# Assignment

**Question 1**: How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer 1:

- Maximal margin classifier works perfectly when the data is linearly separable using hyperplanes, whereas soft margin classifier works fine with non-linear data with the help of slack variables.
- MMC covers all data point whereas SMC leaves out outliers ie. MMC is greatly affected by outliers
- SMC is an extended version of MMC

**Question 2**: What does the slack variable Epsilon (ε) represent?

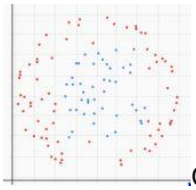
Answer 2: Slack variable is used to control misclassifications. It explains where an data point is located relative to the margin and hyperplane. Its value can vary from 0 to infinity.

- $\varepsilon = 0$ if observation is at safe distance from hyperplane and is falling on the correct side of hyperplane
- $0 < \varepsilon < 1$ if a data point is correctly classified but falls inside the margin.
- $\varepsilon > 1$ if data point is misclassified.

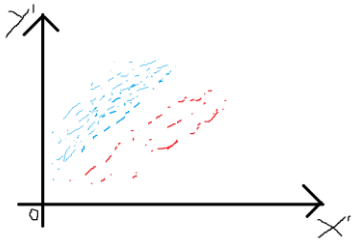
**Question 3**: How do you measure the cost function in SVM? What does the value of C signify?

Answer 3: In the SVM algorithm, the objective is to maximize the margin between the data points and the hyperplane. The parameter cost of misclassification (C) represents the cost of violations to the margin and the hyperplane. C is the sum of all the values of slack variables. C controls misclassifications. If C is large, i.e. we allow a larger number of data points to be misclassified. If C is small, we 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: We will use soft margin classifier and kernels to serve the purpose as they enable the linear SVM model to separate nonlinearly separable data points. Kernels transforms features mathematically and tries to find the hyperplane with those transformed values. Here also, if we square each value, we will get something like this, which can be separated by hyperplane.



**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’. This is done so that non-linear data become linear which can later

be separated into classes using hyperplanes. This leads to exponential increase in dimensions when we do transformation and hence makes model building computationally expensive. This problem is addressed by Kernels.