CS 573 – Homework 4

Priyank Jain (3 slip days used)
jain206@purdue.edu

April 17, 2017

DT stands for decision trees, BT stands for Bagging, RF stands for Random Forests, BDT stands for Boosted Decision Trees, SVM stands for Support Vector Machine.

Code

Figure 1 below shows the output of DT (Decision Tree), BT (Bagging), RF (Random Forests), BDT (Boosted Decision Trees) for test cases of Homework 2.

Analysis

1. (a)

Figure 2 shows the learning curves for DT, BT, RF, BDT and SVM models for varying training set sizes.

1. (b)

Tables 1-4 represent the zero-one loss across the the incremental CV folds for each model and each different training set size.

Let μ_{BT} refer to mean zero-one loss of the bagging classifier and μ_{SVM} refer to mean zero-one loss of the SVM classifier.

Null Hypothesis (H_0): $\mu_{BT} = \mu_{SVM}$ (or $\mu_{DT} = \mu_{SVM}$ or $\mu_{RF} = \mu_{SVM}$ or $\mu_{BDT} = \mu_{SVM}$)

Alternative Hypothesis (H_1): $\mu_{BT} > \mu_{SVM}$ (or $\mu_{DT} > \mu_{SVM}$ or $\mu_{RF} > \mu_{SVM}$ or $\mu_{BDT} > \mu_{SVM}$)

From the graph, we see that the bagging (or DT or RF or BDT) classifier has a higher 0/1 loss for all training set sizes compared to SVM. This difference is significant because the standard error bars of BT (or DT or RF or BDT) do not overlap and are sufficiently far away from that of the SVM classifier.

Alternatively, we can perform a one-tailed paired t-test for each training set size. We will choose our significance $\alpha = 0.05$. In order to correct for testing multiple hypotheses, we apply Bonferroni's correction. We reject the null hypothesis if the p-value is less than $\frac{\alpha}{4} = 0.0125$

From table 5, we see that we reject the null hypothesis for all training percentages.

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.29	0.41	0.31	0.38	0.275	0.425	0.365	0.34	0.4	0.345
BT	0.34	0.46	0.285	0.28	0.285	0.39	0.36	0.35	0.475	0.295
RF	0.28	0.445	0.22	0.335	0.335	0.38	0.29	0.25	0.47	0.325
BDT	0.29	0.41	0.25	0.395	0.385	0.36	0.365	0.34	0.4	0.4
SVM	0.155	0.33	0.225	0.23	0.19	0.22	0.18	0.26	0.37	0.195

Table 1: Zero-one loss for the 10 cross-validations runs for five models with TSS = 0.025

```
priyank@priyank: ~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train0.txt yelp_test0.txt 1
ZERO-ONE-LOSS-DT 0.32
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train0.txt yelp_test0.txt 2
ZERO-ONE-LOSS-BT 0.16333333333333333
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train0.txt yelp_test0.txt 3
ZERO-ONE-LOSS-RF 0.19666666666666666
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train0.txt yelp_test0.txt 4
ZERO-ONE-LOSS-BDT 0.14333333333333334
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train1.txt yelp_test1.txt 1
ZERO-ONE-LOSS-DT 0.11
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train1.txt yelp_test1.txt 2
ZERO-ONE-LOSS-BT 0.08
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train1.txt yelp_test1.txt 3
ZERO-ONE-LOSS-RF 0.07
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_train1.txt yelp_test1.txt 4
ZERO-ONE-LOSS-BDT 0.07
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_data.csv yelp_data.csv 1
ZERO-ONE-LOSS-DT 0.1045
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_data.csv yelp_data.csv 2
ZERO-ONE-LOSS-BT 0.0765
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_data.csv yelp_data.csv 3
ZERO-ONE-LOSS-RF 0.081
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$ python hw4.py yelp_data.csv yelp_data.csv 4
ZERO-ONE-LOSS-BDT 0.0
priyank@priyank:~/Desktop/Dropbox/Spring 2017/DM/My HW Solutions/HW4$
```

Figure 1: Output for testcases of Homework 2

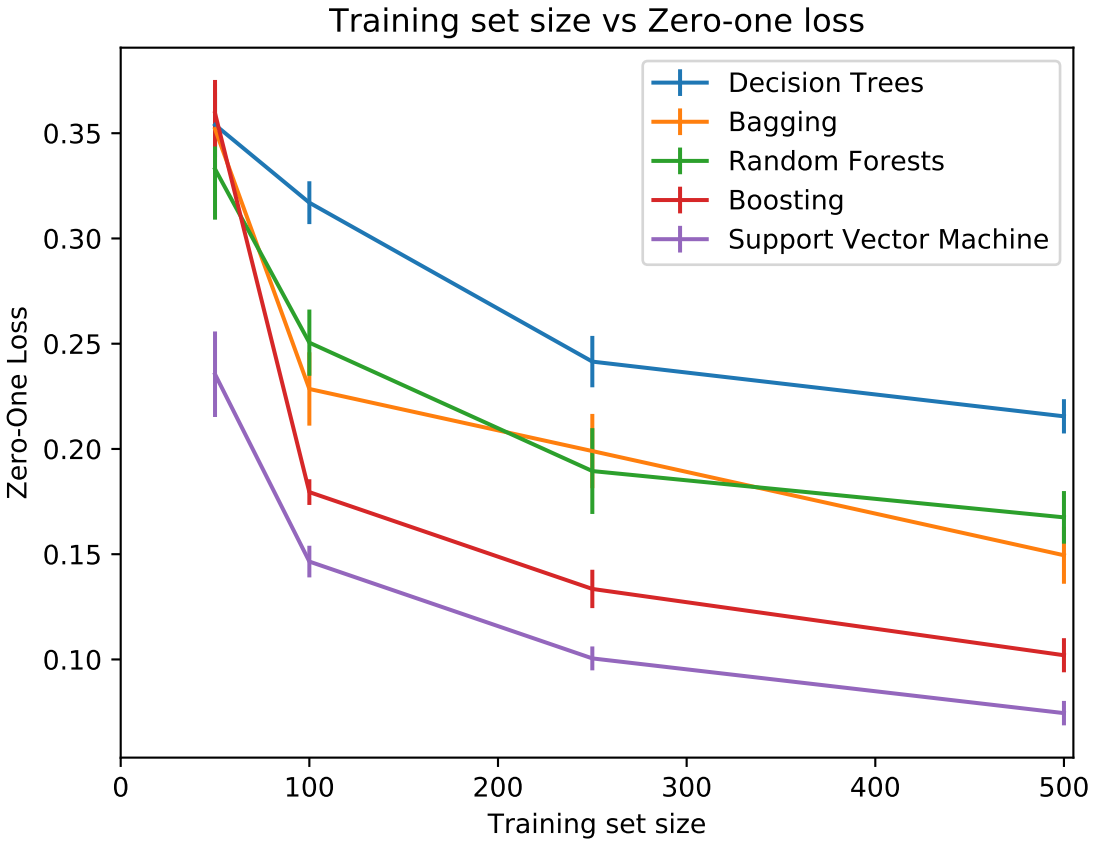

Figure 2: Training set size vs Zero-one Loss

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.285	0.285	0.36	0.325	0.385	0.305	0.33	0.29	0.29	0.315
BT	0.26	0.155	0.305	0.2	0.305	0.24	0.23	0.17	0.15	0.27
RF	0.135	0.255	0.295	0.235	0.295	0.235	0.26	0.215	0.255	0.325
BDT	0.145	0.215	0.17	0.17	0.16	0.175	0.195	0.185	0.18	0.2
SVM	0.125	0.165	0.16	0.145	0.17	0.185	0.135	0.155	0.12	0.105

Table 2: Zero-one loss for the 10 cross-validations runs for five models with TSS = 0.05

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.275	0.205	0.22	0.215	0.265	0.28	0.22	0.195	0.22	0.32
BT	0.165	0.15	0.195	0.235	0.19	0.16	0.295	0.135	0.165	0.3
RF	0.16	0.165	0.11	0.13	0.175	0.235	0.285	0.125	0.2	0.31
BDT	0.155	0.18	0.105	0.1	0.165	0.11	0.12	0.115	0.115	0.17
SVM	0.085	0.125	0.09	0.105	0.13	0.065	0.095	0.11	0.1	0.1

Table 3: Zero-one loss for the 10 cross-validations runs for five models with TSS = 0.125

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.22	0.195	0.245	0.235	0.155	0.22	0.25	0.21	0.215	0.21
BT	0.13	0.105	0.22	0.165	0.115	0.14	0.205	0.12	0.095	0.2
RF	0.12	0.16	0.18	0.16	0.125	0.135	0.25	0.16	0.16	0.225
BDT	0.105	0.105	0.145	0.095	0.07	0.11	0.145	0.095	0.085	0.065
SVM	0.075	0.055	0.11	0.06	0.065	0.095	0.075	0.08	0.045	0.085

Table 4: Zero-one loss for the 10 cross-validations runs for five models with TSS = 0.25

Null Hypothesis/TSS	0.025	0.05	0.125	0.25
$\mu_{DT} = \mu_{SVM}$	3.88e-5	5.52e-8	4.85e-6	2.25e-8
$\mu_{BT} = \mu_{SVM}$	1.6e-5	1.09e-3	3.86e-4	4.23e-5
$\mu_{RF} = \mu_{SVM}$	2.5e-4	1.54e-4	2.23e-3	3.36e-5
$\mu_{BDT} = \mu_{SVM}$	1.16e-4	6.14e-3	1.7e-3	3.54e-3

Table 5: p-values for paired t-tests for varying training set sizes

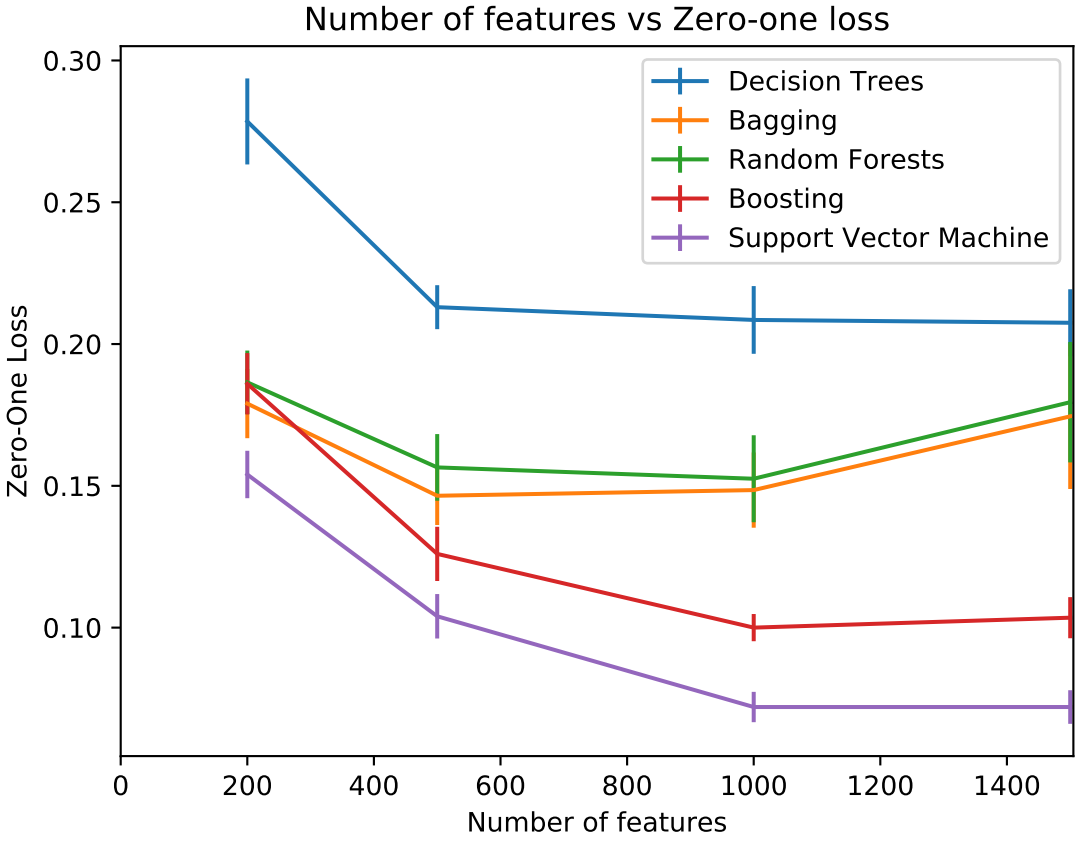

Figure 3: Number of features vs Zero-one loss

2. (a)
Figure 3 shows the learning curves for DT, BT, RF, BDT and SVM models for varying number of features.

2. (b)
Tables 6-9 represent the zero-one loss across the the incremental CV folds for each model and each different number of features.
Let μ_{BT} refer to mean zero-one loss of the bagging classifier and μ_{SVM} refer to mean zero-one loss of the SVM classifier.
Null Hypothesis (H_0): $\mu_{BT} = \mu_{SVM}$ (or $\mu_{DT} = \mu_{SVM}$ or $\mu_{RF} = \mu_{SVM}$ or $\mu_{BDT} = \mu_{SVM}$)

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.295	0.245	0.215	0.26	0.24	0.31	0.375	0.265	0.34	0.24
BT	0.125	0.175	0.17	0.145	0.18	0.19	0.245	0.19	0.24	0.13
RF	0.175	0.22	0.15	0.16	0.17	0.19	0.245	0.19	0.235	0.13
BDT	0.175	0.175	0.165	0.125	0.19	0.265	0.21	0.185	0.17	0.2
SVM	0.17	0.145	0.16	0.115	0.135	0.18	0.195	0.185	0.125	0.13

Table 6: Zero-one loss for the 10 cross-validations runs for five models with 200 features

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.185	0.195	0.215	0.265	0.205	0.235	0.225	0.22	0.175	0.21
BT	0.11	0.12	0.12	0.12	0.175	0.14	0.225	0.15	0.16	0.145
RF	0.16	0.15	0.135	0.22	0.15	0.115	0.22	0.105	0.135	0.175
BDT	0.1	0.13	0.125	0.14	0.15	0.195	0.125	0.08	0.115	0.1
SVM	0.055	0.08	0.095	0.11	0.12	0.13	0.145	0.11	0.11	0.085

Table 7: Zero-one loss for the 10 cross-validations runs for five models with 500 features

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.19	0.175	0.25	0.24	0.18	0.275	0.21	0.205	0.22	0.14
BT	0.105	0.11	0.115	0.17	0.21	0.19	0.195	0.145	0.08	0.165
RF	0.085	0.15	0.13	0.19	0.165	0.24	0.175	0.13	0.07	0.19
BDT	0.08	0.12	0.1	0.11	0.095	0.125	0.105	0.1	0.075	0.09
SVM	0.075	0.08	0.075	0.055	0.07	0.075	0.08	0.11	0.05	0.05

Table 8: Zero-one loss for the 10 cross-validations runs for five models with 1000 features

Alternative Hypothesis (H_1): $\mu_{BT} > \mu_{SVM}$ (or $\mu_{DT} > \mu_{SVM}$ or $\mu_{RF} > \mu_{SVM}$ or $\mu_{BDT} > \mu_{SVM}$)

From the graph, we see that the bagging (or DT or RF or BDT) classifier has a higher 0/1 loss for all different number of features compared to SVM. This difference is significant because the standard error bars of BT (or DT or RF or BDT) do not overlap and are sufficiently far away from that of the SVM classifier (except for BT vs SVM with 200 features and BDT vs SVM with 500 features).
Alternatively, we can perform a one-tailed paired t-test for each different number of features. We will choose our significance $\alpha = 0.05$. In order to correct for testing multiple hypotheses, we apply Bonferroni's correction. We reject the null hypothesis if the p-value is less than $\frac{\alpha}{4} = 0.0125$

From table 10, we see that we reject the null hypothesis for all different number of features, except for BT vs SVM with 200 features and BDT vs SVM with 500 features.

3. (a)
Figure 4 shows the learning curves for DT, BT, RF and BDT models for varying depth limits.

3. (b)
Tables 11-14 represent the zero-one loss across the the incremental CV folds for each model and different depth limits.
Let μ_{BT} refer to mean zero-one loss of the BT classifier and μ_{DT} refer to mean zero-one loss of the DT classifier.

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.205	0.185	0.155	0.245	0.245	0.215	0.275	0.165	0.17	0.215
BT	0.105	0.125	0.15	0.075	0.24	0.26	0.32	0.23	0.065	0.175
RF	0.095	0.21	0.17	0.11	0.285	0.25	0.255	0.2	0.1	0.12
BDT	0.105	0.08	0.12	0.12	0.12	0.135	0.12	0.1	0.065	0.07
SVM	0.07	0.055	0.075	0.065	0.075	0.11	0.085	0.09	0.045	0.05

Table 9: Zero-one loss for the 10 cross-validations runs for five models with 1500 features

Null Hypothesis/Number of features	200	500	1000	1500
$\mu_{DT} = \mu_{SVM}$	7.41e-6	1.96e-7	1.39e-6	1.13e-6
$\mu_{BT} = \mu_{SVM}$	4.48e-2	1.0e-4	2.83e-4	7.89e-4
$\mu_{RF} = \mu_{SVM}$	1.24e-2	2.25e-3	6.12e-4	1.59e-4
$\mu_{BDT} = \mu_{SVM}$	4.25e-3	2.32e-2	7.62e-4	2.71e-5

Table 10: p-values for paired t-tests for varying number of features

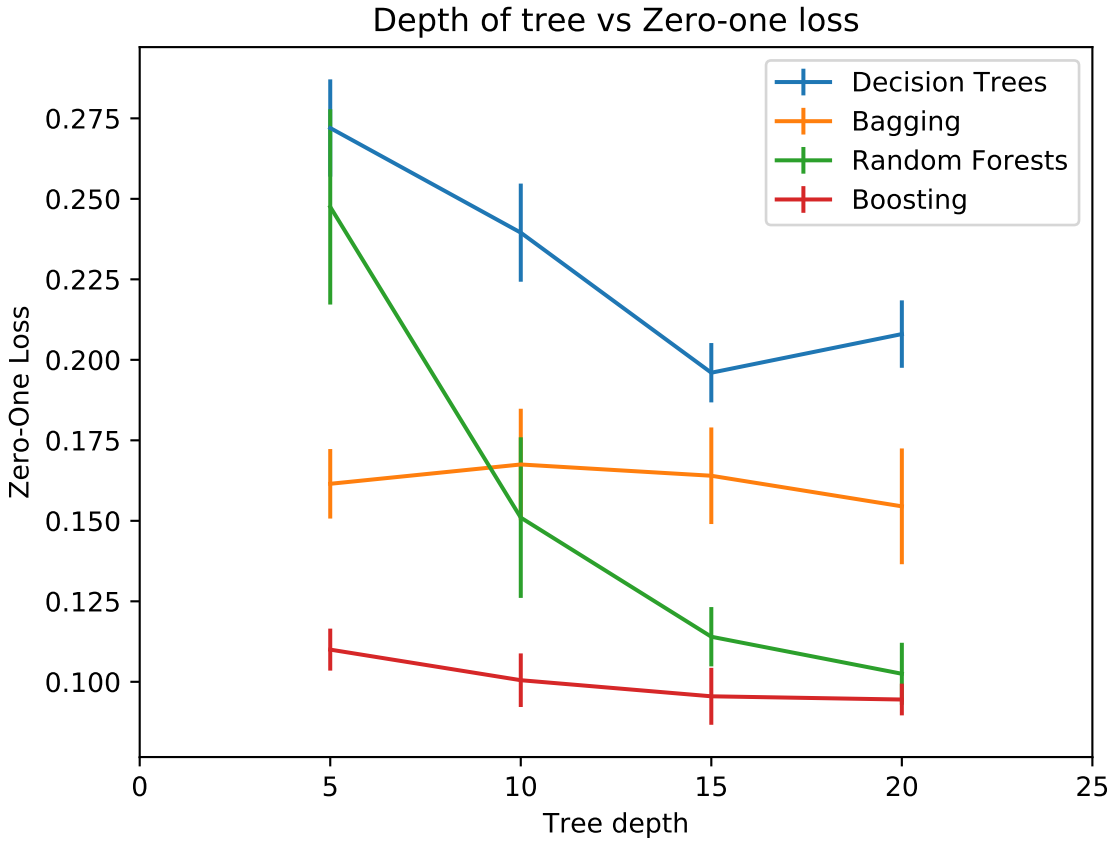

Figure 4: Depth of tree vs Zero-one loss

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.24	0.19	0.325	0.325	0.315	0.285	0.33	0.225	0.25	0.235
BT	0.13	0.16	0.175	0.165	0.105	0.185	0.22	0.135	0.205	0.135
RF	0.155	0.15	0.23	0.35	0.135	0.285	0.405	0.13	0.3	0.335
BDT	0.125	0.085	0.09	0.125	0.095	0.155	0.105	0.12	0.11	0.09

Table 11: Zero-one loss for the 10 cross-validations runs for four models with depth limit 5

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.205	0.285	0.3	0.2	0.295	0.225	0.17	0.24	0.18	0.295
BT	0.145	0.315	0.165	0.19	0.145	0.165	0.11	0.17	0.11	0.16
RF	0.155	0.34	0.125	0.085	0.14	0.22	0.11	0.095	0.05	0.19
BDT	0.1	0.135	0.095	0.08	0.125	0.14	0.08	0.115	0.055	0.08

Table 12: Zero-one loss for the 10 cross-validations runs for four models with depth limit 10

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.17	0.21	0.145	0.215	0.2	0.195	0.195	0.255	0.21	0.165
BT	0.19	0.125	0.13	0.17	0.18	0.16	0.26	0.215	0.1	0.11
RF	0.16	0.12	0.095	0.145	0.09	0.105	0.115	0.155	0.08	0.075
BDT	0.085	0.125	0.08	0.09	0.06	0.135	0.09	0.135	0.05	0.105

Table 13: Zero-one loss for the 10 cross-validations runs for four models with depth limit 15

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.2	0.17	0.165	0.21	0.245	0.215	0.28	0.22	0.19	0.185
BT	0.105	0.23	0.135	0.1	0.14	0.125	0.285	0.125	0.18	0.12
RF	0.05	0.145	0.095	0.075	0.115	0.095	0.155	0.115	0.075	0.105
BDT	0.065	0.115	0.095	0.085	0.085	0.11	0.11	0.11	0.09	0.08

Table 14: Zero-one loss for the 10 cross-validations runs for four models with depth limit 20

Null Hypothesis (H_0): $\mu_{DT} = \mu_{BT}$ (or $\mu_{DT} = \mu_{RF}$ or $\mu_{DT} = \mu_{BDT}$ or $\mu_{BT} = \mu_{RF}$ or $\mu_{BT} = \mu_{BDT}$ or $\mu_{RF} = \mu_{BDT}$)
Alternative Hypothesis (H_1): $\mu_{DT} > \mu_{BT}$ (or $\mu_{DT} > \mu_{RF}$ or $\mu_{DT} > \mu_{BDT}$ or $\mu_{BT} > \mu_{RF}$ or $\mu_{BT} > \mu_{BDT}$ or $\mu_{RF} > \mu_{BDT}$)
From the graph, we see that the DT classifier has a higher 0/1 loss than all other models. This difference is significant (except for DT vs RF with depth 5 and DT vs BT with depth 15) because the standard error bars of DT do not overlap and are sufficiently far away from those of other classifiers. Also, the BT classifier has a higher 0-1 loss than that of BDT, the difference is significant since the errors bars do not overlap and are sufficiently far way. The BT classifier performs better than RF for depth 5, but they both perform equally good for depth 10 since their error bars overlap. Thereafter RF outperforms BT for depths 15 and 20. RF performs that BDT for depth 5, but thereafter their performance are comparable since the corresponding error bars overlap/ are very close to each other.
Alternatively, we can perform a one-tailed paired t-test for each different depth limit. We will choose our significance $\alpha = 0.05$. In order to correct for testing multiple hypotheses, we apply Bonferroni's correction. We reject the null hypothesis if the p-value is less than $\frac{\alpha}{4} = 0.0125$

From table 15, we see that we reject the null hypothesis for all different number of features, except for DT vs BT with depth 15, DT vs RF with depth 5, BT vs RF with depth 10, RF vs BDT with depths 5, 10 and 20.

4. (a)
Figure 5 shows the learning curves for DT, BT, RF and BDT models for varying number of

Null Hypothesis/Depth limit	5	10	15	20
$\mu_{DT} = \mu_{BT}$	5.03e-5	1.48e-3	3.6e-2	8.04e-3
$\mu_{DT} = \mu_{RF}$	0.20	1.96e-3	1.36e-5	4.49e-6
$\mu_{DT} = \mu_{BDT}$	2.12e-6	2.28e-6	4.11e-6	1.79e-6
$\mu_{BT} = \mu_{RF}$	4.05e-3 (t -ve)	1.68e-1	1.94e-3	1.55e-3
$\mu_{BT} = \mu_{BDT}$	1.13e-3	8.65e-4	1.46e-3	2.69e-3
$\mu_{RF} = \mu_{BDT}$	9.6e-4	2.06e-2	5.73e-2	0.146

Table 15: p-values for paired t-tests for varying depth limits

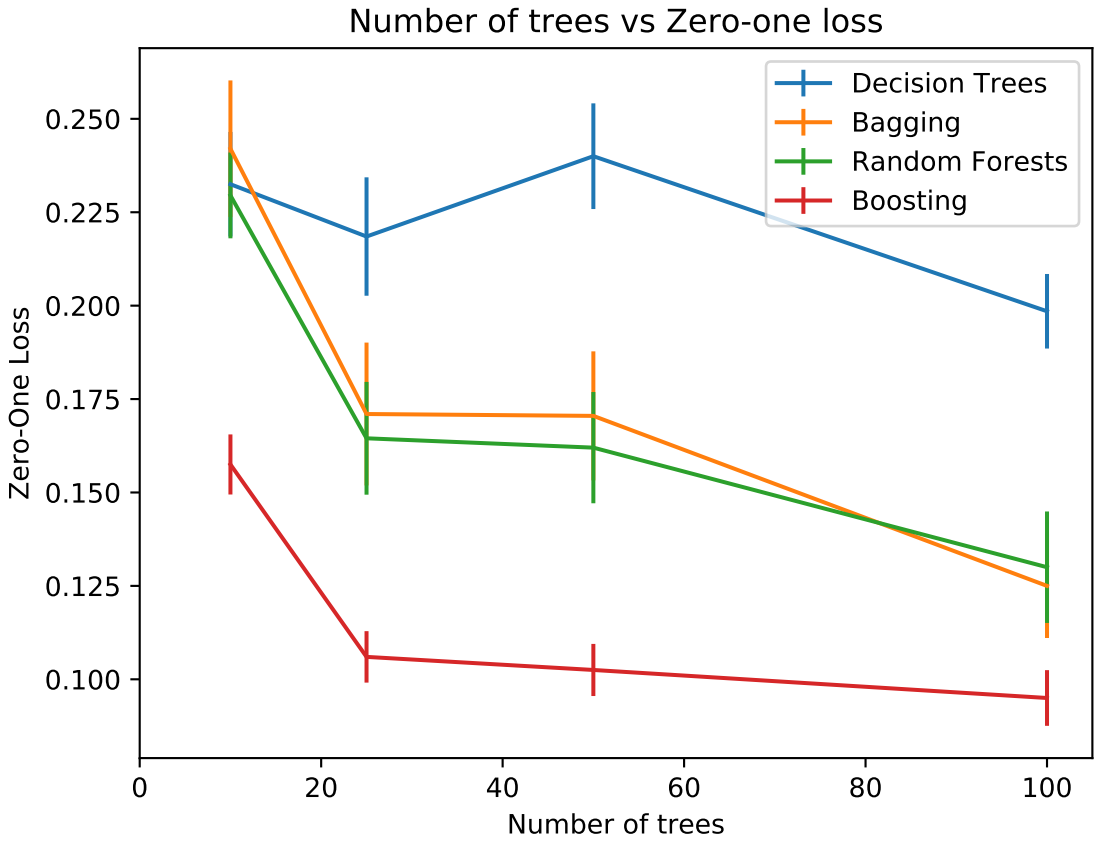

Figure 5: Number of trees vs Zero-one loss

trees.

4. (b)

Tables 16-19 represent the zero-one loss across the the incremental CV folds for each model and each different number of trees.

Let μ_{BT} refer to mean zero-one loss of the BT classifier and μ_{DT} refer to mean zero-one loss of the DT classifier.

Null Hypothesis (H_0): $\mu_{DT} = \mu_{BT}$ (or $\mu_{DT} = \mu_{RF}$ or $\mu_{DT} = \mu_{BDT}$)

Alternative Hypothesis (H_1): $\mu_{DT} > \mu_{BT}$ (or $\mu_{DT} > \mu_{RF}$ or $\mu_{DT} > \mu_{BDT}$)

From the graph, we see that the DT classifier has a higher 0/1 loss than all other models. This difference is significant (except for DT vs RF with 5 trees and DT vs BT with 5 trees)

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.24	0.235	0.165	0.31	0.27	0.22	0.185	0.265	0.175	0.26
BT	0.175	0.225	0.17	0.205	0.225	0.24	0.215	0.34	0.285	0.34
RF	0.205	0.25	0.255	0.185	0.175	0.3	0.225	0.26	0.235	0.205
BDT	0.115	0.2	0.18	0.14	0.145	0.155	0.14	0.16	0.195	0.145

Table 16: Zero-one loss for the 10 cross-validations runs for four models with 10 trees

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.225	0.155	0.33	0.245	0.185	0.195	0.28	0.19	0.19	0.19
BT	0.13	0.17	0.235	0.155	0.14	0.155	0.32	0.135	0.095	0.175
RF	0.12	0.19	0.25	0.15	0.17	0.165	0.235	0.155	0.12	0.09
BDT	0.11	0.12	0.135	0.1	0.125	0.12	0.11	0.105	0.075	0.06

Table 17: Zero-one loss for the 10 cross-validations runs for four models with 25 trees

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.215	0.2	0.295	0.17	0.265	0.255	0.27	0.28	0.28	0.17
BT	0.16	0.205	0.21	0.105	0.115	0.22	0.115	0.245	0.23	0.1
RF	0.105	0.225	0.21	0.105	0.115	0.145	0.155	0.235	0.19	0.135
BDT	0.09	0.125	0.085	0.07	0.13	0.135	0.115	0.11	0.085	0.08

Table 18: Zero-one loss for the 10 cross-validations runs for four models with 50 trees

because the standard error bars of DT do not overlap and are sufficiently far away from those of other classifiers.
Alternatively, we can perform a one-tailed paired t-test for each different number of trees. We will choose our significance $\alpha = 0.05$. In order to correct for testing multiple hypotheses, we apply Bonferroni's correction. We reject the null hypothesis if the p-value is less than $\frac{\alpha}{4} = 0.0125$

From table 20, we see that we reject the null hypothesis for all different number of trees, except for DT vs BT with 5 trees and DT vs RF with 5 trees.

(5)
Let t be the true class of the single example and y be the predicted class. The expected square

Model/Fold	1	2	3	4	5	6	7	8	9	10
DT	0.175	0.14	0.17	0.205	0.23	0.245	0.225	0.175	0.195	0.225
BT	0.08	0.09	0.125	0.225	0.085	0.185	0.105	0.115	0.105	0.135
RF	0.065	0.09	0.17	0.22	0.11	0.135	0.075	0.115	0.185	0.135
BDT	0.06	0.075	0.1	0.075	0.135	0.135	0.09	0.1	0.1	0.08

Table 19: Zero-one loss for the 10 cross-validations runs for four models with 100 trees

Null Hypothesis/Number of trees	10	25	50	100
$\mu_{DT} = \mu_{BT}$	0.335 (t -ve)	6.5e-3	8.74e-4	3.26e-4
$\mu_{DT} = \mu_{RF}$	0.45	1.98e-3	4.05e-4	1.99e-3
$\mu_{DT} = \mu_{BDT}$	2.32e-3	2.74e-5	2.24e-6	5.09e-7

Table 20: p-values for paired t-tests for varying number of trees

loss for the single example can be written as:

$$
\begin{aligned}
E[L_{sq}(t,y)] &= E[(t-y)^2] \\
&= E[((t-E[t]) + (E[t]-E[y]) + (E[y]-y))^2] \\
&\text{Let } f = E[t] = \text{Optimal prediction without noise} \\
&\bar{f} = E[y] = \text{Mean prediction of the model} \\
E[L_{sq}(t,y)] &= E[((t-f) + (f-\bar{f}) + (\bar{f}-y))^2] \\
&= E[(t-f)^2] + E[(f-\bar{f})^2] + E[(\bar{f}-y)^2] + E[2(t-f)(f-\bar{f})] + \\
&E[2(t-f)(\bar{f}-y)] + E[2(f-\bar{f})(\bar{f}-y)] \\
&= E[(t-f)^2] + (f-\bar{f})^2 + E[(\bar{f}-y)^2] + 2(f-\bar{f})E[(t-f)] + \\
&2E[(\bar{f}-y)]E[(t-f)] + 2(f-\bar{f})E[(\bar{f}-y)] \\
&\text{...Assuming independence of t and y (predicted class and true class) and f and } \bar{f} \text{ are constants} \\
&= E[(t-f)^2] + (f-\bar{f})^2 + E[(\bar{f}-y)^2] + 2(f-\bar{f})(E[t]-f) + \\
&2(\bar{f}-E[y])(E[t]-f) + 2(f-\bar{f})](\bar{f}-E[y]) \\
&= E[(t-f)^2] + (f-\bar{f})^2 + E[(\bar{f}-y)^2] + 2(f-\bar{f})(f-f) + \\
&2(\bar{f}-\bar{f})(f-f) + 2(f-\bar{f})](\bar{f}-\bar{f}) \\
&= E[(t-f)^2] + (f-\bar{f})^2 + E[(\bar{f}-y)^2] + 0 + 0 + 0 \\
&= E[(t-f)^2] + (f-\bar{f})^2 + E[(\bar{f}-y)^2] \\
&= Noise + (Bias)^2 + Variance
\end{aligned}
$$

Thus the decomposition is complete.
The bias term is $f - \bar{f} = E[t] - E[y]$.
The noise term is $E[(t-f)^2] = E[(t-E[t])^2]$.
The variance term is $E[(\bar{f}-y)^2] = E[(E[y]-y)^2]$.