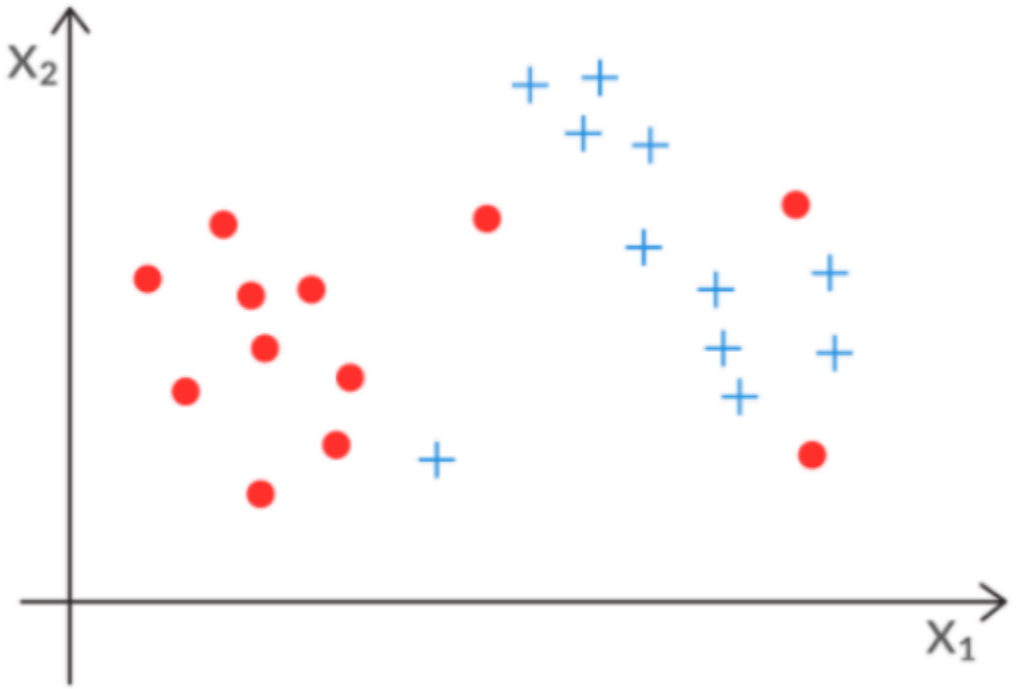
Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Maximum Margin Classifier considers the hyperplane which is the largest possible equal distance from the nearest points of the hyperplane between the classes. This is beneficial if the points are clearly differentiated, but when the points are intermingled as given below, we use soft margin classifier.



Soft margin classifier i.e. the support vector classifier classifies the data points such that some of the data points are intentionally misclassified considering only the points which are closer to the hyperplane.

Question 2

What does the slack variable Epsilon (ε) represent?

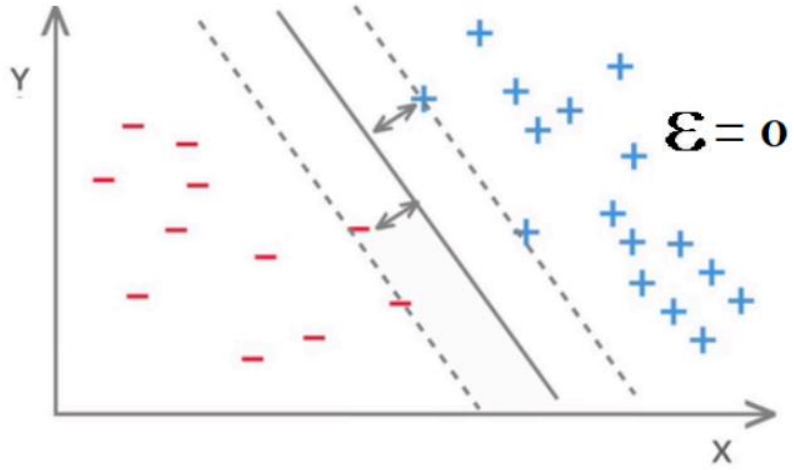
The Maximum Margin Classifier equation is given by :

$$(l_i X(W_i.Y_i)) \geqslant M,$$

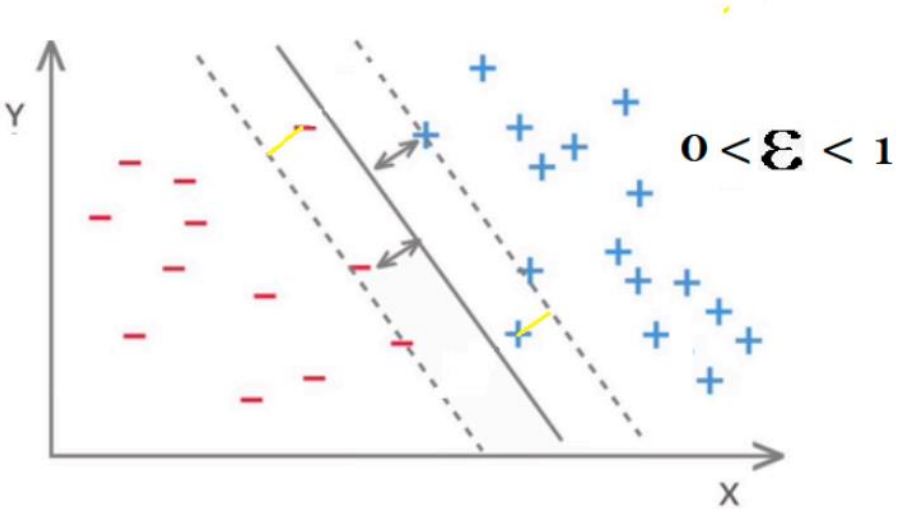
Hence, we are introducing a level of misclassifications, the slack variable ε is introduced to let the Classification to literally slack a little bit, thereby allowing it to misclassify.

$$(l_i X(W_i.Y_i)) \geqslant M\ (1 - \varepsilon)$$

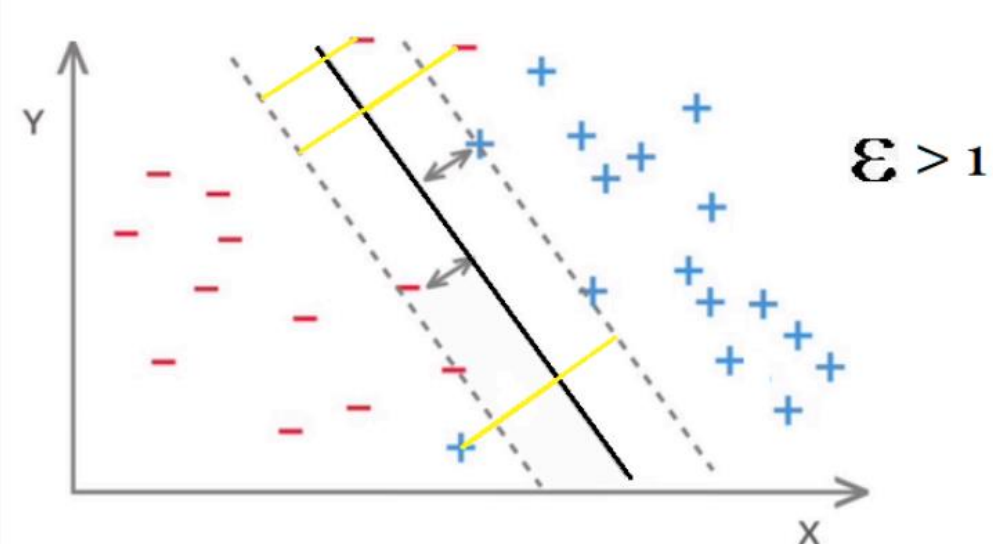
ε takes the value 0 when there are no misclassifications



ε takes the value between 0 and 1 i.e. still residing in the limits of hard margin, when the points are classified properly but lies within the margin.



ε takes the value greater than 1, violating the hard margin and gets misclassified.



Question 3

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

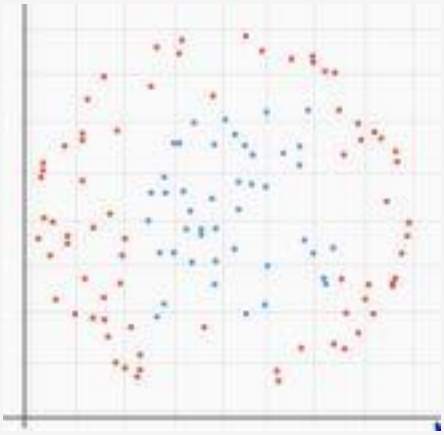
The cost function C is measured by the sum of all the slack variables,

$$\sum \epsilon_i \leq C$$

The cost function has a positive linear relationship with the value ϵ, i.e., when C is large, ϵ is also large hence there are more misclassifications leading to a higher bias and lower variance, thus this model is never overfitted leading to a very simple model and vice versa, when C is small , ϵ is also small hence there are lesser misclassifications leading to a lower bias and higher variance, thus this model is complex and overfit.

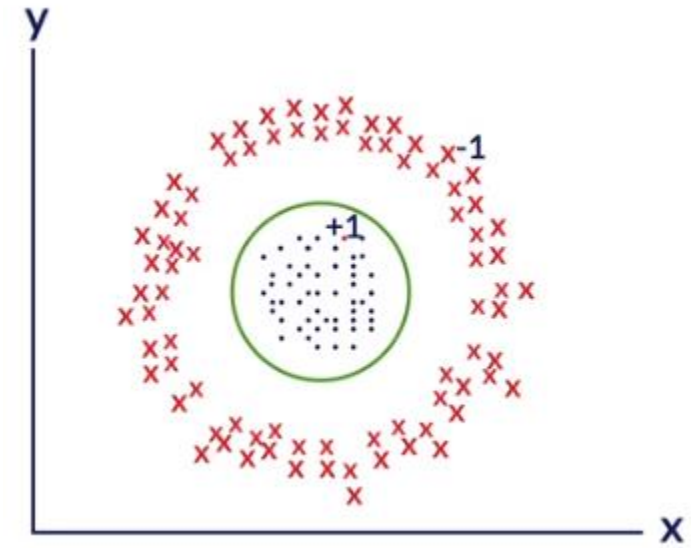
The ideal value of C should not be too high or too low.

Question 4



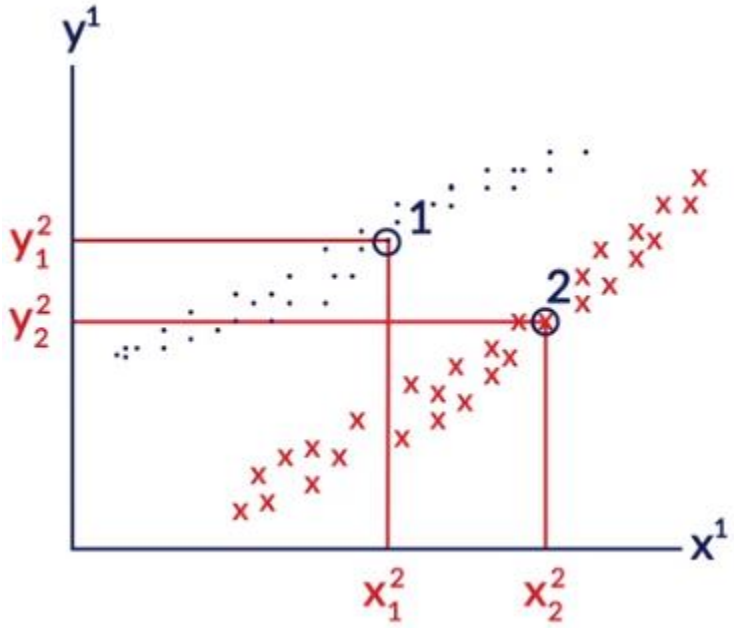
Given the above dataset where red and blue points represent the two classes, how will you use SVM to classify the data?

The aforementioned model is a non-linear model i.e of the form $x^2 / a + y^2 / b = c$ is a non linear kind of data which has a non linear boundary as given below:



Separating the boundary with a hyperplane in this kind of data is difficult as this data set is not linear.

Thus, to mark a clear boundary with a proper hyperplane, we transform the given dataset space into a different space by transforming the features from x,y to x^2,y^2 by plotting:



Question 5

What do you mean by feature transformation?

Process of transforming original attributes to attributes in the feature space is called feature transformation. For any non-linear problem there is an equivalent n dimension space where the features can be transformed making a linear equation.