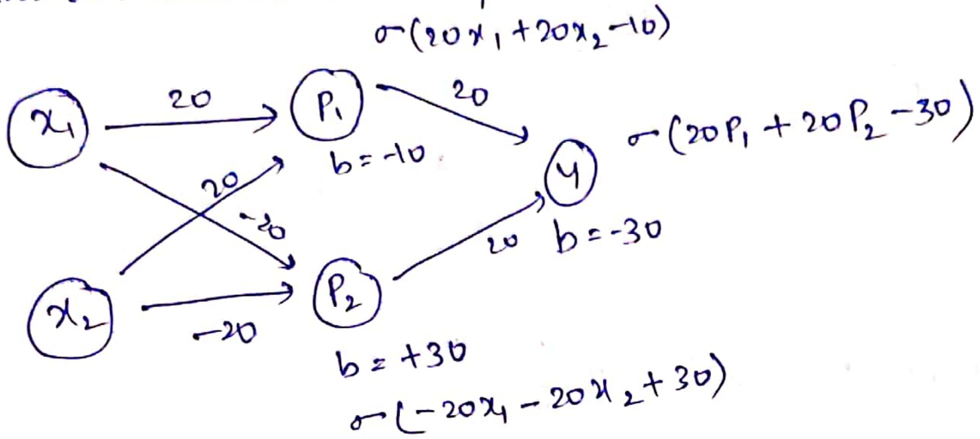
Neural Network

1. a) Considering following neural net with a hidden layer (weights and bias are considered as per convienience)



Truth Table :-

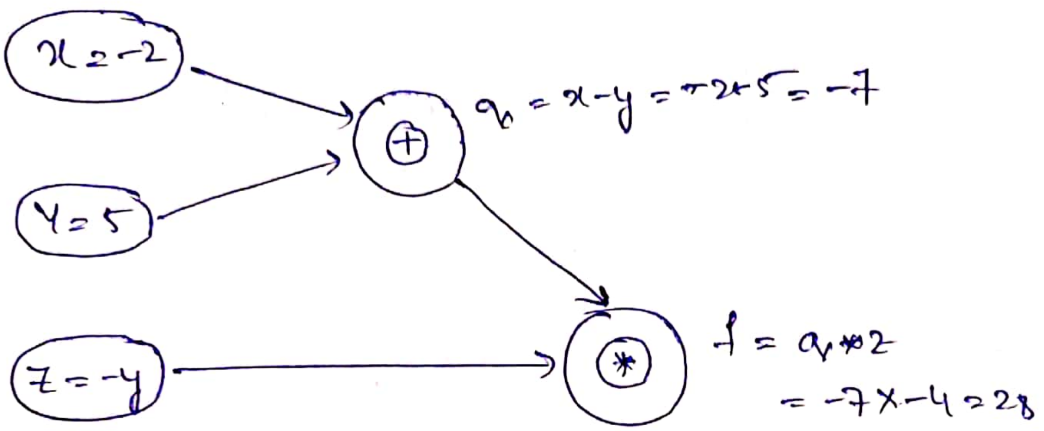
x_1	x_2	P_1	P_2	y
0	0	$\sigma(20(0)+20(0)-10) = 0$	$\sigma(-20(0)-20(0)+30) = 1$	$\sigma(20(0)+20(1)-30)=0$
1	1	$\sigma(20(1)+20(1)-10)=1$	$\sigma(-20(1)-20(1)+30) = 0$	$\sigma(20(1)+20(0)-30)=0$
1	0	$\sigma(20(1)+20(0)-10)=1$	$\sigma(-20(0)-20(0)+30)=1$	$\sigma(20(1)+20(1)-30)=1$
0	1	$\sigma(20(0)+20(1)-10)=1$	$\sigma(-20(0)-20(1)+30) = 1$	$\sigma(20(1)+20(1)-30) = 1$

Simplifying the above table we get

x_1	x_2	y	T/F
0	0	0	False
1	1	0	False
1	0	1	True
0	1	1	True

Hence XOR function returns true only when one of the arguments is true & other is false, otherwise it returns false. Hence XOR is solved.

b) Graphical Representation :-



Gradient of f wrt x, y, z :-

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}((x-y)*z) = z = -4$$

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}((x-y)*z) = -z = 4$$

$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z}((x-y)*z) = x-y = -7$$

$$\therefore \frac{\partial f}{\partial x} = -4 \quad \frac{\partial f}{\partial y} = 4 \quad \frac{\partial f}{\partial z} = -7$$

2) Given Extension of cross entropy error function for multiclass classification problem ⇒

$$E(w) = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{kn} \ln y_k(x_n, w) \quad -①$$

Here k = no of classes

N = No of data samples

Also given $y_k(x_n, w) = P(t_k = Y_x) = \dfrac{\exp(a_k(x,w))}{\sum \exp(a_k(x,w))}$ $-②$

Here $0 \le y_k \le 1$ $\quad \sum y_k = 1$

and a_k are presoftmax activations of output layer neurons (called logits)

Now, consider

$$\frac{\partial E}{\partial w_{ki}} = \sum_{c=1}^{C} \frac{\partial E}{\partial y_c} \cdot \frac{\partial y_c}{\partial a_k} \cdot \frac{\partial a_k}{\partial w_{ki}} \quad - (3)$$

Here, $\frac{\partial a_k}{\partial w_{ki}} = x_i$ (as usual) — (4)

and $\frac{\partial E}{\partial y_c} = \frac{-t_c}{y_c}$ — (5)

Now look at $\frac{\partial y_c}{\partial a_k}$ we look at two cases

when $c=k$ and when $c \neq k$

Case (i) When $c=k$, we apply the quotient rule of differentiation

$$\frac{\partial y_k}{\partial a_k} = \frac{\sum \exp(a_j)\exp(a_c) - \exp(a_c)\exp(a_c)}{(\sum \exp(a_j))^2}$$

$$\boxed{\frac{\partial y_c}{\partial a_k} = y_c - y_c^2 = y_c(1-y_c)} \quad - (6)$$

Now when $c \neq k$

$$\frac{\partial y_c}{\partial a_k} = \frac{-\exp(a_c)\exp(a_k)}{(\sum \exp(a_j))^2} = -y_c y_k \quad - (7)$$

We can combine these using the kronecker delta,

$\delta_{ck} \Rightarrow$ ($\delta_{ck} = 1$ when $c=k$ &
$\delta_{ck} = 0$ when $c \neq k$)

$$\boxed{\therefore \frac{\partial y_c}{\partial a_k} = y_c(\delta_{ck} - y_c)} \quad — (8)$$

Now substituting eq (4), (5), (8) in eq (3) we get

$$\frac{\partial E}{\partial w_{ki}} = \sum_{c=1}^{C} \frac{-t_c}{y_c} \times y_c(\delta_{ck} - y_c) \times x_i$$

$$\frac{\partial E}{\partial w_{ki}} = \sum_{c=1}^{C} + t_c(y_c - \delta_{ck})\, x_i \quad — (9)$$

since here given $t_n = [0, 0, 1, \ldots\ldots]$

so $\boxed{\sum t_n = 1}$ everytime in any expected output

$\therefore$ eq (9) can be written as

$$\frac{\partial E}{\partial w_{ki}} = (y_c - \delta_{kc})\, x_i \quad — (10)$$

also, $\boxed{\delta_{kc} = t_c}$ $\rightarrow$ target

$$\frac{\partial E}{\partial w_{ki}} = (y_c - t_c)\, x_i$$

$$\frac{\partial E}{\partial a_k} \cdot \frac{\partial a_k}{\partial w_{ki}} = (y_c - t_c)\, x_i$$

$$\frac{\partial E}{\partial a_k} \cdot \cancel{\frac{\partial a_k}{\partial w_{ki}}} = (y_c - t_c)\, \cancel{\frac{\partial a_k}{\partial w_{ki}}}$$

$$\boxed{\frac{\partial E}{\partial a_k} = (y_k - t_k)} \quad — (11)$$

Hence Proved.

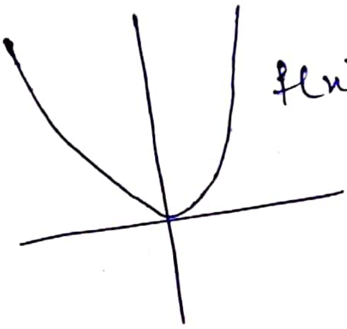
3) Consider a convex fuction $f(x)=x^2$ and sum of squares error

$$E_{AV} = \frac{1}{M}\sum_{m=1}^{M} E_x\left[(y_m(x) - f(x))^2\right]$$

of the members of an esemble model and the expected error of the esembly satisfy

$$E_{SNS} = E_x\left[\frac{1}{M}\sum_{m=1}^{M} (y_m(x) - f(x))^2\right]$$

$$E_{SNS} \leq E_{AN}$$



$f(x)=x^2$

let we know that

$$E_{SNS} = E_x\left[\frac{1}{M}\sum_{m=1}^{M} (y_m(x) - f(x))^2\right]$$ and

$$E_{AV} = \frac{1}{M}\sum_{m=1}^{M} E_x\left[(y_m(x) - f(x))^2\right]$$

Now rearranging the both values

$$E_x\left[\frac{1}{M}\sum_{m=1}^{M}(y_m(x) - f(x))^2\right] \leq \frac{1}{M}\sum_{m=1}^{M} E_x\left[(y_m(x) - f(x))^2\right]$$

Because all terms of E_{SNS} is containers. in E_{AN} and

hence proved

$$E_{SNS} \leq E_{AN}$$

It is hold for any error function $E(y)$ not just sum of squares.