Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

->

Maximal margin classifier classifies two classes based on a best line that maintains the largest possible equal distance from the nearest points of both the classes.

This will do great on a training set but on test/ unseen data it will not perform very well, as they may have outliers and its a drawback of it.

Meanwhile the Soft margin classifier overcomes the drawbacks of the above mentioned by allowing some of the observations to be misclassified, as there are supposed to be outliers.

And we can control the amount of misclassifications using C, which represents the maximum value of the summation of the slack variable.

According to the values of C we can make sure the models flexibility and generalization, over fitting, bias and all.

Question 2

What does the slack variable Epsilon (ε) represent?

➔

A slack variable is ideally used to control the misclassifications; it identifies the points / observations, (where they lie?) with respect to margin and hyperplane. Each observation has a slack associated with it, it typically lies between 0 and +infinity.

The lower values are the more preferable ones; $\varepsilon = 0$ represents a correct classification and if its greater than that, it implies an incorrect one.

Values meaning:

1. If observations are at a safe distance from hyperplane then slack = 0.
2. If observations are classified correctly but is inside the margin(or not), then the value is between 0 and 1.
3. And for incorrect classifications its > 1.

Question 3

How do you measure the cost function in SVM? What does the value of C signify?

->

This could be the cost function:

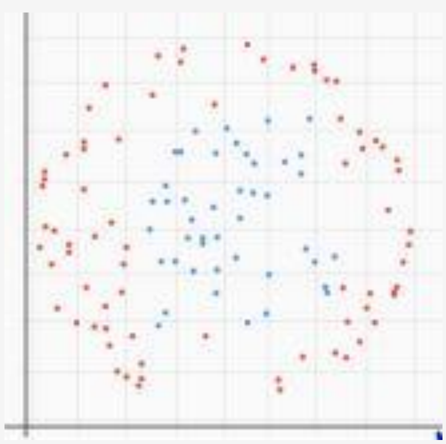
(li * (Wi . Yi)) ⩾ M (1 - **ϵi**)

∑ϵi ≤ C

This is how we measure the cost function in SVM.

Value of C signifies: when C is very small, the tolerance for error is very small, optimization is strict separator. On the other hand if C is large then lot of errors are going to loose, anything is exactable (bais high), it will create a generalized model.

Question 4



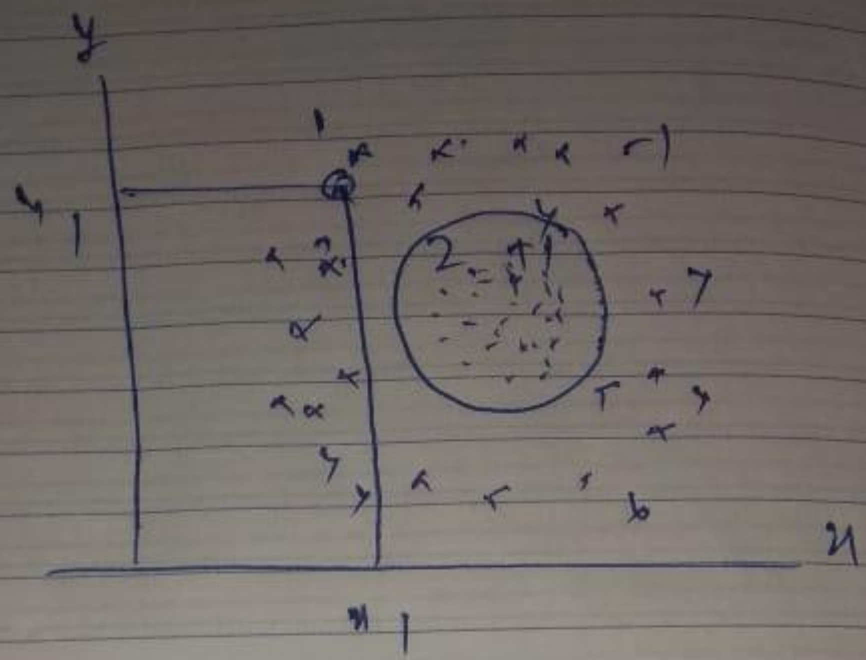
Given the above dataset where red and blue points represent the two classes, how will you use SVM to classify the data?

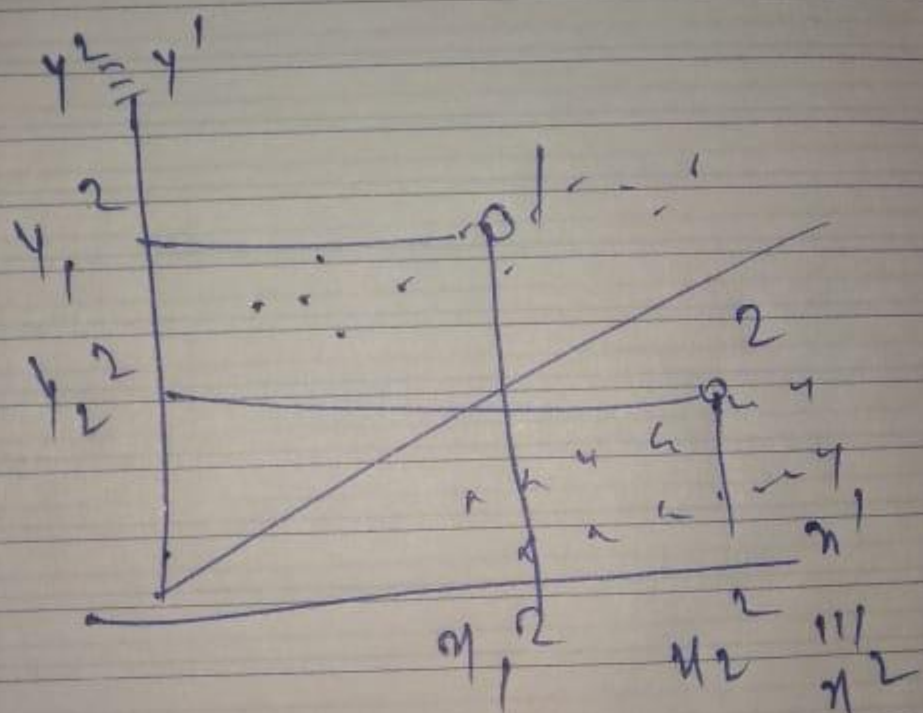
->

By using the boundary classification we will apply or use SVM on the transformed data (feature space).

We have to transform the non linear data to linear so that SVM can be applied/used.

To apply boundary transformation we represent all the points on the mentioned graph on a transformed scale by squaring them up and that will eventually represent a straight line as separator, as shown in the equation below the original equation of circle or ellipse becomes a line after performing this.





$$\frac{x^2}{a} + \frac{y^2}{b} = c \Rightarrow \frac{x^1}{a^1} + \frac{y^1}{b^1} = c^1$$

Question 5

What do you mean by feature transformation?

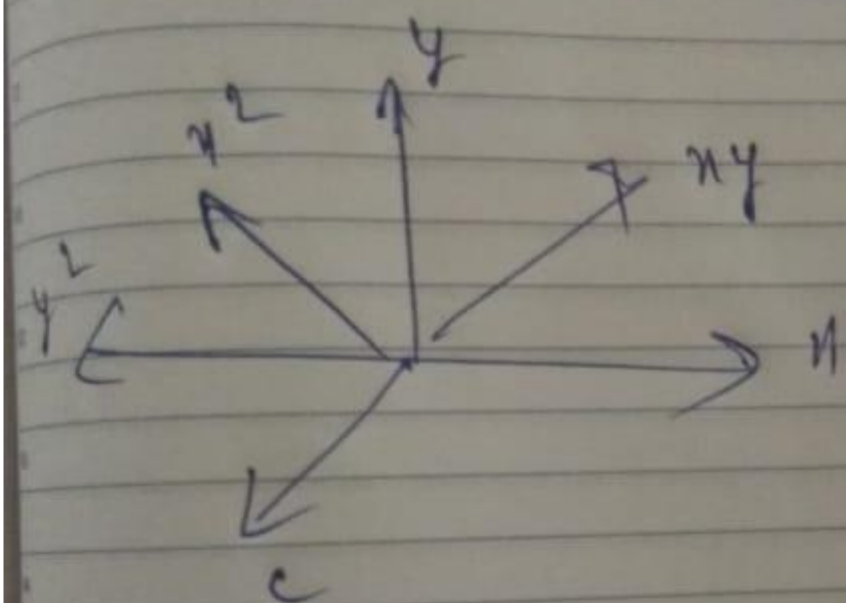
- ➔ Feature transformation is the process of transforming the original attributes into a new feature space.

Like for example we did the same in the question above,

Let say we want a variable x to be feature transformed to quadratic following will be the features of it, which would be computationally expensive though.

The quadratic equation

$$ax^2+by^2+cxy+dx+ey+fc=0$$



features

Now transformation is

$$(x,y) = (x^2, xy, y^2, x, y, c)$$