Nam Nguyen – npn190000

Link: <https://colab.research.google.com/drive/1ik4DYMFV7EnfoOz74CKtfKWHPUX6uoFG#scrollTo=OEN55SDcN3-J>

**Exercise 1**

Construct the following models on the same dataset:

Bagging

Random Forest

Adaboost

Compare their performance and write a short paragraph on which one is the best. You are free to change the hyperparameters.

Answer:

# Create Bagging Classifier

clf = BaggingClassifier( n\_estimators=10, random\_state=0)

Table

Description automatically generated

Accuracy: 0.7881567973311092

# Create RandomForest Classifier

clf = RandomForestClassifier(max\_depth=3, random\_state=0, n\_estimators = 10)

Table

Description automatically generated

Accuracy: 0.6105087572977481

# Create Adaboost Classifier

clf = AdaBoostClassifier(n\_estimators=10, random\_state=0)

Table

Description automatically generated

Accuracy: 0.5696413678065054

After training and testing the models, the Bagging model outperforms both the Random Forest and Ada boost models. However, It also depends how to optimal the hyperparameters to create a best version for each algorithm. The bagging model is a simple ensemble technique that relies on multiple models to make predictions. The bagging model is less likely to overfit the data than a single model. The random Forest model and The Ada boost model are more sophisticated ensemble technique that relies on multiple models to make predictions

**Exercise 2**

The accuracy for this dataset is quite low. Can you try any other method that increases the accuracy. You can try either Random Forest or Adaboost. What do you notice?

Answer:

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated with medium confidence

Ada Boost is an algorithm that can be used with many different types of classifiers, such as decision trees. Ada boost work by weights instances in the training data set according to their importance. The weighting of instances is done according to weighting function. Both Ada boost and Random Forest achieved high accuracy on this dataset. Ada boost slightly outperformed Random Forest, but this is not significant.

**Exercise 3**

Try other combination of hyperparameters for Random Forest and AdaBoost models and check how good of an accuracy you can obtain.

Answer:

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Both Ada boost and Random Forest achieved high accuracy on this dataset. Ada boost slightly outperformed Random Forest, but the difference is not significant. However, It also depends how to optimal the hyperparameters to create a best version for each algorithm. Boosting is combining a set of "weak" classifiers to produce a strong classifier. Start with equal weights given to all training instances. Create a final hypothesis by taking a weighted average of these k hypotheses. Random Forest the main hyperparameters I tuned were the number of trees in the forest (n\_estimators) and the maximum depth of each tree (max\_depth). Ada boost main hyperparameters I tuned were the number of trees in the forest (n\_estimators) and the learning rate of each tree (Learning\_rate). We can create a better version by adjust the number to have a optimal version for this data set.