<u>ML Assignment - 3</u>

Solution C :

1. ReLU is the best activation function out of all activation functions because the model gave the highest accuracy when on it.
   analysis :
   accuracy = 0.76 (activation func = sigmoid)
   accuracy = 0.86 (activation func = relu)
   accuracy = 0.75 (activation func = tanh)
   accuracy = 0.80 (activation func = linear)

   (on validation set)
   accuracy = 0.81 (activation func = sigmoid)
   accuracy = 0.84 (activation func = relu)
   accuracy = 0.74 (activation func = tanh)
   accuracy = 0.81 (activation func = linear)

2. Learning rates = [0.1, 0.01, 0.001]
   accuracy = 0.81, at alpha = 0.1 (its fast, can miss the optimum points)
   accuracy = 0.84, at alpha = 0.01 (it's pretty much faster than) previous)
   accuracy = 0.56, at alpha = 0.001 (it is too slow)

   So, the 0.01 is the best learning rate

3. When we decrease the number of neurons in each layer it leads to decrease the time taken of the model to train it.
   As we see in the all three plots at different hidden layers (252,30), (244,30), (228,18) when hidden layers decrease, train loss decreases sharply as the number of epochs increases.

4. This algorithm is used to find out the best parameters for the model in order to get better predictions.

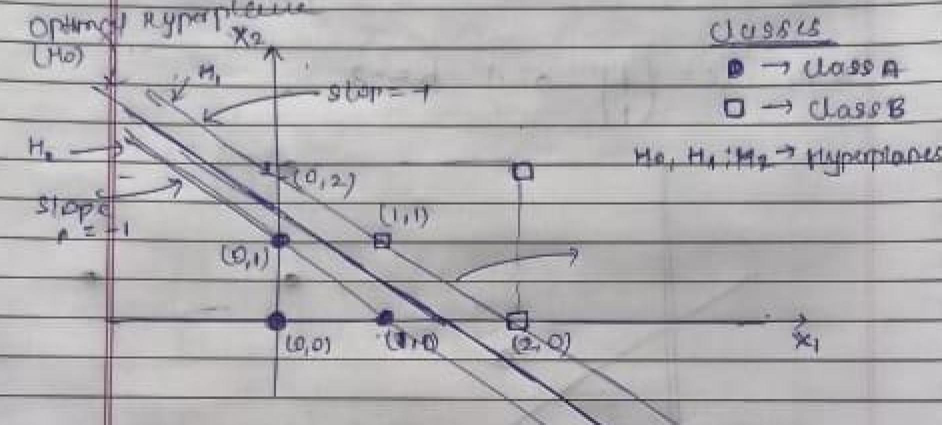
   accuracy = 0.72
   accuracy = 0.73 (on validation set)

Solution A :

Solution 1 :

Class	x_1	x_2
A	0	0
A	1	0
A	0	1
B	1	1
B	2	2
B	2	0

part 1)



Yes, we can see easily in the above graph the points are linearly separable.

part 2) Mid-point of pts $(1,0)$ & $(2,0)$ $= \left(\frac{1+2}{2}, \frac{0+0}{2}\right)$

$= (3/2, 0)$

Mid-point of pts $(0,1)$ & $(0,2)$ $= \left(\frac{0+0}{2}, \frac{1+2}{2}\right)$

$= (0, 3/2)$

RANKA
DATE / /
PAGE

eqn for H_0 :

pts $(3/2, 0)$ & $(0, 3/2)$

$$(y-0) = \frac{(3/2-0)}{(0-3/2)} \times (x-3/2)$$

$$y = -1(x - 3/2)$$

$$-y = x + \frac{3}{2}$$

$$\boxed{x + y = 3/2}$$

$$\Rightarrow \cancel{x_1 + x_2 + 0 = 3/}$$

$$x_1 + x_2 - 3/2 = 0$$

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 3/2 = 0$$

$$w \cdot x + b = 0$$

$$\therefore \; w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; b = -3/2$$

So, the weight vector corresponding to the max^m margin hyperplane is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

support vectors = $(0,1), (1,0), (1,1)$ & $(2,0)$

Part 3) ~~[illegible]~~ the optimal margin for the given dataset get increase when we ~~[illegible]~~ remove the either of support vectors (1, 0 or (1,1) and remain same ~~for~~ when either of ~~the~~ the support ~~[illegible]~~ vectors got removed.

Part 4) In general, it is expected ~~[illegible]~~ that ~~the~~ the optimal margin will increase when we ~~[illegible]~~ remove support vectors. because whenever we remove ~~[illegible]~~ either support vectors it would not ~~[illegible]~~ impact the plot for the dataset. (~~[illegible]~~ problem).