Question1:

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer 1:

1. Maximum Margin Classifier separates two classes perfectly but Soft Margin Classifier allows certain points in both classes to be deliberately misclassified.

2. Soft Margin Classifier can work for linear as well as non linear data but Maximum Margin Classifier works well only for Linear data.

3. Soft Margin Classifier works equally good with train and test data but Maximum Margin Classifier works poorly with test data(unseen data)

4. Soft Margin Classifier can handle noise class boundaries but Maximum Margin Classifier fails to do so.

Question 2:

What does the slack variable Epsilon (ε) represent?

Answer 2 :

A slack variable(ε) tells you where an observation is located relative to the margin and hyperplane.It is used to control **misclassifications**.

Each data point has a slack value associated to it, according to where the point is located. Generally lower values of slack are better than higher values

The value of slack lies between 0 and +infinity where

1. slack = 0 implies a correct classification
2. slack > 1 implies an incorrect classification
3. slack within 0 and 1 classifies correctly but violates the margin

Question 3:

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

Answer3 :

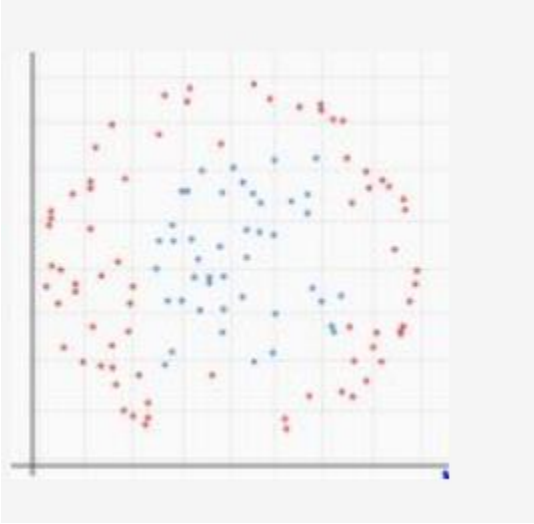
The cost function in SVM is measured by C - the sum of all the values of slack variables (epsilons) of each data point. It is denoted as.
$\sum \epsilon i \leq C$

The parameter cost of misclassification (C) represents the cost of violations to the margin and the hyperplane. Thus, the parameter C is called the tuning parameter.

If C is large, the slack variables (epsilons(ϵ)) can be large, i.e. you allow a larger number of data points to be misclassified or violate the margin

If C is small, you force the individual slack variables to be small, i.e. you do not allow many data points to fall on the wrong side of the margin or the hyperplane

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:

1. From graph, it is clear that both classes cannot be separated by linear SVC model.
2. So First we should check the accuracy of different non linear kernels like rbf, poly etc. The Non linear model providing highest accuracy should be considered.
3. With Help of GridSearch, find the combination of gamma and C which provides highest accuracy
4. With Obtained SVC Kernel, Gamma & C we can build a model.
5. This model will classify the dataset into two categories

Question 5:

What do you mean by feature transformation?

Answer 5:

The process of transforming the original attributes into a new feature space is called Feature Transformation.

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.

Some common techniques for Feature Transformation are:

1. Scaling or normalizing features within a range, say between 0 to 1.
2. Principle Component Analysis or SVD.
3. SVM also transforms features internally.
4. Transforming categorical features to numerical.