CS 57300

Hengrui Zhang

4/15/17

Analysis:

1. Assess whether ensembles improve performance.

a)

Learning curves for the five models based on incremental CV:

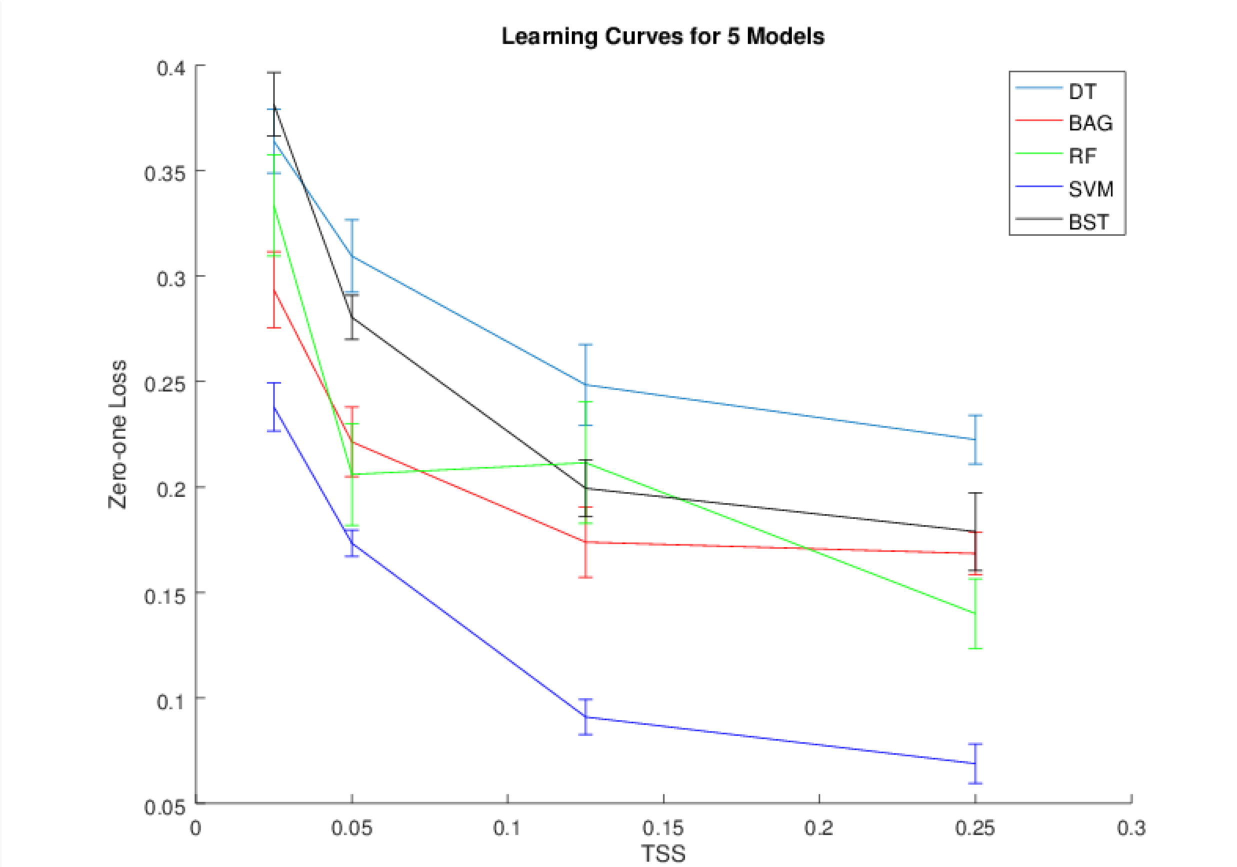


Figure1. Learning curves for 5 models based on different training set size

b).

In this question, SVM was compared to Random Forest regarding the performance on Zero-one Loss based on incremental cross validation. The hypothesis is:

$$H_0 : \mu_{SVM} = \mu_{RF}$$
$$H_a : \mu_{SVM} < \mu_{RF}$$

Here, µ refers to the mean of Zero-one Loss. A paired T test will be suitable in this situation. The situations of four different sizes of training size will be considered respectively. The procedure will be in the procedure as follows. Take TSS = 0.05 as an example:

k	μ_{svm}	μ_{RF}	Diff
1	0.16	0.33	-0.17
2	0.19	0.28	-0.09
3	0.215	0.23	-0.015
4	0.15	0.115	+0.035
5	0.185	0.175	+0.01
6	0.175	0.165	+0.01
7	0.165	0.15	+0.015
8	0.16	0.12	+0.04
9	0.155	0.19	-0.035
10	0.19	0.305	-0.115

Choose α = 0.05 as the significant level and calculate the t test statistics, which equals to

$$t^* = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = -1.412$$

Check the t table and found p = 9.5789e-2 (p>0.05). Hence, H_0 is concluded at α = 0.05 significant level.

For different training set sizes:

TSS	P-value
0.025	4.615e-3
0.05	9.5789e-2
0.125	3.491e-3
0.25	7.86e-4

Using Bonferroni correction here for multiple comparisons.

$$\alpha' = \alpha / 4 = 0.0125$$

Compare p-values with α’, we notice that H_0 are rejected for all of the training dataset except when the TSS is 0.05.

2. Assess whether the number of features improve performance.

a).

Learning curves for the five models with different feature size:

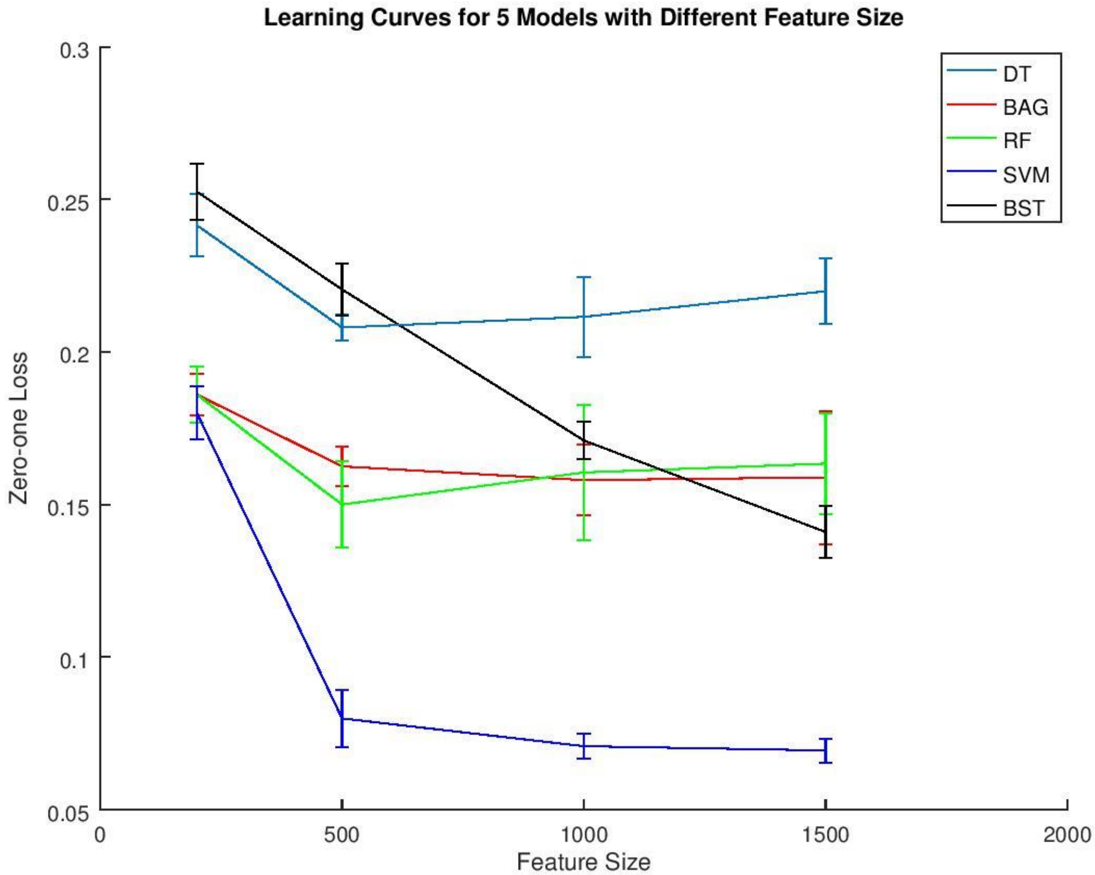


Figure2. Learning curves for 5 models with different feature size

b).

In this question, SVM was compared to Bagged Decision Tree regarding the performance on Zero-one Loss based on varying numbers of features from 200, 500, 1000 to 1500. The hypothesis is:

$$H_0 : \mu_{SVM} = \mu_{BAG}$$
$$H_a : \mu_{SVM} < \mu_{BAG}$$

Same as question one, one side paired T test was applied in this situation.

Choose α = 0.05 as the significance level and calculate the t test statistics and relative P-values.

For different number of features:

Number of Features	P-value
200	3.1454e-1

500	7.543e-6
1000	5.265e-5
1500	7.4931e-4

Using Bonferroni correction here for multiple comparisons.

$$\alpha' = \alpha / 4 = 0.0125$$

Compare p-values with α', we notice that H_0 are rejected for all of the training dataset except when the number of features is 200.

3. Assess whether the depth of the tree affects performance.

a). Learning curves for the four models with different depth of tree:

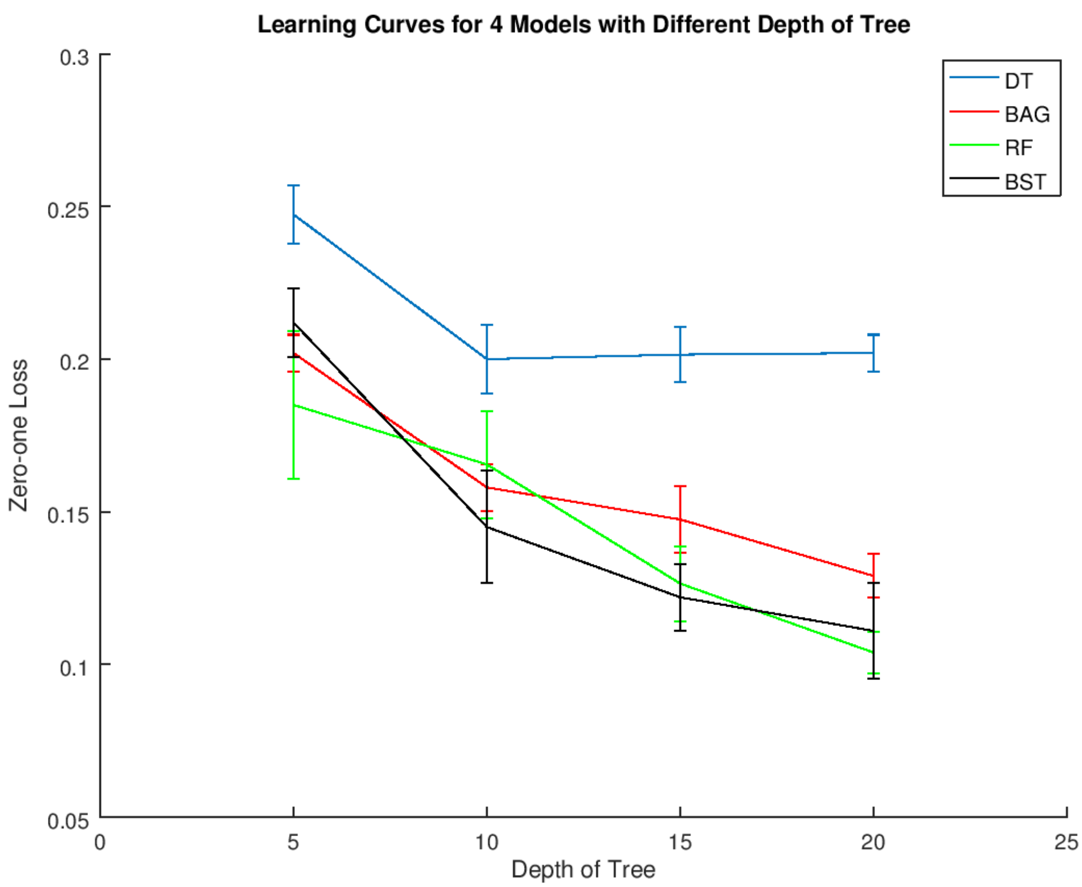


Figure3. Learning curves for 5 models with different depth of tree

b).

In this question, Bagged Decision Tree was compared to the Random Forest with respect to the mean of Zero-one Loss. The hypothesis is:

$$H_0 : \mu_{RF} = \mu_{BAG}$$
$$H_a : \mu_{RF} < \mu_{BAG}$$

One side paired T test was applied in this situation.

Choose α = 0.05 as the significance level and calculate the t test statistics and relative P-values.

For different depths of tree:

Depth of Tree	P-value
5	2.6205e-1
10	4.6774e-2
15	1.1376e-4
20	2.7121e-2

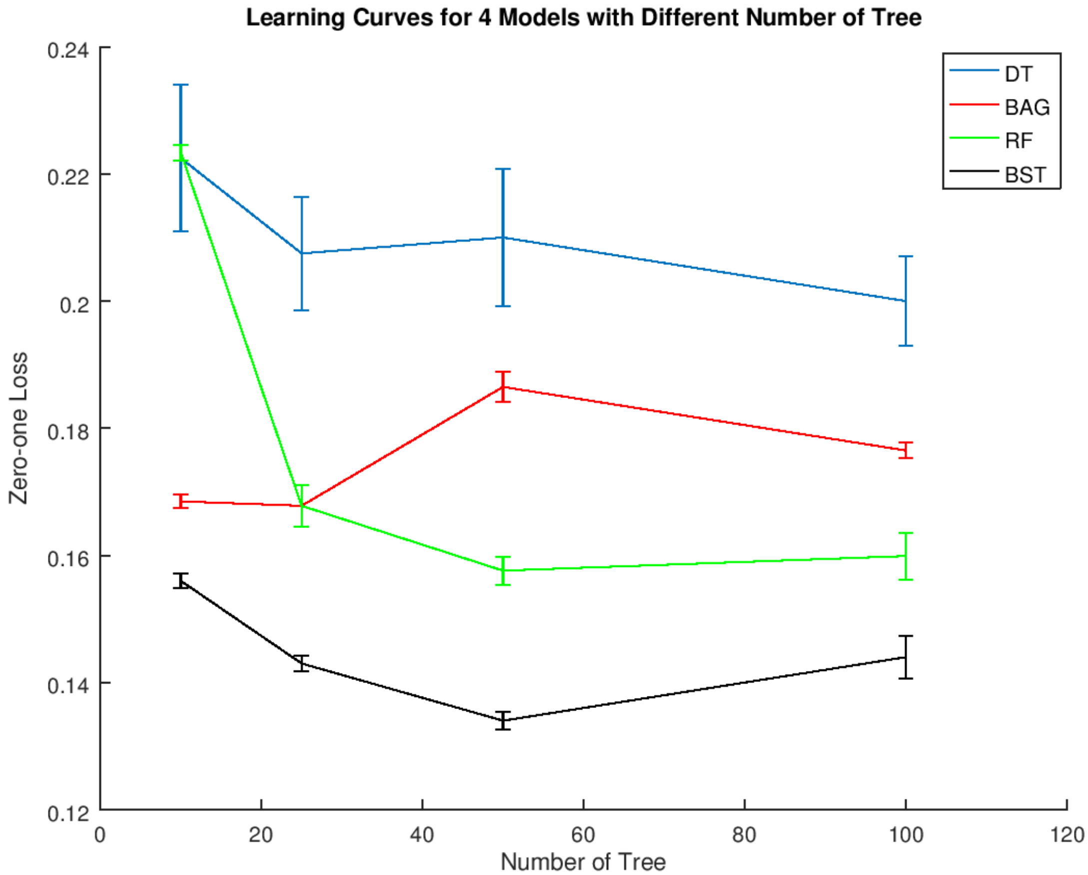
Using Bonferroni correction here for multiple comparisons.

$$\alpha' = \alpha / 4 = 0.0125$$

Compare p-values with α’, we notice that H_0 are rejected when the depth of tree is 15, otherwise will be accepted.

4. Assess whether the number of trees affects performance.

a). Learning curves for the four models with different number of tree:



b).

In this question, Bagged Decision Tree was compared to single Decision Tree with respect to the mean of Zero-one Loss. The hypothesis is:

$$H_0 : \mu_{BAG} = \mu_{DT}$$
$$H_a : \mu_{BAG} < \mu_{DT}$$

One side paired T test was applied in this situation.

Choose α = 0.05 as the significance level and calculate the t test statistics and relative P-values.

For different depths of tree:

Number of Trees	P-value
5	0.91212
10	0.87822
15	0.02223
20	0.0348

Using Bonferroni correction here for multiple comparisons.

$$\alpha' = \alpha / 4 = 0.0125$$

Compare p-values with α', we notice that H_0 are all rejected.

5. Prove that the expected squared loss for a single example can be decomposed into bias/variance/noise.

Let (x^*,y^*) denotes a single example in the dataset, $h(x^*)$ denotes the predicted value of x^*, according to the formula of mean square error, we have:

$$
\begin{aligned}
E[(h(x^*)-y^*)^2] &= E[h(x^*)^2 - 2h(x^*)y^* + y^{*2}] \\
&= E[h(x^*)^2] - 2E[h(x^*)]E[y^*] + E[y^{*2}] && \boxed{\text{(lemma)}} \\
&= E[(h(x^*)-\overline{h(x^*)})^2] + \overline{h(x^*)}^2 - 2\overline{h(x^*)}f(x^*) + E[(y^* - f(x^*))^2] + f(x^*)^2 && \boxed{\text{(lemma)}} \\
&= E[(h(x^*)-\overline{h(x^*)})^2] + && \boxed{\text{[Variance]}} \\
&\quad (\overline{h(x^*)} - f(x^*))^2 + && \boxed{\text{[Bias}^2\text{]}} \\
&\quad E[(y^* - f(x^*))^2] && \boxed{\text{[Noise]}}
\end{aligned}
$$

6. Bonus.

Please refer to Figure1 to Figure 4 to see the result of Boosting Decision Tree compared to other algorithms.

(1) Compare Boosting to SVM

Here we compare Boosting to SVM with respect based on incremental CV.

The hypothesis is:

$$
\begin{aligned}
H_0 &: \mu_{SVM} = \mu_{BST} \\
H_a &: \mu_{SVM} < \mu_{BST}
\end{aligned}
$$

One side paired T test was applied in this situation.

Choose α = 0.05 as the significance level and calculate the t test statistics and relative P-values.

For different training set sizes:

TSS	P-value
0.025	5.64e-5
0.05	1.812e-4
0.125	2.1232e-4
0.25	1.234e-2

Using Bonferroni correction here for multiple comparisons.

$$\alpha' = \alpha / 4 = 0.0125$$

Compare p-values with α', we notice that H_0 are all rejected. But take a look at the situation when TSS equals 0.25, the p-value is quite close to α'. Besides, the whole tendency of p-value is getting close to α'. Probably because when TSS is getting larger, the performance of Boosting will get better.

(2) Compare Boosting to Random Forest

Here we compare Boosting to Random Forest with respect based on incremental CV.

The hypothesis is:

$$H_0 : \mu_{RF} = \mu_{BST}$$
$$H_a : \mu_{RF} < \mu_{BST}$$

One side paired T test was applied in this situation.

Choose α = 0.05 as the significance level and calculate the t test statistics and relative P-values.

For different training set sizes:

TSS	P-value
0.025	6.7888e-1
0.05	2.3444e-3
0.125	3.2322e-1
0.25	2.7633e-4

Using Bonferroni correction here for multiple comparisons.

$$\alpha' = \alpha / 4 = 0.0125$$

Compare p-values with α', we notice that H_0 are accepted when the TSS are 0.025 and 0.125, rejected when TSS are 0.05 and 0.25.