

CS 6375 Machine Learning Programming Assignment 4

Prepared By:

Ankit Sahu (AXS210226)

Yogesh Raisinghani (YXR210018)

Problem: Implement Bagging and AdaBoost based on the decision tree code that you developed in Programming Assignment 1. This code must be modified to work with ensemble methods.

PART A.

(**Bagging**, 20 points) Construct four models for each combination of maximum depth $d = 3, 5$ and bag size ($k = 10, 20$). Report the confusion matrix for these four settings.

The required confusion matrices are as follows:

Bagging: Max Depth 3 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	216.0	12.0
Actual Negative	0.0	204.0

Figure 1: Bagging - Depth 3, Size 10

Bagging: Max Depth 3 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	218.0	10.0
Actual Negative	0.0	204.0

Figure 2: Bagging - Depth 3, Size 20

Bagging: Max Depth 5 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	212.0	16.0
Actual Negative	6.0	198.0

Figure 3: Bagging - Depth 5, Size 10

Bagging: Max Depth 5 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	216.0	12.0
Actual Negative	0.0	204.0

Figure 4: Bagging - Depth 5, Size 20

PART B.

(**Boosting**, 20 points) Construct four models for each combination of maximum depth $d = 1, 2$ and bag size ($k = 20, 40$). Report the confusion matrix for these four settings.

The required confusion matrices are as follows:

Boosting: Max Depth 1 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	216.0	12.0
Actual Negative	0.0	204.0

Figure 5: Boosting – Depth 1, Size 20

Boosting: Max Depth 1 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	216.0	12.0
Actual Negative	0.0	204.0

Figure 6: Boosting – Depth 1, Size 40

Boosting: Max Depth 2 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	220.0	8.0
Actual Negative	4.0	200.0

Figure 7: Boosting – Depth 2, Size 20

Boosting: Max Depth 2 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	220.0	8.0
Actual Negative	4.0	200.0

Figure 8: Boosting – Depth 4, Size 40

PART C.

(**scikit-learn**, 40 points) Use scikit-learn's bagging and AdaBoost learners and repeat the experiments as described in parts (a) and (b) above. Report the confusion matrices for these sets of settings.

The required confusion matrices are as follows:

Sklearn Bagging: Max Depth 3 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	218.0	10.0
Actual Negative	6.0	198.0

Figure 9: skBagging – Max Depth 3, Size 10

Sklearn Bagging: Max Depth 3 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	216.0	12.0
Actual Negative	2.0	202.0

Figure 10: skBagging – Max Depth 3, Size 20

Sklearn Bagging: Max Depth 5 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	214.0	14.0
Actual Negative	8.0	196.0

Figure 11: skBagging – Max Depth 5, Size 10

Sklearn Bagging: Max Depth 5 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	210.0	18.0
Actual Negative	2.0	202.0

Figure 12: skBagging – Max Depth 5, Size 20

Sklearn Boosting: Max Depth 1 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	217.0	11.0
Actual Negative	16.0	188.0

Figure 13: skBoosting – Max Depth 1, Size 20

Sklearn Boosting: Max Depth 1 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	218.0	10.0
Actual Negative	16.0	188.0

Figure 14: skBoosting – Max Depth 1, Size 40

Sklearn Boosting: Max Depth 2 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	209.0	19.0
Actual Negative	10.0	194.0

Figure 15: skBoosting – Max Depth 2, Size 20

Sklearn Boosting: Max Depth 2 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	209.0	19.0
Actual Negative	8.0	196.0

Figure 16: skBoosting – Max Depth 2, Size 40