HW2.2 Ensemble Learning Report
Name: Atena Jafari Parsa
Course: AIN3001 Machine Learning

Introduction:

Ensemble learning is a powerful technique that combines the predictions of multiple machine learning models to enhance overall performance and robustness. In This assignment, a Voting Classifier is implemented in both hard, and soft voting configurations, using the Scikit-Learn library.
The ensemble is composed of three diverse classifiers: Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN).

Individual Classifiers:
RandomForestClassifier Accuracy: 0.896
SVC Accuracy: 0.856
KNeighborsClassifier Accuracy: 0.912
Hard VotingClassifier Accuracy: 0.904

According to the output above, the most accurate classifier is KNeighbors classifier with 0.912 accuracy. The Hard Voting Classifier accuracy is 0.904 which is better than the accuracy of the Random Forest classifier and the SVC, but is not the most accurate option.

The Bagging Classifier, employing an ensemble of 500 Decision Tree classifiers, each trained on 100 instances randomly sampled from the training set with replacement, demonstrated an accuracy score of 0.904 on the test set based on the output:
Bagging Accuracy Score: 0.904

Tree Classifier Accuracy Score: 0.856

Observing and comparing the performance of both the Tree Classifier (0.856 accuracy) and the Bagging Classifier (0.904 accuracy), clarifies how the ensemble approach has potentially improved the overall predictive accuracy.

Next, there is the out-of-bag evaluation, highlighting its role in providing a reliable estimate of the model's performance without the need for a separate validation set or cross-validation.

Out-of-Bag (oob) Score: 0.9253333333333333
Out-of-Bag (oob) Bagging Accuracy Score: 0.904

By looking at the output below, it is clear that both the Random Forest Classifier and the Bagging Random Forest Classifier achieved the same accuracy score of 0.912 on the test set. This result is consistent with the expectation that the RandomForestClassifier class in scikit-learn, is essentially a more convenient and optimized implementation of the bagging method for Decision Trees.

Random Forest Classifier Accuracy Score: 0.912
Bagging Random Forest Classifier Accuracy Score: 0.912

Soft Voting on SVM and KNeighbor Classifier Accuracy: 0.904
Soft Voting on RandomForest Classifier and KNeighbor Classifier Accuracy: 0.896
Soft Voting on SVM and Random Forest Classifier Accuracy: 0.896
Soft Voting on SVM and KNeighbor Classifier and KNeighbor Accuracy: 0.912

Soft Voting on the 3 classifiers SVM, Random Forest and KNeighbor had the highest accuracy (0.912).