Dr. Tony Diana
DATA 602 Introduction to Machine Learning
Homework I Week 7

Ensemble Model

- Use the iris data set from sklearn.
- Load the iris dataset. What are the features?
- Create a DataFrame of given iris dataset.
- Split the dataset into training and test sets (30%).
- Create a Gaussian RandomForestClassifier as clf (2,000 estimators and a depth of 2).
- Determine the feature importance. Which one is the most important?
- Use scikitlearn to determine the accuracy level. What is your assessment?
- Use the Gradient Boosting algorithm to fit the model and predict test data.
- Compute the accuracy.
- Provide the feature importance.
- Did the Gradient Boosting model perform better? Are there any reservations about GB and why?