Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Ans: Soft Margin Classifier allows errors to be made unlike Maximum Margin Classifier which allows zero errors. Because of this Maximum Margin Classifier works better on the data that is linearly separable without any errors, where as Soft Margin Classifier solves this problem by introducing slack variables that is by deliberately allowing some misclassifications.

Question 2

What does the slack variable Epsilon (ε) represent?

Ans: Epsilon (ε) represents the degree to which the constraint on the data point can be violated. Each data point has a slack variable associated to it according to where the point is located.The value of slack variable lies between 0 and infinity

ε =0 implies data point is at safe distance, that is at a distance more than margin 'M' from the hyperplane

ε>1 implies data point is on the wrong side of the hyperplane,which results in incorrect classification

0< ε<1 implies data point is on the correct side of the hyperplane but falls inside the margin 'M'

Question 3

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

Ans: The cost function in SVM is measured by adding the slack variables of all the data points. The value of C represents the cost of violations to the margin and the hyperplane.

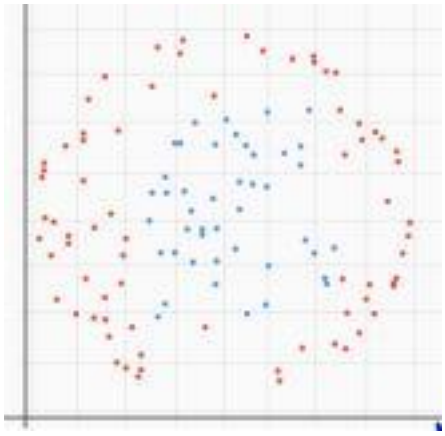
Cost function in SVM = $\sum_{i=1}^{n} \varepsilon$

Where

- 'i' represents the data point
- 'ε' represents the slack variable of each data point

- Theoretically large value C indicate that model allows more misclassifications, so the model is more generalisable.But chances of more errors.
- Small value of C indicate that model allows less number of misclassifications. so the model is less generalisable and more chances of overfitting.

Question 4



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

Ans: I would use the Radial Basis Function as the Kernel function to classify the data set as the data points are non-linearly distributed. Then fine tune the hyperparameters (Gamma - which represents the non linearity, C- which represents the cost function of SVM),so that the SVM model correctly fits the data set

Question 5

What do you mean by feature transformation?

Ans: The process of transforming the original attribute space to new high dimensional feature space by transforming the features so that the data points which were non-linearly distributed in the original attribute space become linearly separable is called Feature Transformation