**AP3 Report**

The Breast Cancer dataset from the University of Wisconsin contained cancer patient data in which various attributes were recorded, including the condition of a tumor. These attributes were comprised of sample code number, clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, mitoses, and class. By analyzing this dataset, a prediction can be made from these attributes as to whether or not a tumor could be considered benign or malignant. After breaking the data points into training and test sets using a 75%/25% split, a Support Vector Machine was created and trained on the training data in order predict the results of the test set using the Scikit-Learn linear SVM. The performance of the SVM on the test set was fairly successful, including high values for prediction accuracy, precision, and recall:

The linear SVM coefficients can be interpreted as a representation of a vector that is perpendicular to the plane which separates the classes of malignant and benign tumors in the dataset. The direction of the vector is used to interpret the class being predicted by calculating the dot product of the vector and each data point to determine if the value is positive or negative. The importance of each feature on the prediction can be measured by the absolute value of each coefficient in relation to the others in the dataset, indicating its significance for separating the data. The feature that was the most important to the prediction was gamma. If gamma is too large, the algorithm’s learning rate can overshoot the correct value. Additionally, a very small gamma value can greatly increase processing time. An appropriate gamma value must be assigned in order to efficiently make an accurate prediction.