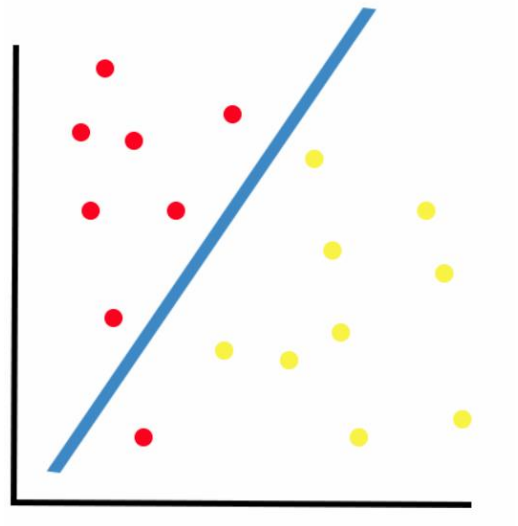
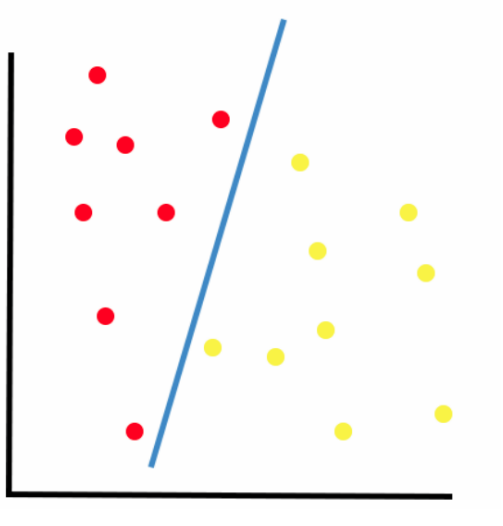
Assignment - Support Vector Machines

Part - II

Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer 1

Soft Margin Classifier	Maximum Margin Classifier
Here there is no strict rules to follow when it comes to drawing the hyperplane.	Here there is a strict rule to follow when it comes to drawing the hyperplane.
It allows some data points to cross into the other side of the margin.	It does not allow any data points to cross into the other side of the margin.
This is used in day to day situations since it is more practical.	This is not used in situations where accuracy is important since it is very theoretical.
This gives us a fair classification with a lower chance of over fitting	The classification here tends to over fit
While testing on an unseen data set, this model will get us more accuracy	While testing on an unseen data set, this model will get us lesser accuracy
	

[The graph represents two variables in each axis, the data points are show as dots and they are coloured depending on its classification. The blue line is the hyperplane that separates the two.]

Question 2

What does the slack variable Epsilon (ε) represent?

Answer 2

Every data point has a slack value(Ɛ). It is the Euclidian distance between the data point and the margin drawn. The slack value can lie between 0 and infinity. A lower slack value is better than higher values.

1. Ɛ = 0: This means than all the points are at a safe distance from the hyperplane and do not fall anywhere in the margin. (Usually the best situation since accuracy will be high)
2. 0< Ɛ<1: This means than there are a few points within the margin but are classified properly. (This is a fine and more practically occurring situation)
3. Ɛ>1: This means than there are a few points that has crossed the margin and are also not classified properly. (This model will provide a lesser accuracy)

Question 3

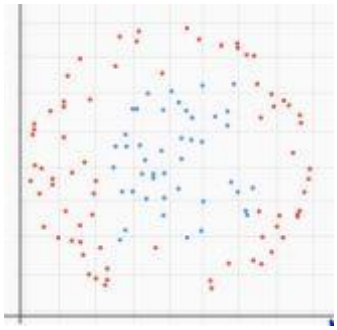
How do you measure the cost function in SVM? What does the value of C signify?

Answer 3

As the cost function increases more and more points gets classified in the wrong group. But this is required since otherwise we might tend to over fit the model. The cost function (C) controls the misclassifications.

1. If the value of cost function (C) is large – the slack variable (Ɛ) will also be large – Means there are a lot of misclassifications.
2. If the value of cost function (C) is small – the slack variable (Ɛ) will also be small – Means there are a very few misclassifications.

Question 4



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

Answer 4

It is clear from the graph above that it will not be possible to have a very accurate linear boundary. One way is to mathematically convert this into a linear dataset (using feature transformation) and then solve it easily.
The other method is using a kernel (which is usually used). It will create a margin according to the dataset or convert the existing non-linear data set into a linear one, thus allowing us to easily separate them. There are three types of kernels:

1. Linear kernel: It creates a linear hyperplane and splits the data.
2. Polynomial Kernel: It creates a polynomial hyperplane that splits the data.
3. Radial basis function (RBF): This creates elliptical decision boundaries and usually convert the radially divide dataset into a linear one.

In the given graph, using the RBF will be the best idea since the data set is radial in nature.

Question 5

What do you mean by feature transformation?

Answer 5

When we do not have a linear data set, we convert the non-linear data set into a linear one by increasing the dimensions. This way the points will end up getting located in a linear form but in many dimensions. This is called feature transformation. So, a two dimensioned data set might turn into a 15 dimensioned data set. This increase is usually exponential. That means doing this manually when we have data sets of 100+ attributes is not going to be possible. So we usually use kernel method for this process.