School of Computer Science Engineering and Technology

Course- BTech	Type- Core
Course Code- 301	Course Name-AIML
Year- 2022	Semester- Even
Date- 24-02-2022	Batch- 4th Sem (SPL)

6 - Lab Assignment No. 6.2

Objective: To Implement Adaboost and GradientBoost Classification models

Problem Statement: The goal is to implement the Adaboost and GradientBoost Classification models to predict whether the given patient is having Malignant or Benign tumor based on the attributes in the given dataset.

About Dataset: It is a dataset of Breast Cancer patients with Malignant and Benign tumor.

Steps

1. Download the dataset from the link https://drive.google.com/drive/folders/1TzuzYWFkeUaXwsoWt33wNNjHXOUR4BrA?usp=sharing
2. Remove unwanted features such as id
3. Check the presence of Null values present in the dataset
4. Handle Null values using fillna() (Student may use sklearn Lib i.e., SimpleImputer class)
5. Split the dataset into training and testing using 70-30 division
6. Perform feature scalling using Standardization approach
7. A) Train and Test the Adaboost model (Use SkLearn to train a AdaBoost)
   B) Create the confusion matrix and print the accuracy
   C) Playing with the models Change the following parameters of the AdaBoost and analyze their performance for training and testing using the evaluation measures
   I. base_estimator
   II. n_estimators
   III. learning_rate
   IV. algorithm
   V. random_state
8. A) Train and Test the Gradient Boosting Classifier model (Use SkLearn to train a GradientBoostingClassifier)
   B) Create the confusion matrix and print the accuracy
   C) Playing with the models Change the following parameters of the GradientBoostingClassifier and analyze their performance for training and testing using the evaluation measures
   I. n_estimators
   II. learning_rate
   III. max_depth
   IV. random_state
9. Compare the performance of the AdaBoost and GradientBoost classification model with other classification models such as Decision Tree and Random Forest etc.

Suggested Platform: Python: Azure Notebook/Google Colab Notebook, packages such as numpy, nltk, regular expression package re.

Additional (Not a part of the evaluation**)**

*** You can try different strategies to see whether testing error comes down or not. Strategies can be different 1. removal of some features, 2. Change normalization methods, 3. Shuffling of training samples. Check the model error for the testing data for each setup.