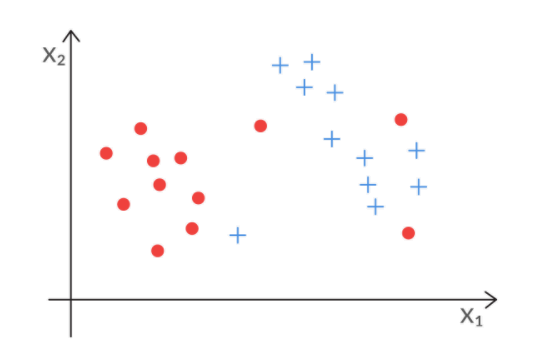
**Question 1**

How is Soft Margin Classifier different from Maximum Margin Classifier?

 To classify data using Maximum Margin Classifier, the data for two classes should be perfectly separated. If the data is partially intermingled as shown below , Maximum Margin Classifier is not possible in this case. We use Soft Margin Classifier that exists at equidistance from support vectors. Support vectors are the points that lie close to classifier on each class of data.



**Question 2**

What does the slack variable Epsilon (ε) represent?

Epsilon (ε) represents the location data point relative to margin and classifier. This is used to control misclassification.

Epsilon = 0 represents data points away from classifier with distance >= M(Margin) (strict classifier)

0< Epsilon < 1 represents data points away from classifier with distance < M(Margin) (strict classifier with margin violation)

Epsilon > 1 represents data point that is misclassified that lies on other side of classifier.(incorrect classifier)

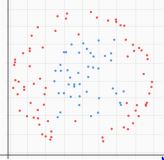
**Question 3**

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

Cost function / Cost of misclassification is sum of slack variables. C value represents the number of points that are classified wrongly.

High C value represents the higher number of points that are misclassified, So the obtained model will be more generalisable and does not overfit.. So model variance will be less but there will high bias. For lower value of C, It is more likely to overfit with high model variance and less bias.

**Question 4**

****

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

 The above figure shows that Maximum margin classifier or soft margin classifier cannot classify data. The only way is to use feature transformation of converting nonlinear data into linear data and then use classifier to classify data. SVM has kernel methods to do this trick and classify data. There are three Kernel functions:

Linear kernel – Same as SVC , does not perform any transformation.

Polynomial Kernel – Creates non linear decision boundaries.

Radial basis function – Capable of transforming highly nonlinear feature spaces to linear ones. Capable of creating enclosed decision boundaries also.

The above data set can be classified using RBF kernel function available in SVC.

**Question 5**

What do you mean by feature transformation?

The process of transforming the original attributes into a new feature space is called ‘**feature transformation**’.

In the above question the nonlinear feature space can be transformed to linear space to classify using SVM

Below example shows how attribute space can transformed into feature space and classify using SVM.

The new transformed feature space axis is (x-a)^2 instead of x. similarly for y. so that non linear classifier becomes linear classifier. So transforming X, Y to n-dimensional feature space based on the equation/shape of nonlinear classifier to get linear classifier is called Feature Transformation.

