Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer=

Although Maximum margin classifier is an extended version of soft margin classifier but they are different from each other.

The one that maximizes the distance to the closest data points from both classes , is the hyper plane with maximum margin hence maximum margin classifier is the most efficient one in SVM.

However, In reality a data is complex and messy so cannot be separated perfectly with a hyper plane, the constraint of maximizing the margin of the line that separates the classes must be relaxed and this is called the soft margin classifier.

In soft margin classifier the values of epsilon approximate the number of misclassifications.

Question 2

What does the slack variable Epsilon (ε) represent?

Answer=

A slack variable is a variable that is added to an inequality constraint to convert it into equality one. These are used particularly in linear programming as with other variables in the constraints, it cannot take on negative values as algorithm requires them to be positive or Zero. If a slack variable is positive at a particular candidate solution, the constraint is non- binding there, as the constraint does not restrict the possible changes from that point. And if variable is zero at a particular solution the constraint is binding there, as constraints restrict possible changes from that point. If the point is negative at some point , the point is not allowed as it does not satisfy the constraint.

Question 3

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

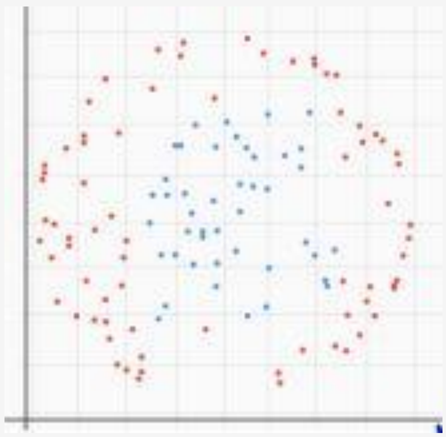
Answer=

$$\theta = C \sum_{i=1}^{m} \left[y^{(i)} cost_1(\theta^T x^{(i)}) + (1 - y^{(i)}) cost_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$$

Cost function is used to train SVM. By lowering theta value, it can be ensured that the SVM is as accurate as possible. In the above equation, functions cost1 and cost0 refer to an example where y=1 and cost for an example where y=0 . cost is defined by kernel(similarity) functions.

Signification of C = For large value of C, the optimization will choose a smaller-margin hyperplane, conversely , a very small value of C will cause the optimizer to look for a larger margin separating hyperplane, even if that hyperplane misclassifies more points. For very tiny values of C , we should get misclassified examples, even if our training data is linearly seperable.

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=

In above question, since it is a non-linear hyper plane , RBF(Radial Basis Function Kernel) will be used . RBF can map an input space in infinite dimensional space. Here gamma is parameter which ranges from 0 t0 1,A higher value of gamma will perfectly fit the training dataset.

In situation like above SVM transform the input space to a higher dimensional space. The data points are plotted on X axis and Z axis, (Z is the squared sum of both x and y: z=x^2=y^2). Now we can easily segregate these red and blue points using linear separation.

Question 5

What do you mean by feature transformation?

Answer=

Feature transformation (FT) refers to family of algorithms that create new features using the existing features. These new features may not have the same interpretation as the original features, but they may have more discriminatory power in a different space than the original space. This can also be used for feature reduction. FT may happen in many ways, by simple/linear combinations of original features or using non-linear functions. Some common techniques for FT are:

- Scaling or normalizing features within a range, between 0 to 1.
- Principle Component Analysis and its variants.
- Random Projection.
- Neural Networks.
- SVM also transforms features internally.
- Transforming categorical features to numerical .

.