Date......../....../..........

HW 3

ISHMEET KAUR
2015042

ques 1. No. XOR is not linearly seperable.
Neural network with linear activation can't classify it

$y = f(w_2(w_1 x + b_1) + b_2)$ is SVM with linear function

$f(x) = B(Ax + a) + b$
$= \underbrace{BA}_{const}x + \underbrace{Ba + b}_{const}$

$\therefore f(x) = C_1 x + C_2$

ques 2 Relu : performs better than sigmoid
- large gradient results in update weight that gradient flowing could be 0.
- this kills neurons
- So, keep the learning rate low.

Sigmoid : when sigmoid neuron's activation saturates at 0/1, gradient → 0
ie sigmoid saturation and gradients killed
: no flow through neurons
: Initialise weight by extra 1
: start with low weights

ques 3. For quadratic, error ↑, Learning rate ↓

$C = \frac{(y - a)^2}{2}$

$a = \sigma(z)$

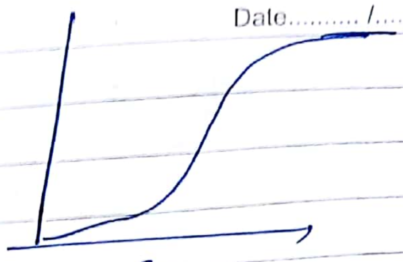
$\frac{\partial C}{\partial w} = (a - y)\,\sigma'(z)\,x = a\sigma'(z)$

"The best Vitamin to be a Happy"

Evergreen

Date......../....../..........

$$\frac{\partial C}{\partial b} = a\sigma'(z)$$



$\therefore \frac{\partial C}{\partial w}, \frac{\partial C}{\partial b}$ are v. low (slow learning)

cross entropy const

$$C = -\frac{1}{n}\sum_x [y \ln a + (1-y)\ln(1-a)]$$

I. $C > 0$

II. output closer to actual value then entropy $\to 0$.

$$\Rightarrow \frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum\left(\frac{y}{\sigma(z)} - \frac{(1-y)}{1-\sigma z}\right)\sigma'(z)y$$

$$= \frac{1}{n}\sum_x x_j(\sigma(z) - y).$$

$\therefore$ learning rate of weights $= \sigma(z) - y$

large error $\Rightarrow$ more learning !!