AP3 Report

The dataset was created using data gathered from Breast Cancer research by the University of Wisconsin. The dataset contained 682 data points that covered information on a variety of features, but specifically whether a tumor was Malignant or Benign.

These bits of data were split into a training and testing set (split 75% training/25% testing). The training data was used to create and train a SVM and the test set was used to assess it. It was found that its prediction precision was at 92.1%, its accuracy was at 96.5%, and its recall was at 98.3%. So, the performance of the SVM was very high.

The coefficients given by the SVM represent the coordinates of a vector which separates the plotted dataset. Negative coefficients represent a negative class and positive coefficients point to a positive class. In the initial dataset that was taken, only one coefficient was negative while the rest were positive.

The most important feature involved in the predictions would be the specification of the linear kernel; this is because the coefficients become very difficult to interpret correctly without it.