Rajarshi Chattopadhyay

IMPLEMENTATION OF BAGGING AND BOOSTING USING ID3 AND COMPARISON WITH SCIKIT-LEARN

Bagging: For the mushroom dataset, used self-implemented bagging and boosting algorithm that uses id3 and computed the confusion matrix on the test set for tree depths 3, 5 and bag size 5, 10.

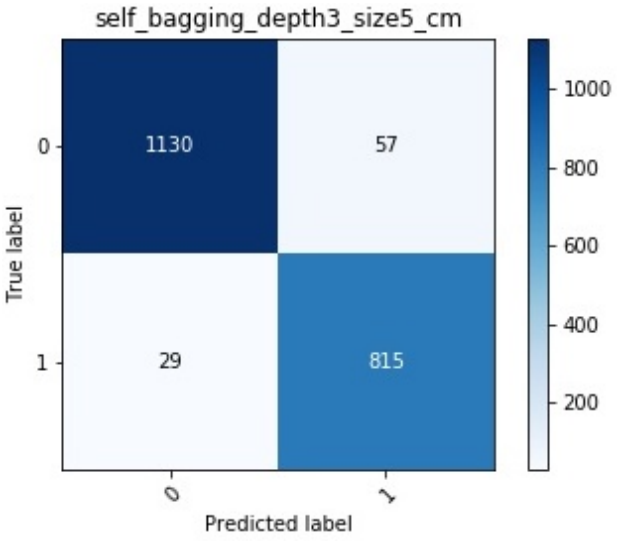
Scikit Learner: For mushroom, used scikit-learn's BaggingClassifier and AdaBoostClassifier that use DecisionTreeClassifier to learn a tree using criterion='entropy' and computed the confusion matrix on the test set for stump depths 1, 2 and ensemble size 5, 10.

Self-implementation on mushroom

Bagging

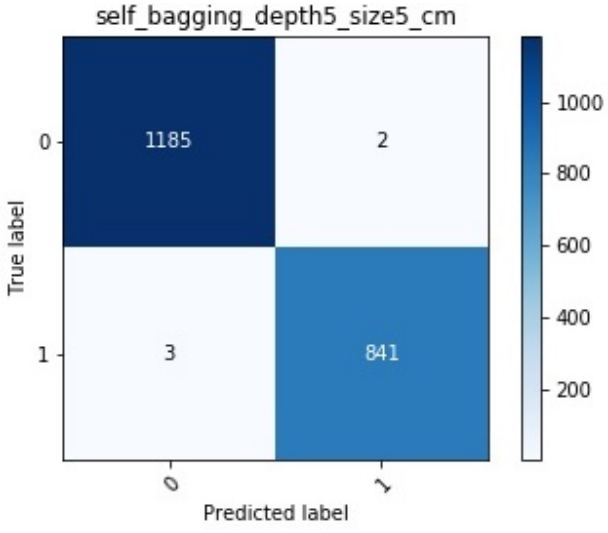
Bag Size 5 | Tree Depth 3

(tn, fp, fn, tp) = (1130, 57, 29, 815)



Bag Size 5 | Tree Depth 5

(tn, fp, fn, tp) = (1185, 2, 3, 841)

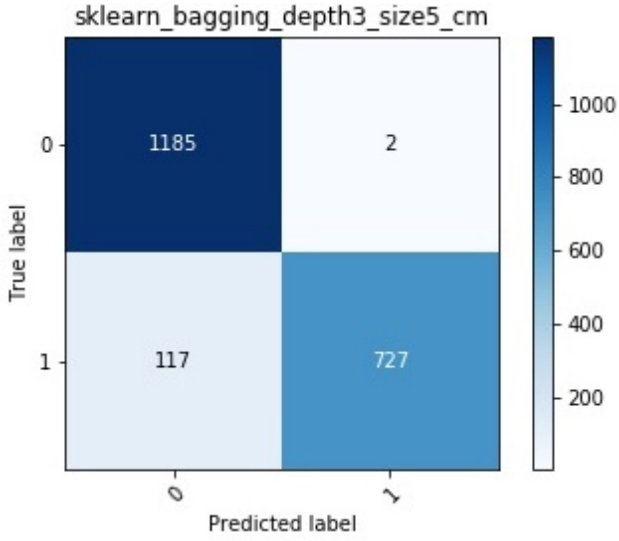


Scikit-learn's implementation on mushroom

Bagging

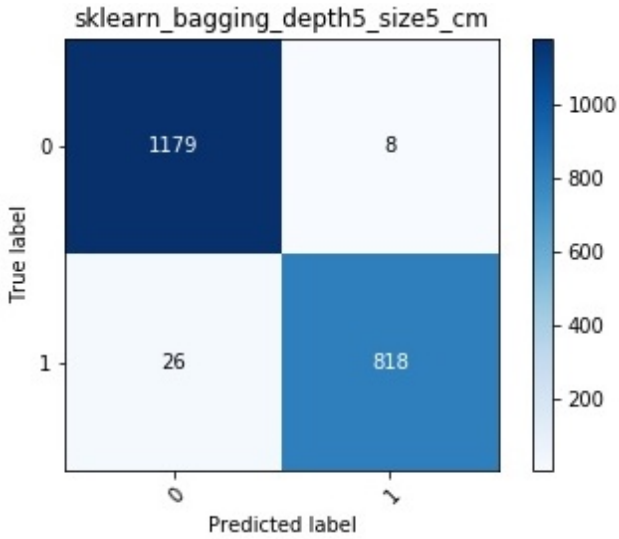
Bag Size 5 | Tree Depth 3

(tn, fp, fn, tp) = (1185, 2, 117, 727)



Bag Size 5 | Tree Depth 5

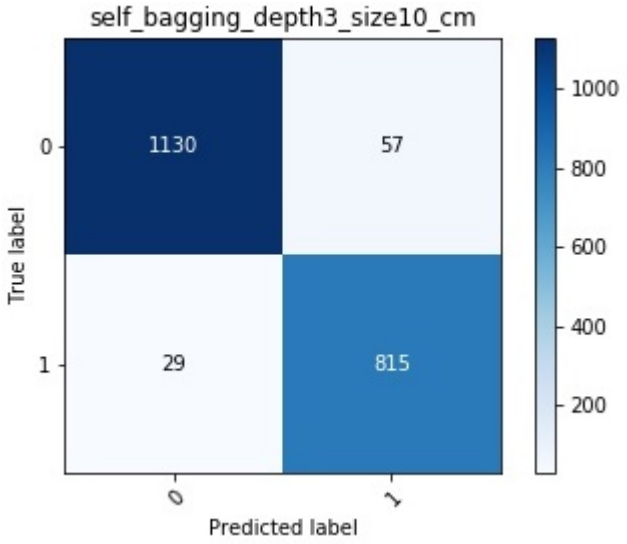
(tn, fp, fn, tp) = (1179, 8, 26, 818)



Self-implementation on mushroom

Bag Size 10 | Tree Depth 3

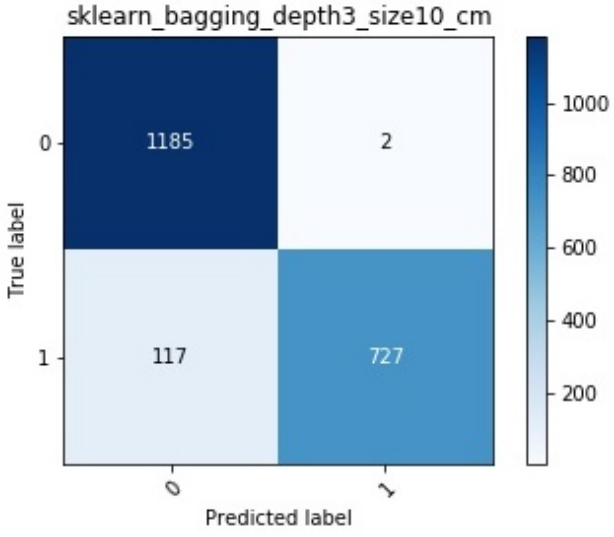
(tn, fp, fn, tp) = (1130, 57, 29, 815)



Scikit-learn's implementation on mushroom

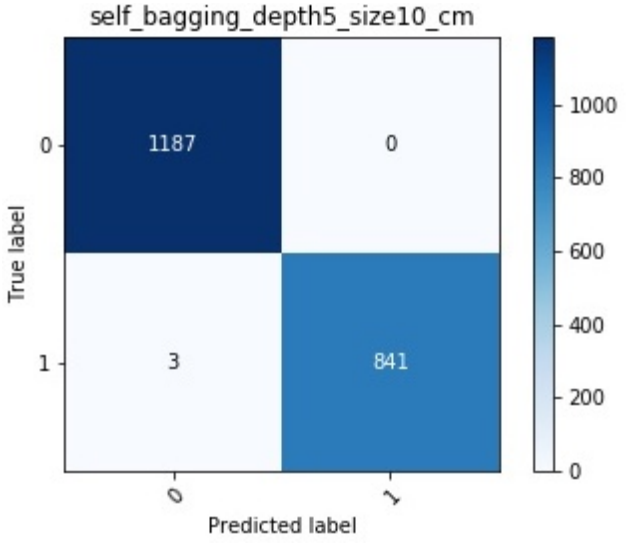
Bag Size 10 | Tree Depth 3

(tn, fp, fn, tp) = (1185, 2, 117, 727)



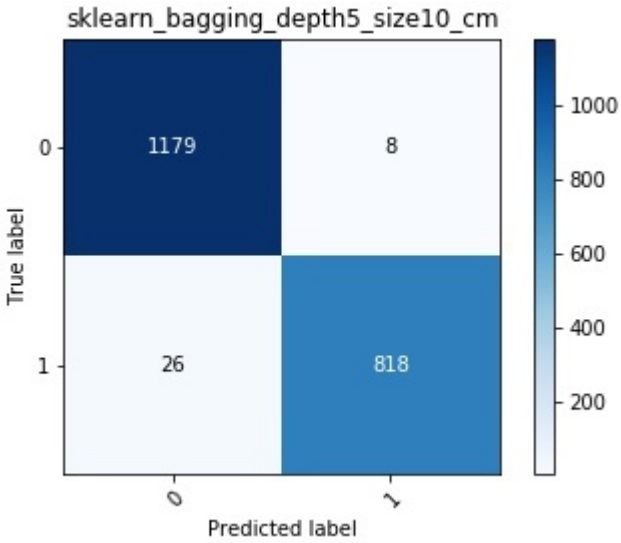
Bag Size 10 | Tree Depth 5

(tn, fp, fn, tp) = (1187, 0, 3, 841)



Bag Size 10 | Tree Depth 5

(tn, fp, fn, tp) = (1179, 8, 26, 818)

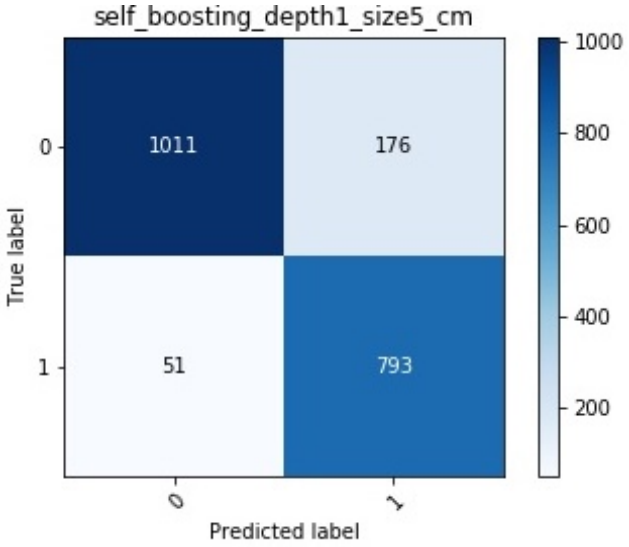


Self-implementation on mushroom

Boosting

Ensemble Size 5 | Stump Depth 1

(tn, fp, fn, tp) = (1011, 176, 51, 793)

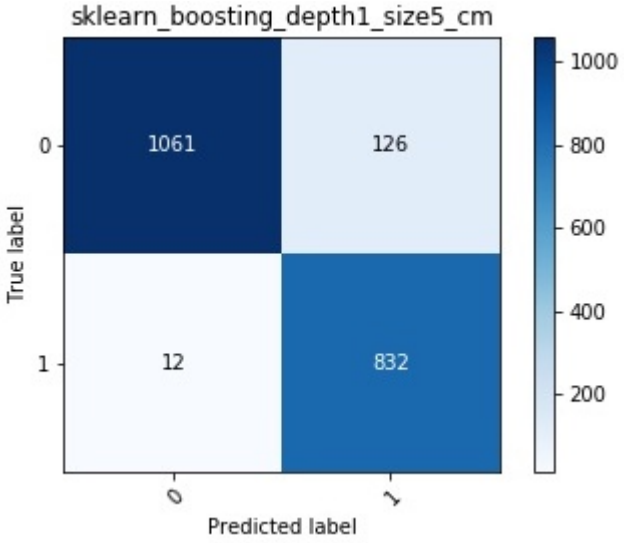


Scikit-learn's implementation on mushroom

Boosting

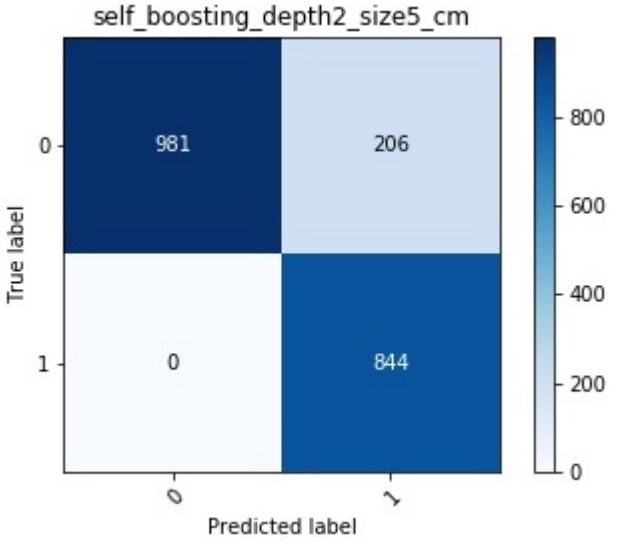
Ensemble Size 5 | Stump Depth 1

(tn, fp, fn, tp) = (1061, 126, 12, 832)



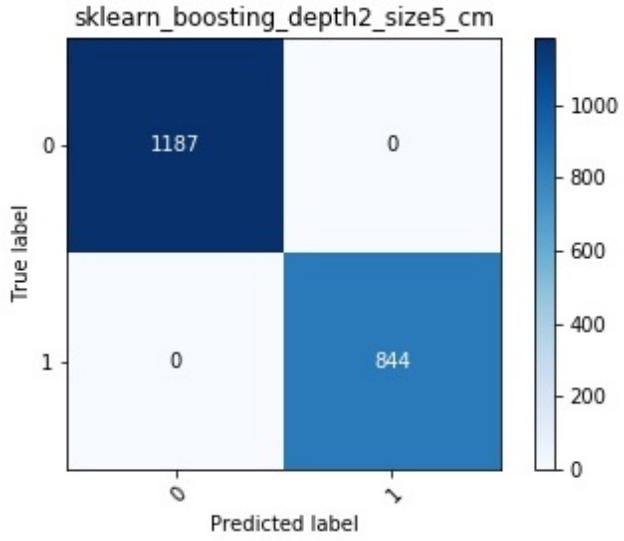
Ensemble Size 5 | Stump Depth 2

(tn, fp, fn, tp) = (981, 206, 0, 844)



Ensemble Size 5 | Stump Depth 2

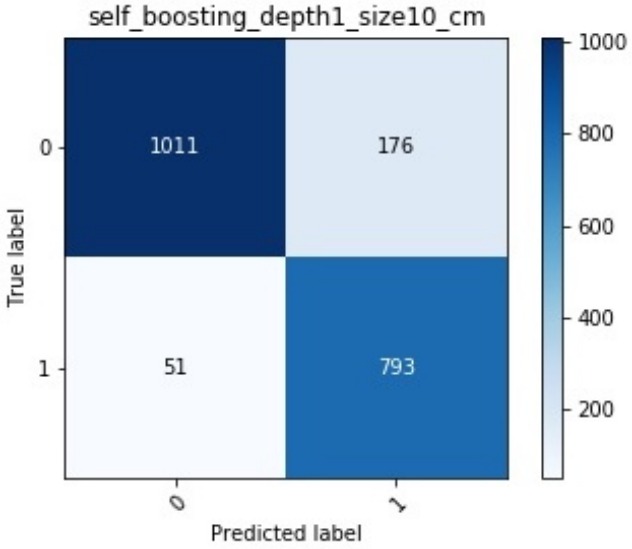
(tn, fp, fn, tp) = (1187, 0, 0, 844)



Self-implementation on mushroom

Ensemble Size 10 | Stump Depth 1

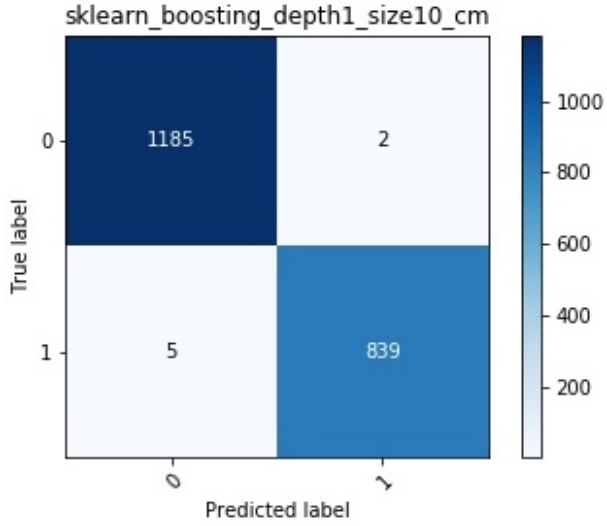
(tn, fp, fn, tp) = (1011, 176, 51, 793)



Scikit-learn's implementation on mushroom

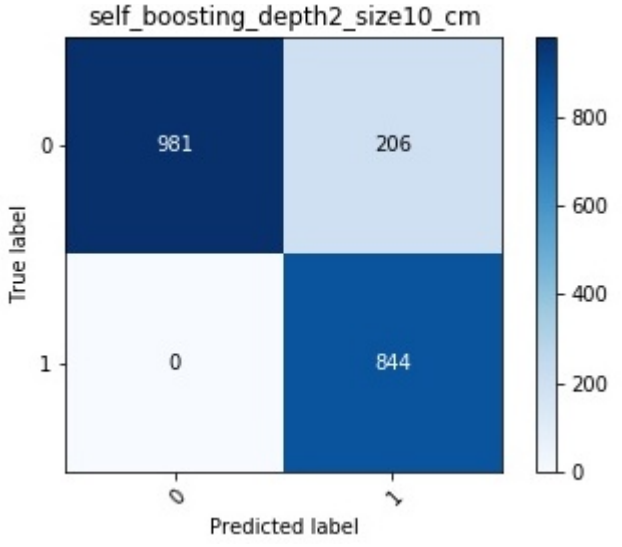
Ensemble Size 10 | Stump Depth 1

(tn, fp, fn, tp) = (1185, 2, 5, 839)



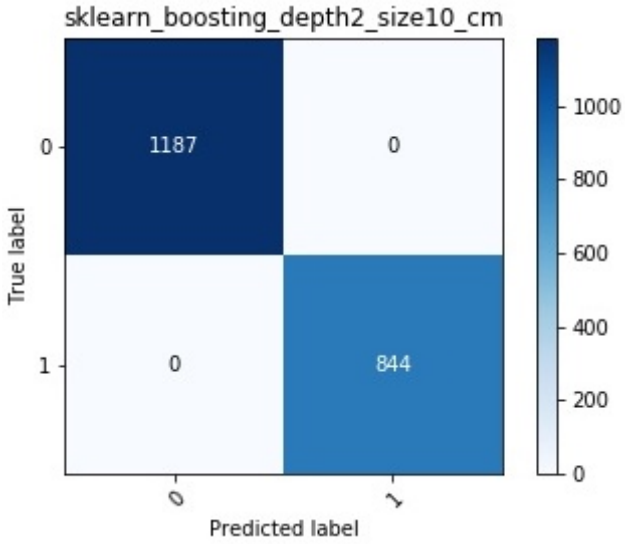
Ensemble Size 10 | Stump Depth 2

(tn, fp, fn, tp) = (981, 206, 0, 844)



Ensemble Size 10 | Stump Depth 2

(tn, fp, fn, tp) = (1187, 0, 0, 844)



Bagging on mushroom dataset

Bag Size 5 | Tree Depth 3

	tn	fp	fn	tp	Accuracy	Remarks
Self	1130	57	29	815	95.76%	classifies Positives better
Scikit	1185	2	117	727	94.14%	classifies Negatives better

Bag Size 5 | Tree Depth 5:

	tn	fp	fn	tp	Accuracy	Remarks
Self	1185	2	3	841	99.75%	classifies both Positives and negatives better
Scikit	1179	8	26	818	98.32%	-

Bag Size 10 | Tree Depth 3:

	tn	fp	fn	tp	Accuracy	Remarks
Self	1130	57	29	815	95.76%	classifies Positives better
Scikit	1185	2	117	727	94.14%	classifies Negatives better

Bag Size 10 | Tree Depth 5: ()

	tn	fp	fn	tp	Accuracy	Remarks
Self	1187	0	3	841	99.85%	classifies both Positives and negatives better
Scikit	1179	8	26	818	98.32%	-

Boosting on mushroom dataset

Ensemble Size 5 | Stump Depth 1:

	tn	fp	fn	tp	Accuracy	Remarks
Self	1011	176	51	793	88.82%	-
Scikit	1061	126	12	832	93.20%	classifies both Positives and negatives better

Ensemble Size 5 | Stump Depth 2:

	tn	fp	fn	tp	Accuracy	Remarks
Self	981	206	0	844	89.85%	classifies negatives accurately
Scikit	1187	0	0	844	100%	100% accurate overall classification

Ensemble Size 10 | Stump Depth 1:

	tn	fp	fn	tp	Accuracy	Remarks
Self	1011	176	51	793	88.82%	-
Scikit	1185	2	5	839	99.65%	classifies both Positives and negatives better

Ensemble Size 10 | Stump Depth 2: ()

	tn	fp	fn	tp	Accuracy	Remarks
Self	981	206	0	844	89.86%	classifies negatives accurately
Scikit	1187	0	0	844	100%	100% accurate overall classification