ASSIGNMENT - 2

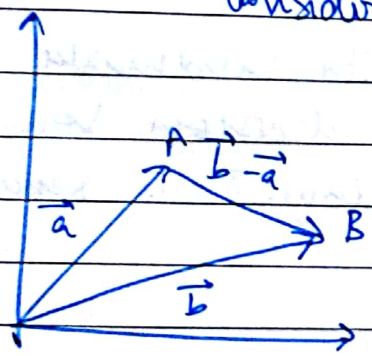
Q1

$$\min \frac{1}{2} |\theta|^2 + C \times \sum_{i=1}^{M} \xi_i$$

where $\xi_i = \max(0, y(\vec{\theta}^T x + b))$

The parameter C takes care of overfitting. Larger values of C train the classifier such that it has not false predictions thus reducing the margin. However this can lead to overfitting. Thus by taking smaller values of C we can reduce the effect of overfitting.

Q2



Considering there are only two points.

Now, if we have just two data points, then the hyper plane that would be predicted would be equidistant from ~~the~~ each of these datapoints.

To define a ~~p~~ hyper plane we need one point on the plane & the equation of the normal to the hyper plane.

The eqn of normal is given by,

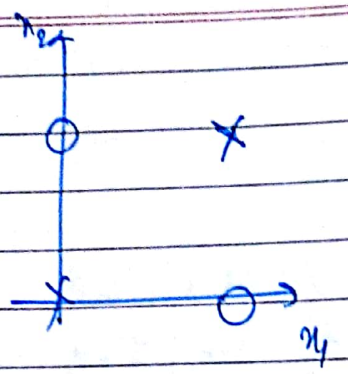
$$\frac{\vec{b} - \vec{a}}{|\vec{b} - \vec{a}|} = \hat{n}$$

the equation of the point on the plane is given by,

$$\left(\frac{x_{11} + x_{21}}{2}, \frac{x_{12} + x_{22}}{2}, \ldots, \frac{x_{1n} + x_{2n}}{2} \right)$$

$\therefore$ We can define our ~~b~~ hyperplane as $\hat{n} \cdot p = 0$

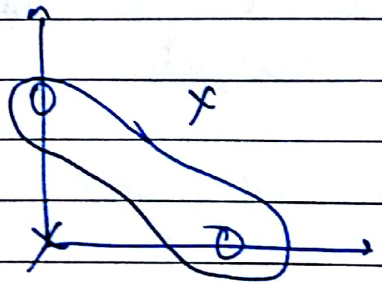
Q4.



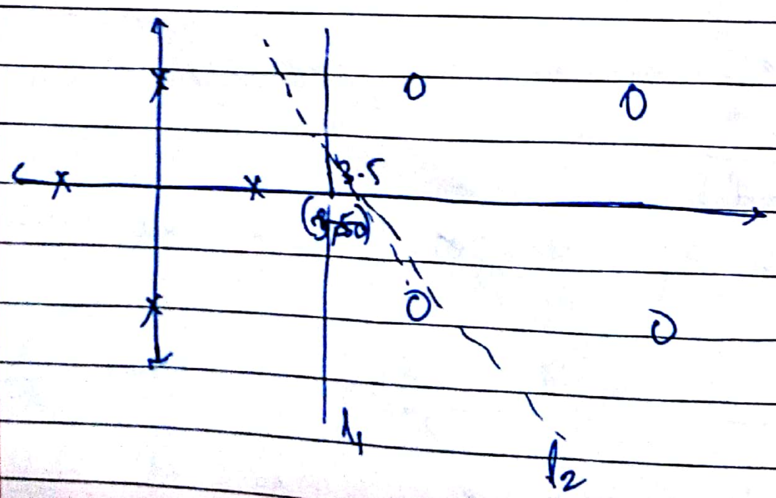
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

The data cannot be classified using SVM as it isn't linearly separable and whatever the decision boundary is there will always be at least one false classification. which out of just 4 points is very high. This can be modeled using a gaussian or a rbf kernel.

Using a gaussian kernel with landmarks same as the given points. The decision boundary would like as shown below. Thus separating them.



Q3.



The decision boundary is given by the line l_1

$$x = 3.5.$$

This gives the max. margin possible.

max margin = 5 - 2 = 3

When we remove the data point (5, -2) we would get a slanted classifier as shown by the curve l_2

max margin = $\sqrt{3^2 + 2^2} = \sqrt{13}$