Regression

Problem

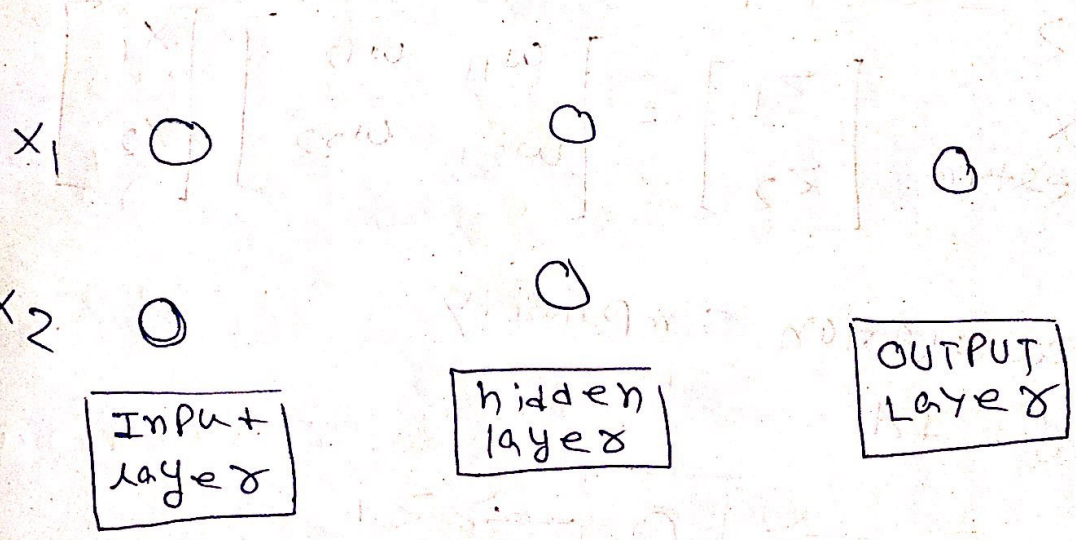
x_1	x_2	y
1.1	2	7
7.2	4	9

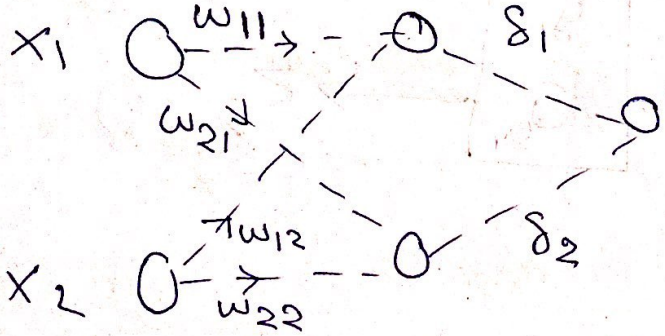
sample data

we ~~created a~~ will name to create a Network

Input size = 2 [since only two features]

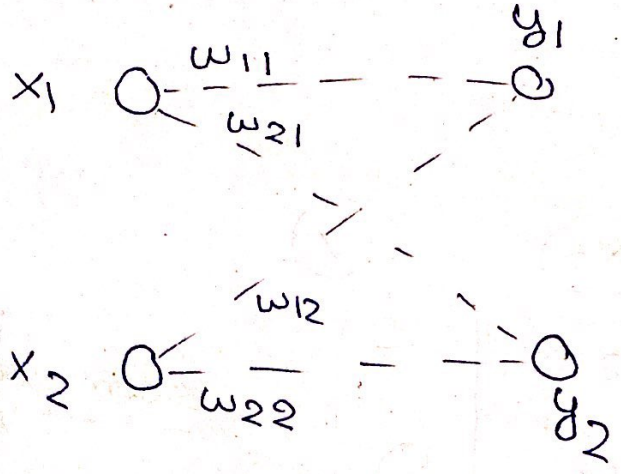
output size = 1 [dim(y) = 1]

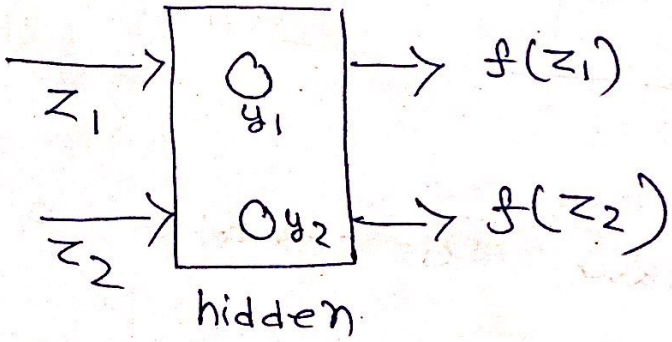




No bias
so # parameters
= 6

w_{ij} = ith node of the hidden layer
" " " input layer





$$z_1 = w_{11}x_1 + w_{12}x_2$$

$$z_2 = w_{21}x_1 + w_{22}x_2$$

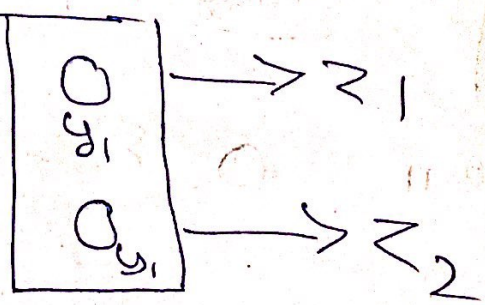
matrix representation

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

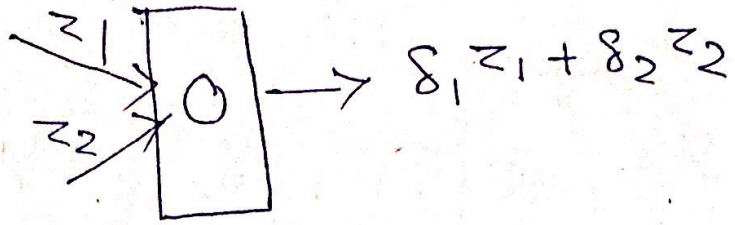
For calculation simplicity

$f = Id$

$\Rightarrow$



OUTPUT NODE



True y predicting $\delta_1 z_1 + \delta_2 z_2$

$$\text{LOSS} = \left[y - (\delta_1 z_1 + \delta_2 z_2)\right]^2 = \alpha$$

$$\text{LOSS}\left[w, \delta_1, \delta_2\right]$$

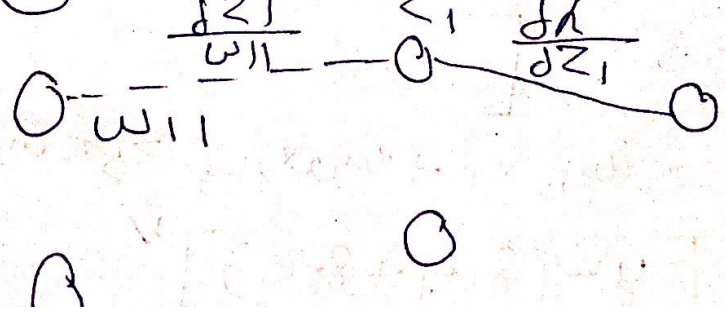
Update : [SGD]

$$\frac{d\alpha}{d\delta_1} = 2\left[\delta_1 z_1 + \delta_2 z_2 - y\right] z_1$$

$$\frac{d\alpha}{d\delta_2} = 2\left[\delta_1 z_1 + \delta_2 z_2 - y\right] z_2$$

we need to update w as well,

but w is related to α directly



$$\frac{\partial d}{\partial w_{11}} = \frac{\partial d}{\partial z_1} \frac{\partial z_1}{\partial w_{11}}$$

$$\frac{\partial d}{\partial z_1} = 2\left[\delta_1 z_1 + \delta_2 z_2 - y\right] \delta_1$$

$$\frac{\partial z_1}{\partial w_{11}} = x_1$$

$$\Rightarrow \frac{\partial d}{\partial w_{11}} = 2\left[\delta_1 z_1 + \delta_2 z_2 - y\right] \delta_1 x_1$$

~~6: until~~

while (not converged):

$$\delta_1 = \delta_1 - \eta \frac{\partial d}{\partial \delta_1}$$

$$\delta_2 = \delta_2 - \eta \frac{\partial d}{\partial \delta_2}$$

$$w_{11} = w_{11} - \eta \frac{\partial d}{\partial w_{11}}$$

$$\vdots$$

$$w_{22} = w_{22} - \eta \frac{\partial d}{\partial w_{22}}$$

~~Loss calculate Loss~~

~~Loss = [~~

$z_1 = w_{11} x_1 + w_{12} x_2$; $z_2 = w_{21} x_1 + w_{22} x_2$ ✓

$d = \left[y - (\delta_1 z_1 + \delta_2 z_2)\right]$

Classification:

Problem

x_1	x_2	y
1.1	2	0
7	9.8	1

sample data

0 → class 1

1 → class 2

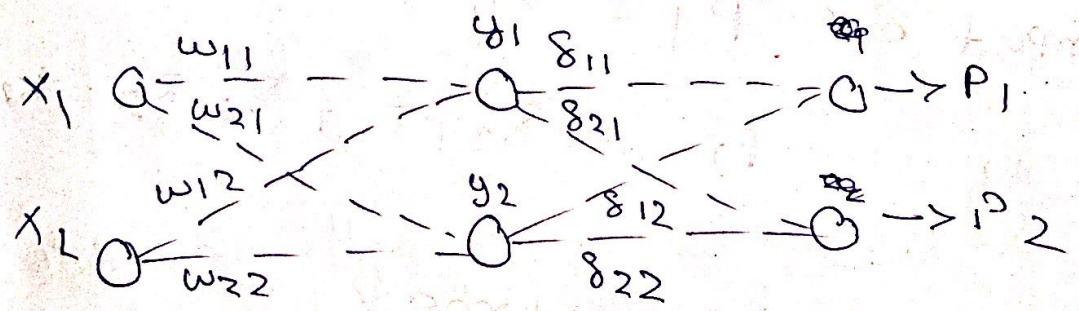
Input size → 2

Output size → 2 [2 classes are there]

we modify the vector or output y

0 → [1, 0]

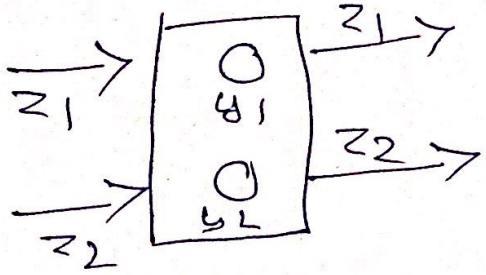
1 → [0, 1]



Probability

P_1 → of class 1

P_2 → Probability of class 2



$$z_1 = w_{11} x_1 + w_{12} x_2$$

$$z_2 = w_{21} x_1 + w_{22} x_2$$

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

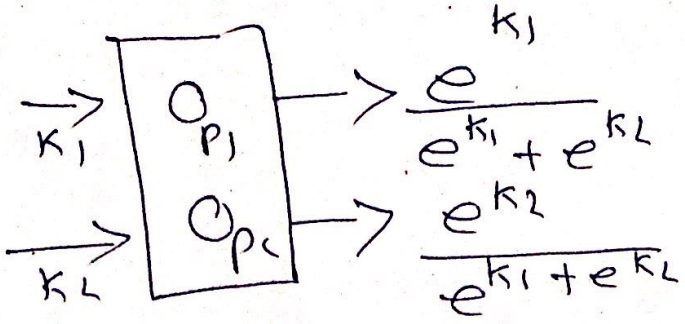
Output NODE:

Input of $P_1 = \delta_{11} z_1 + \delta_{12} z_2 = k_1$

" " $P_2 = \delta_{21} z_1 + \delta_{22} z_2 = k_2$

Finally apply softmax

Output Node



$$(a, b) \rightarrow \left(\frac{e^a}{e^a+e^b} \quad \frac{e^b}{e^a+e^b}\right)$$

$$(K_1, K_2) \rightarrow \left(\underset{\tilde{P}_1}{\frac{e^{K_1}}{e^{K_1}+e^{K_2}}} \quad \underset{\tilde{P}_2}{\frac{e^{K_2}}{e^{K_1}+e^{K_2}}}\right)$$

Loss function:

$\tilde{P}_1$ probability of class 0

$\tilde{P}_2$ " " " 1

~~if the actual~~

y	S_1	S_2
0 →	1	0
1 →	0	1

$$\text{Loss} \rightarrow -\left[S_1 \log \tilde{P}_1 + S_2 \log \tilde{P}_2\right]$$