Johnny Nguyen

Student ID: 801119047

Homework #3

GitHub Link: https://github.com/Norumai01/Intro_Machine_Learning/tree/main/HW_3

Problem 1:

- Standardization scaling was uses for the following breast cancer dataset for optimized training model. Comparing the precision, recall and accuracy of the Naive Bayes' and logistic regression's training model, they have a 2-3 percent difference. However, they are generally the same range of accuracy so does not make any difference to use either model.

Problem 2:

- Using the PCA extraction, the dataset's features are reduced and standardization scaling was used for the logistic regression's training model. The number of components being used seems to result in variety of different accuracy, precision and recall. Generally, the more components were used for the training model, lesser the accuracy was for the classification report. As less components were used for the training model, the more accuracy that the training model had.

Problem 3:

- Using the PCA extraction, the dataset's features are reduced, and standardization scaling was used for the Naïve Bayes' training model. The classification report was generally the same with problem #2.