

## Homework 3 Report

### Problem 1

Unfortunately, due to the way the built-in library calculates the logistic regression model, it seems that there is now way to graph the model. However, a logistic regression model could be made for the feature maps and the results for accuracy, precision, and recall are as follows:

Accuracy: 0.9649122807017544

Precision: 0.9565217391304348

Recall: 0.9850746268656716

### Problem 2

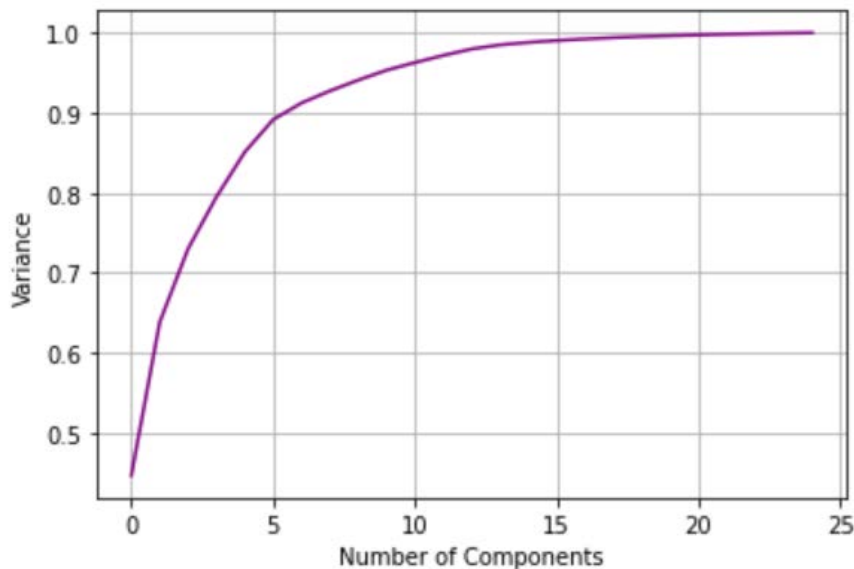
Problem 2 repeated the process of problem 1 using PCA feature extraction. This yielded the accuracy, precision, and recall results as follows:

Accuracy: 0.9824561403508771

Precision: 0.9722222222222222

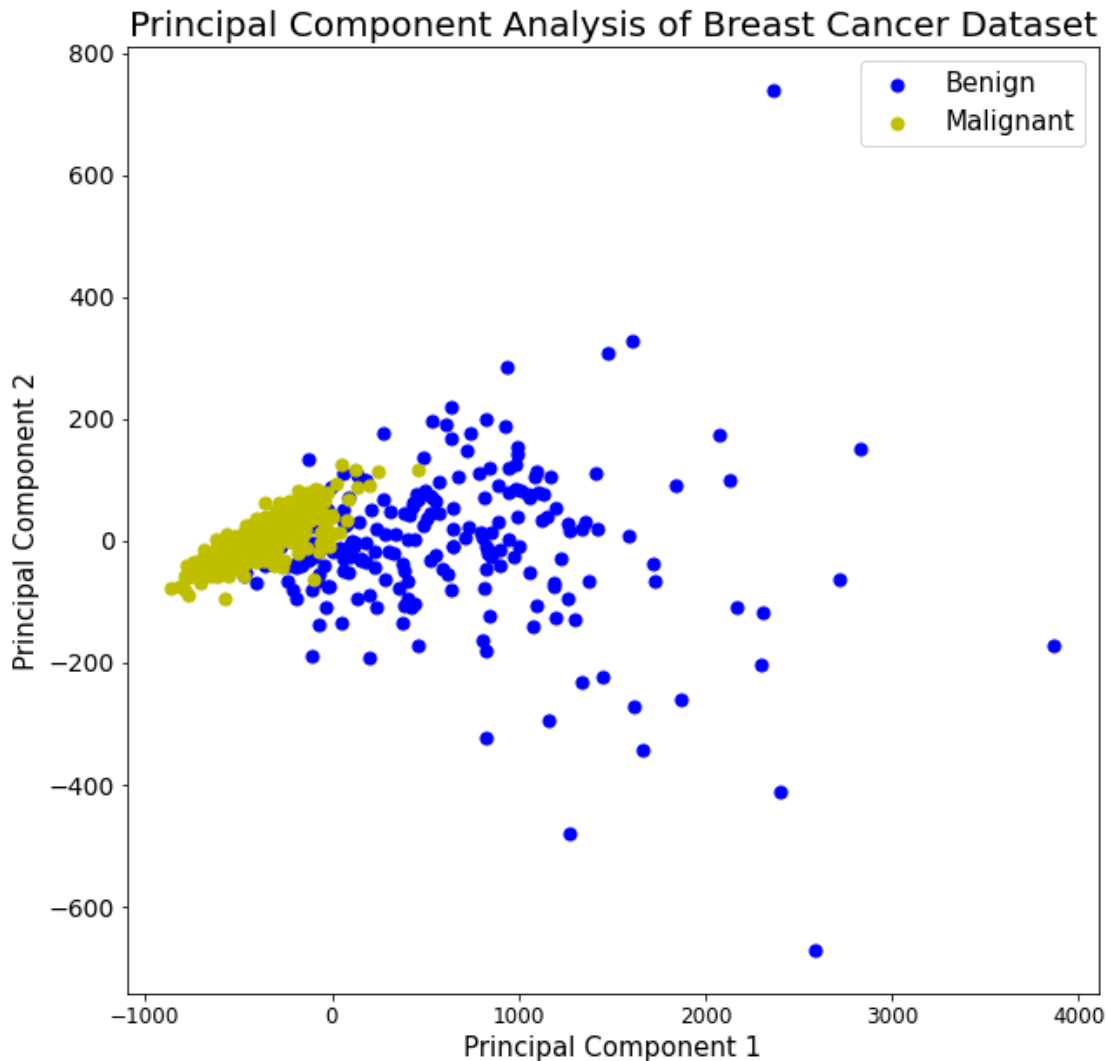
Recall: 1.0

The code ran through 25 values of N to find the optimum number of principal components to achieve the highest classification accuracy. The graph below shows the variance as the value of N increases.



After observing the graph, the optimum value of N seems to be around 20 as that is where the graph seems to plateau.

The graph below shows the scatterplot for the two principal components of the PCA, benign and malignant.



### **Problem 3**

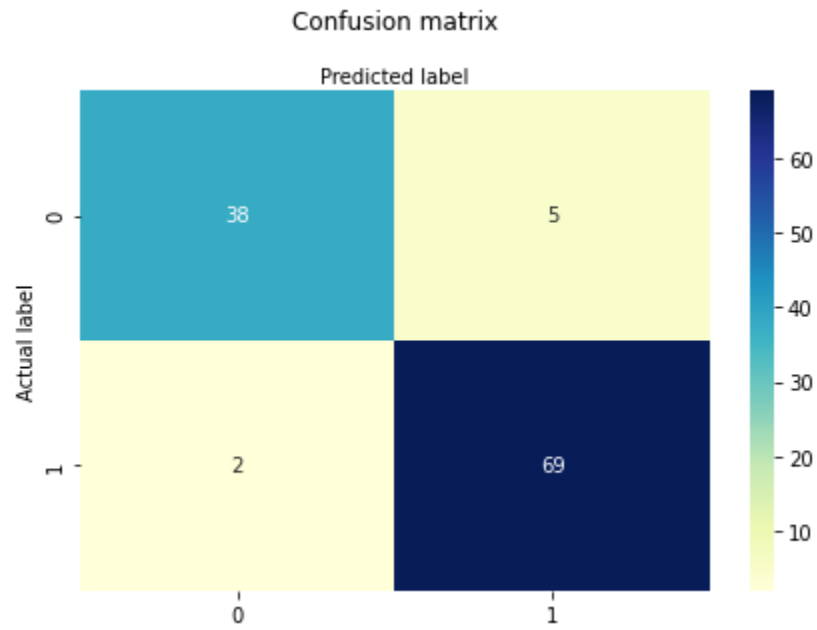
Problem 2 was repeated using LDA feature extraction. The results for the accuracy, precision, and recall are as follows:

Accuracy: 0.9385964912280702

Precision: 0.9324324324324325

Recall: 0.971830985915493

The code also yielded the confusion matrix below.



GitHub: <https://github.com/willwoodard16/Intro-to-ML-4105>