CS – 6375(ML Spring 2019) - PA 2

submitted by-
Anant Vaibhav (axv180022)

Implementation of bagging and boosting in python based on the decision tree creat ed in PA1.

The implementation is in the python file 'BaggingBoosting.py', which has all the code for the assignment.

Below is the output of the code.
It contains part PA2- > a,b, and c.

```
======================================================================
PART A - Four Combination of bagging size 10, 20 and tree depth 3,5
======================================================================
Number of Bags: 10 , Depth of Decision Tree: 3
Testing error=3.58%
Number of Bags: 10 , Depth of Decision Tree: 5
Testing error=0.21%
Number of Bags: 20 , Depth of Decision Tree: 3
Testing error=5.40%
Number of Bags: 20 , Depth of Decision Tree: 5
Testing error=0.13%
```

Below is the confusion matrix for various configurations.

Number of Bags: 10, Depth of Decision Tree: 3

	Classifier Positive	Classifier Negative
Actual Positive	2449.0	83.0
Actual Negative	135.0	3426.0

Number of Bags: 10, Depth of Decision Tree: 5

	Classifier Positive	Classifier Negative
Actual Positive	2519.0	13.0
Actual Negative	0.0	3561.0

Number of Bags: 20, Depth of Decision Tree: 3

	Classifier Positive	Classifier Negative
Actual Positive	2209.0	323.0
Actual Negative	6.0	3555.0

Number of Bags: 20, Depth of Decision Tree: 5

	Classifier Positive	Classifier Negative
Actual Positive	2530.0	2.0
Actual Negative	6.0	3555.0

```
================================================================================
PART B - Four Combination of Boosting Number of stumps of size 10, 20 and tre
e depth 1,2
================================================================================
Number of Stumps: 20 , Depth of Decision Tree: 1
Testing error=9.49%
Number of Stumps: 20 , Depth of Decision Tree: 2
Testing error=55.70%
Number of Stumps: 40 , Depth of Decision Tree: 1
Testing error=9.49%
Number of Stumps: 40 , Depth of Decision Tree: 2
Testing error=55.70%
```

Below is the confusion matrix for various configurations.

Number of Stumps: 20, Depth of Decision Tree: 1

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	578.0	2983.0

Number of Stumps: 20, Depth of Decision Tree: 2

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	3394.0	167.0

Number of Stumps: 40, Depth of Decision Tree: 1

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	578.0	2983.0

Number of Stumps: 40, Depth of Decision Tree: 2

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	3394.0	167.0

```
======================================================================
PART C - Sklearn bagging
======================================================================
Number of Stumps: 10 , Depth of Decision Tree: 3
Testing error=5.40%
Number of Stumps: 10 , Depth of Decision Tree: 5
Testing error=1.67%
Number of Stumps: 20 , Depth of Decision Tree: 3
Testing error=5.40%
Number of Stumps: 20 , Depth of Decision Tree: 5
Testing error=1.00%
```

Below is the confusion matrix for various configurations.

Number of Stumps: 10, Depth of Decision Tree: 3

	Classifier Positive	Classifier Negative
Actual Positive	2209.0	323.0
Actual Negative	6.0	3555.0

Number of Stumps: 10, Depth of Decision Tree: 5

	Classifier Positive	Classifier Negative
Actual Positive	2454.0	78.0
Actual Negative	24.0	3537.0

Number of Stumps: 20, Depth of Decision Tree: 3

	Classifier Positive	Classifier Negative
Actual Positive	2209.0	323.0
Actual Negative	6.0	3555.0

Number of Stumps: 20, Depth of Decision Tree: 5

	Classifier Positive	Classifier Negative
Actual Positive	2495.0	37.0
Actual Negative	24.0	3537.0

```
==========================================================================
PART C - Sklearn boosting
==========================================================================
Number of Stumps: 20 , Depth of Decision Tree: 1
Testing error=0.10%
Number of Stumps: 20 , Depth of Decision Tree: 2
Testing error=0.00%
Number of Stumps: 40 , Depth of Decision Tree: 1
Testing error=0.00%
Number of Stumps: 40 , Depth of Decision Tree: 2
Testing error=0.00%
```

Below is the confusion matrix for various configurations.

Number of Stumps: 20, Depth of Decision Tree: 1

	Classifier Positive	Classifier Negative
Actual Positive	2530.0	2.0
Actual Negative	4.0	3557.0

Number of Stumps: 20, Depth of Decision Tree: 2

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	0.0	3561.0

Number of Stumps: 40, Depth of Decision Tree: 1

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	0.0	3561.0

Number of Stumps: 40, Depth of Decision Tree: 2

	Classifier Positive	Classifier Negative
Actual Positive	2532.0	0.0
Actual Negative	0.0	3561.0

Comparison of our model to the sklearn:

- Execution time performance: The execution time of sklearn's model is way better than our model.
- Accuracy: The Accuracy of our model is less as compared to sklearn's model.