Johnny Nguyen

Student ID: 801119047

Homework #4

GitHub: https://github.com/Norumai01/Intro_Machine_Learning/tree/main/HW_4

Problem #1:

- Importing the breast cancer dataset from sklearns' library, PCA feature extraction and standardization scaling was used for the following training model. The graph is plotted with the measurement of the precision, recall and accuracy over the different numbers of components using PCA extraction. There are three graphs that uses three different kernels (Linear, Polynomial and Radial Basis Function) from the SVM classifier.
- For the highest classification accuracy of the three different kernels, the linear model highest accuracy was 0.93 when reduced to 22 principal components. For the polynomial model, the highest accuracy was 0.78 when reduced to 1 component. Lastly, the radial basis function model highest accuracy was 0.97 when reduced to 21 components.
- When looking at the different kernels that was used, the linear and radial basis function model gave the more optimal precision, recall and accuracy over the different number of components. Overall, the linear model seems more stabilize than other kernels.
- Comparing the linear model to the logistic regression model over the numbers of components, the precision, recall, and accuracy follows a steadily pace on the linear model with the SVM classifier. If there was a specific training model to select, it would preferably be the SVM classifier for general usage.

Problem 2:

- Importing the housing dataset, PCA extraction and standardization scaling was used similarly to Problem #1. The graph is also displayed with similar processes but only with regression accuracy in Problem #1.
- For different kernels with SVM regression, the linear model had the highest accuracy of 0.69 when reduced to 8 components. The polynomial kernel gave the highest accuracy of 0.54 when reduced to 5 components. Lastly, the radial basis function gave the highest accuracy of 0.58 when reduced to 10 components.
- Similarly with Problem #1, the SVM regression with the kernels, linear and radial basis function, had optimal rises of accuracy over the number of components. If a training model was to be selected, the linear model has a high accuracy on predicting the housing price that can be considered for generalized usage.