 The dataset was comprised of eleven columns, with the first being the index data. The last column labeled “Class” represented the known outcome. A ‘2’ indicates benign and a ‘4’ indicates malignant. Every other column represented an attribute of the area in question. The range of the data was between one and ten. A small sample of the data is shown below with column headings.

In my first attempt I tried using “svm.SVC(kernel=’linear’)” to setup my SVM. However, the program was still trying to fit the data after approximately 15 or 20 minutes. I instead switched to svm.LinearSVC, which produced results almost instantly. I started off getting around 65% accuracy rates and realized I was including the Sample column, greatly skewing the data. After removing that column I was above 95%. I further adjusted the code by varying ‘C’ between .0001 and 10000 by factors of ten. I found a value of one was about the best I could get, yielding an accuracy of 98.25%. The coefficients I arrived at are shown below. The best predictor based off the coefficients was Mitoses, followed closely by the Clump Thickness.

