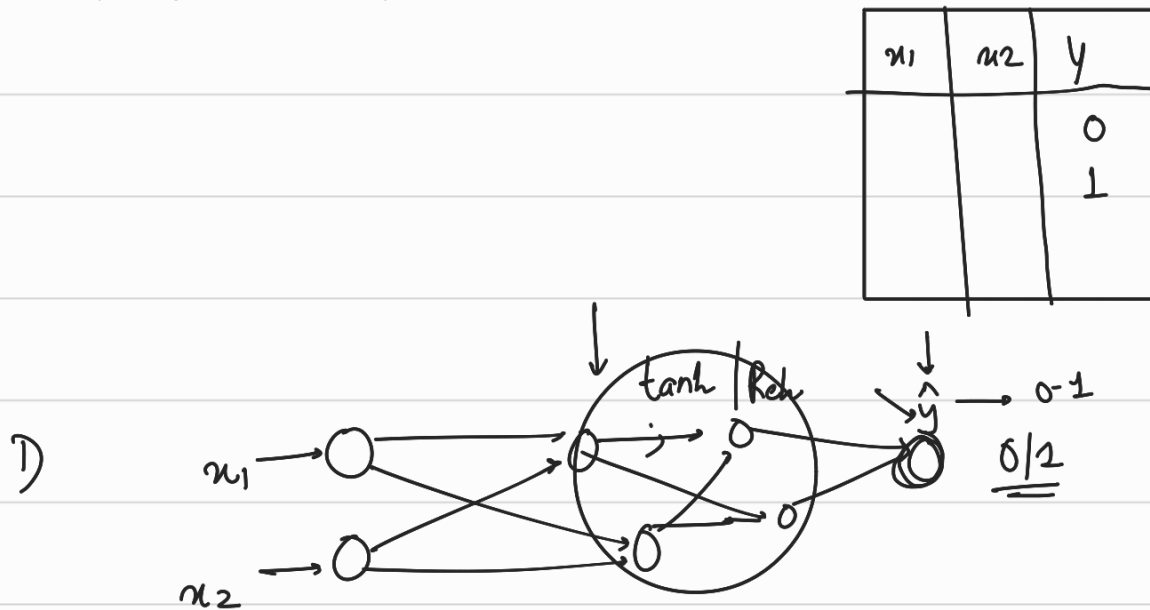
We forward propagate, calculate the loss (mse for regression)

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):



Changes: 1) activation function at the O/L.

Sigmoid  $\rightarrow$

2) Loss  $\rightarrow$  diff ( $y, \hat{y}$ )

$0/1, \hat{y} \rightarrow 0-1$

Log loss

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

Diagram illustrating the calculation of Log Loss for two examples:

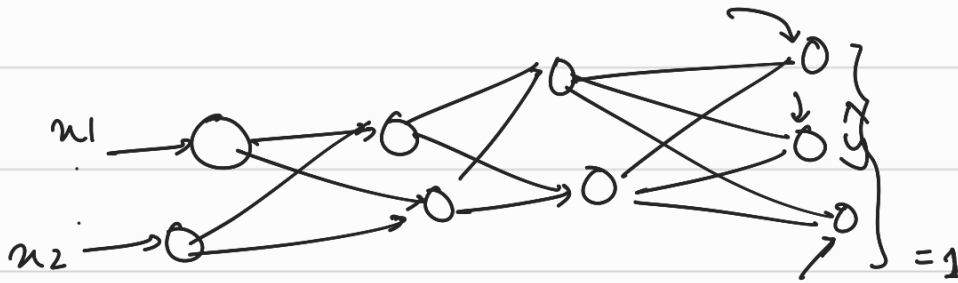
- Example 1:  $y_i = 1, \hat{y}_i = 0.95$   
 $y_i \log \hat{y}_i = 1 \cdot \log(0.95) = -0.0513$   
 $(1 - y_i) \log (1 - \hat{y}_i) = 0 \cdot \log(0.05) = 0$   
 $\text{Log Loss} = -(-0.0513) = 0.0513$  (labeled as 0.022 in the diagram)
- Example 2:  $y_i = 0, \hat{y}_i = 0.05$   
 $y_i \log \hat{y}_i = 0 \cdot \log(0.05) = 0$   
 $(1 - y_i) \log (1 - \hat{y}_i) = 1 \cdot \log(0.95) = -0.0513$   
 $\text{Log Loss} = -(-0.0513) = 0.0513$  (labeled as 0.022 in the diagram)

Another example shown:  $y_i = 1, \hat{y}_i = 0.2$   
 $y_i \log \hat{y}_i = 1 \cdot \log(0.2) = -1.6094$   
 $(1 - y_i) \log (1 - \hat{y}_i) = 0 \cdot \log(0.8) = 0$   
 $\text{Log Loss} = -(-1.6094) = 1.6094$  (labeled as 0.69 in the diagram)

2) Multiclass Problem:

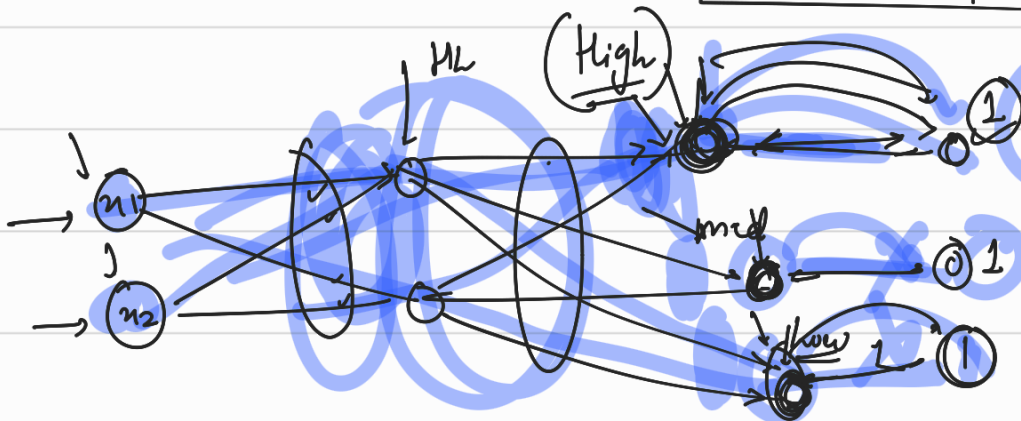
$x_1$	$x_2$	$y$
		High
		Med
		Low

X



One hot encoding  
↓

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



$$\hat{y} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.33 & 0.33 & 0.33 \end{bmatrix}$$

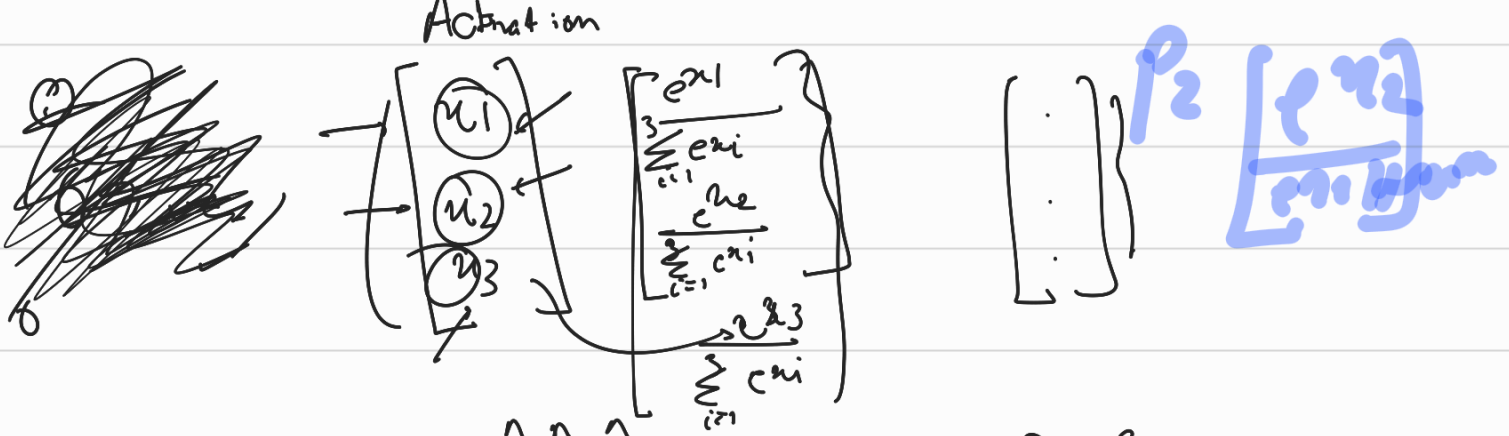
Softmax

$$H \quad M \quad L$$

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

$P_1$

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



$i=0$   $\hat{y}_i = \begin{bmatrix} p_1 & p_2 & p_3 \\ 0.8 & 0.1 & 0.1 \end{bmatrix}$   $y = \begin{bmatrix} p_1 & p_2 & p_3 \\ 0 & 1 & 0 \end{bmatrix}$

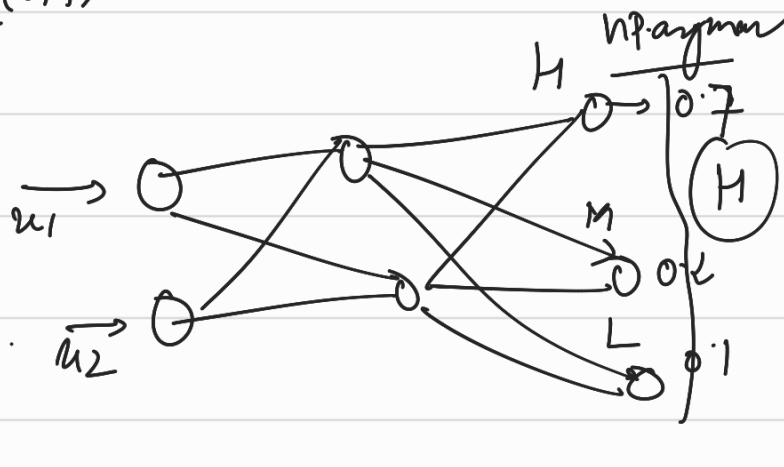
Log loss  $= - \left( p_1 \times \log \hat{p}_1 + p_2 \times \log \hat{p}_2 + p_3 \times \log \hat{p}_3 \right)$

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

	$x_1$	$x_2$	$y$			$\hat{y}$			
			H	M	L	H	M	L	
①	✓	✓	①	0	0	0.2	0.3	0.5	$-\log 0.2$
②	✓	✓	②	1	0	0.3	0.6	0.1	$-\log 0.6$
③	.	.	.	.	.	.	.	.	$\vdots$
									$\sum_{i=1}^n (-\log \hat{p}_i)$

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

Training  $\frac{dL}{dy}$



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

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

$$p_1 =$$

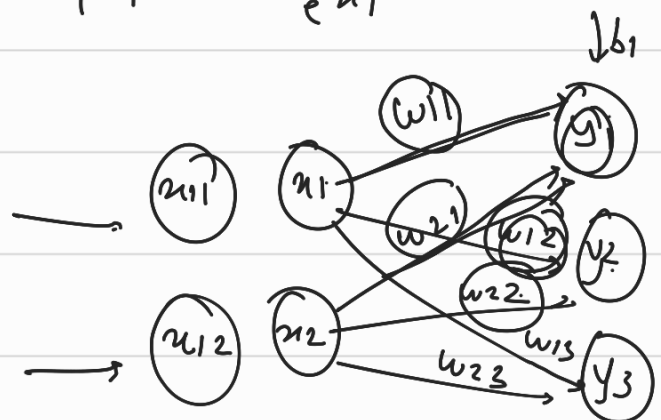
$$p_1 + p_2 = 1$$

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

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

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

$$x = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \vdots \\ x_{2001} & x_{2002} \end{bmatrix}$$

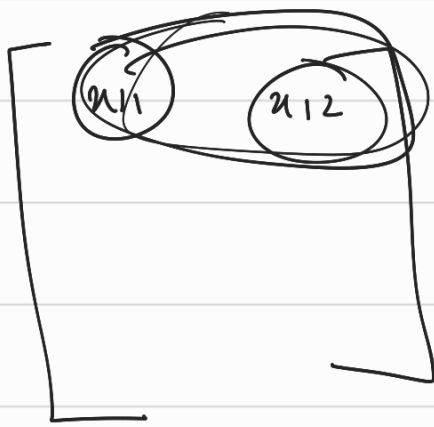


$$x_1$$

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

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

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



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

The diagram illustrates a matrix multiplication and addition operation. A 2x3 weight matrix is added to a 1x3 bias vector to produce a 1x3 output vector. The weights are arranged in two columns: the first column contains  $w_{11}$  and  $w_{21}$ , the second column contains  $w_{12}$  and  $w_{22}$ , and the third column contains  $w_{13}$  and  $w_{23}$ . The bias vector contains  $b_1$ ,  $b_2$ , and  $b_3$ . The output vector contains  $y_{11}$ ,  $y_{12}$ , and  $y_{13}$ .

$x_1$	$x_2$	$y$
2	3	1

label encoded.

$$-\log \left[ \frac{1}{0.6 + 0.2 + 0.2} \right]$$

The diagram shows the calculation of the negative log-likelihood for a specific output. The denominator is the sum of the probabilities for all possible outputs, which are 0.6, 0.2, and 0.2. The probability for the correct output (1) is 0.2.

