Problem Statement – II

Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer 1:

The maximal margin classifier is able to separate two classes perfectly, it will perform very well with the seen/training data but with unseen/test data it will perform poorly. Also there are chances where the two classes are not separated perfectly.

On the other hand soft margin classifier will allow data points to be deliberately misclassified by doing so it will able to classify most of the unseen data it is more robust and generalized.

Instead of accurately classify each and every point correctly it will allow some points to fall on the other side. The point which are close to the hyperplane are only considered for constructing the margin which are also called as support vectors.

Question 2

What does the slack variable Epsilon (ε) represent?

Answer 1:

Slack variable is a value which is attached to every data point, this is used to control the misclassification it will tell you where the data point lies according to margin and hyperplane.

The value of slack variable lies between 0 and +infinity.

Case 1: If you selected a support vector that do not allows misclassification such that the value of epsilon = 0 then all the point will be correctly classified but this will result in overfitting.

Case 2: If you selected a support vector in such a way where epsilon value is grater then 0 but less ten 1 then the point will fall near the hyperplane.

Case 3: if you selected a support vector where the epsilon value is grater then 1 then the point will be misclassified and will fall on the other side.

Question 3

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

Answer 3:

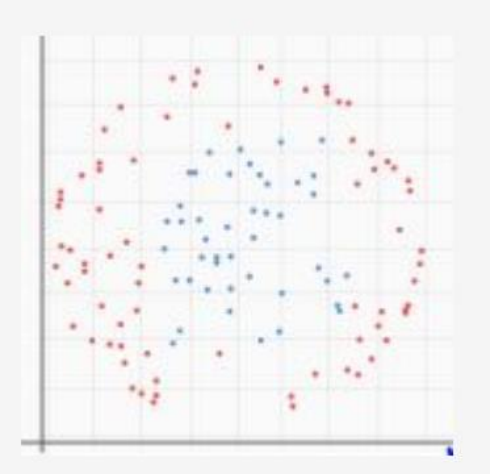
Cost function in SVM is denoted by 'C' it's basically summation of all the epsilon of each data point.

We can compare the models using the summation of epsilon which is the cost function, we will choose the model with the less cost.

Case 1: If C is large we are allowing the epsilon to be large in other words we are allowing the data points to be misclassified , this will reduce the model complexity and the model will less likely to over fit.

Case 2: if C is small we are restricting the epsilon to be small we are forcing the model to not allow the data point to fall on the wrong side , in this situation the bias will be less but the model variance will be high with some new data point the model will change drastically.

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: In the above scenario I will use radial basis function (RBF) kernel: This is the most complex one, which is capable of transforming highly nonlinear feature spaces to linear ones. It is even capable of creating elliptical (i.e. enclosed) decision boundaries.

Internally it uses the kernel trick normal computation is very expensive but in this case the learning algorithm only needs the inner product of the observations (XTi.Xj). It never uses the individual data points X1, X2 etc. in silo.

Question 5

What do you mean by feature transformation?

Answer 5:

Basically it is the process of transforming original attributes to a new feature space, with this there will be exponential increase in the dimension in the transformed feature space, Suppose

We have 4 variables transformation of polynomial degree 2 will result in 15 attribute in the transformed feature space with the increase in the number of feature the computation will become more expensive.