

Reading Assignment 3

Q1. A hyperplane is a decision boundary or a separator line that is drawn in between data points of two classes, where it's position is also maximally far away from the nearest data points of any given class. Each axis in the feature space represents distance.

Q2. SVM seeks to minimize classification error or uncertainties through maximizing the geometric margin. Since the final standard formulation is in the form of a quadratic function, it is therefore trying to optimize the quadratic function subject to linear constraints.

Q3. A kernel is an efficient method of mapping original feature space to some higher dimensional space so that the training set is separable and yet relevant relationships between data points are preserved. The benefit of using kernel is to handle cases where data set cannot be classified by simply using a linear classifier. There are also different forms of kernels that would allow us to model more complex feature characteristics.

Q4. Soft margin is another form of maximum margin idea that allows the bulk of data sets to be separated into classes while ignoring a few mislabeled special cases. It works through the introduction of slack variables in the event where not all data sets can be linearly separated. A maximum margin will still be used to separate the data sets into classes as cleanly as possible and at the same time, since there will be a few remaining data sets that cannot be separated (those that fall within the margin), the degree of misclassification will be accounted for or measured by the slack variables. This transforms the optimization problem into a tradeoff between a large geometric margin versus errors.

*Submitted by:
Wee Teck Tan*