Dimensionality reduction

Before doing the regression on our dataset we do some preprocessing where we standardize the dataset. From line 59 to 63, we do the reduction where we use Principal Component analysis to reduce to 2 dimensions as the assignment stated. We only use the dataset for 5,6 and 8.

Classification

To classify the test set, we used Logistic Regression, with a reduction down to 2 dimension we only get an accuracy of 67.741%, compared to if we only reduce to 100 dimensions, where we get 95.503%

Looking at the graph underneath, which is the results from the classification, it seems like 5 and 8 are clustering at the same area, which might explain the low accuracy, due to them being hard to distinguish between.