CS573 Data Mining
Homework 4

Fangda Li
li1208@purdue.edu

April 12, 2017

1 Analysis 1

The learning curves for the four models plus SVM can be found in Figure 1. Similar to what has been done in the solution of HW3, we first let μ_{BDT} refer to the mean zero-one-loss of Bagged Decision Trees and μ_{SVM} refer to the mean zero-one-loss of Support Vector Machine. Then, the following null and alternative hypotheses are formed:

$$H_0 : \mu_{BDT} = \mu_{SVM},$$

$$H_0 : \mu_{BDT} < \mu_{SVM}.$$

In order to test the hypothesis, we perform a paired one-tailed t-test for each training set size, with the significance value $\alpha = 0.05$. Moreover, Bonferroni's correction is used and therefore, the null hypothesis can only be rejected if the p-value is smaller than $\frac{\alpha}{4} = 0.0125$. In my experiment, the following p-values are obtained across the training set sizes: 9.816e-07, 1.187e-05, 1.999e-05, 1.828e-06. Since all the p-values appear to be less than 0.0125, we can reject the null hypothesis on every training set size.

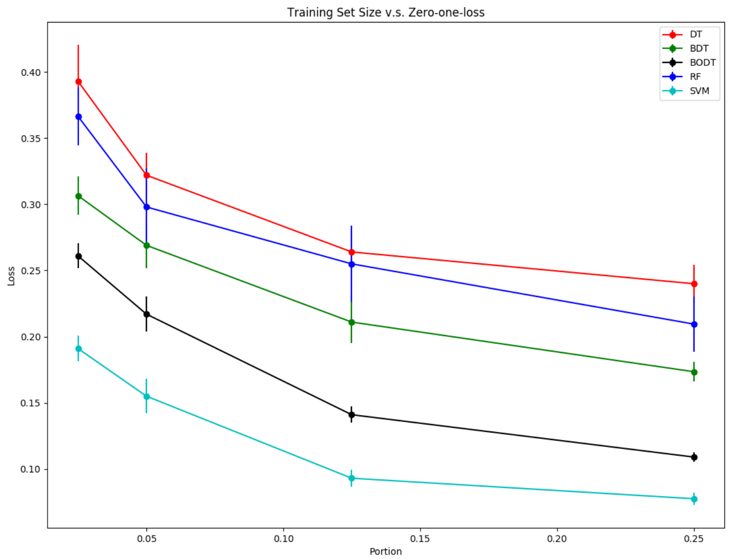

Figure 1: Training set size v.s. zero-one loss. The data point and error bar represent the mean and standard error across 10 folds, respectively.

2 Analysis 2

The learning curves for the four models plus SVM can be found in Figure 2. Identical to Analysis 1, we first let μ_{BDT} refer to the mean zero-one-loss of Bagged Decision Trees and μ_{SVM} refer to the mean zero-one-loss of Support Vector Machine. Then, the following null and alternative hypotheses are formed:

$$H_0 : \mu_{BDT} = \mu_{SVM},$$

$$H_0 : \mu_{BDT} < \mu_{SVM}.$$

In order to test the hypothesis, we perform a paired one-tailed t-test for each number of features, with the significance value $\alpha = 0.05$. Moreover, Bonferroni's correction is used and therefore, the null hypothesis can only be rejected if the p-value is smaller than $\frac{\alpha}{4} = 0.0125$. In my experiment, the following p-values are obtained across the training set sizes: 5.112e-04, 3.206e-06, 2.201e-08, 1.290e-06. Since all the p-values appear to be less than 0.0125, we can reject the null hypothesis on every number of features.

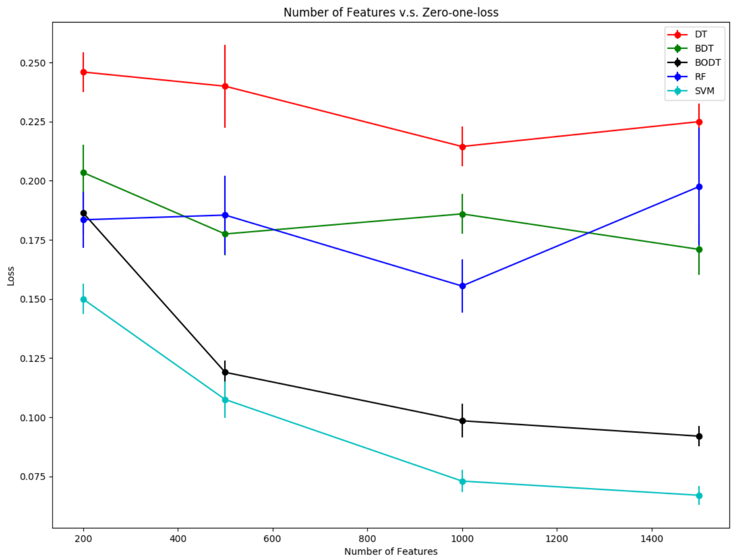

Figure 2: Number of features v.s. zero-one loss. The data point and error bar represent the mean and standard error across 10 folds, respectively.

3 Analysis 3

The learning curves for the four models plus SVM can be found in Figure 3. Identical to Analysis 1, we first let μ_{BDT} refer to the mean zero-one-loss of Bagged Decision Trees and μ_{RF} refer to the mean zero-one-loss of Random Forest. Then, the following null and alternative hypotheses are formed:

$$H_0 : \mu_{BDT} = \mu_{RF},$$

$$H_0 : \mu_{BDT} < \mu_{RF}.$$

In order to test the hypothesis, we perform a paired one-tailed t-test for each depth limit, with the significance value $\alpha = 0.05$. Moreover, Bonferroni's correction is used and therefore, the null hypothesis can only be rejected if the p-value is smaller than $\frac{\alpha}{4} = 0.0125$. In my experiment, the following p-values are obtained across the depth limits: 0.114 0.123 0.035 0.04. Since all the p-values appear to be greater than 0.0125, we cannot reject the null hypothesis on any depth limit.

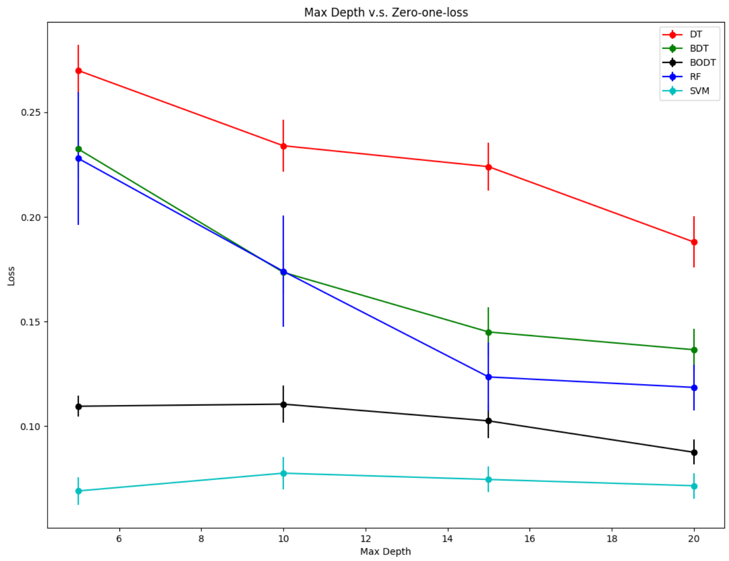

Figure 3: Max depth v.s. zero-one loss. The data point and error bar represent the mean and standard error across 10 folds, respectively. SVM is not affected by the max depth parameter.

4 Analysis 4

The learning curves for the four models plus SVM can be found in Figure 4. Identical to Analysis 1, we first let μ_{DT} refer to the mean zero-one-loss of Decision Tree and μ_{BDT} refer to the mean zero-one-loss of Bagged Decision Trees. Then, the following null and alternative hypotheses are formed:

$$H_0 : \mu_{DT} = \mu_{BDT},$$

$$H_0 : \mu_{DT} < \mu_{BDT}.$$

In order to test the hypothesis, we perform a paired one-tailed t-test for each number of trees, with the significance value $\alpha = 0.05$. Moreover, Bonferroni's correction is used and therefore, the null hypothesis can only be rejected if the p-value is smaller than $\frac{\alpha}{4} = 0.0125$. In my experiment, the following p-values are obtained across the numbers of trees: 0.006, 0.001, 0.001, 0.005. Since all the p-values appear to be less than 0.0125, we can reject the null hypothesis on all numbers of trees.

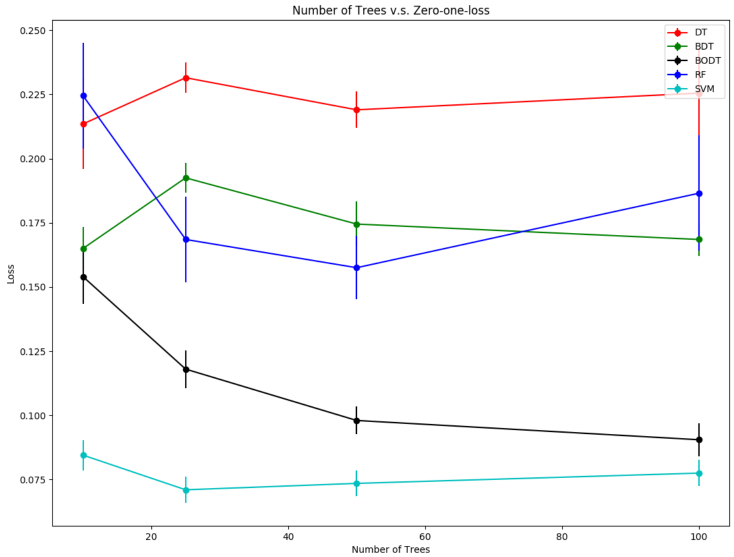

Figure 4: Number of trees v.s. zero-one loss. The data point and error bar represent the mean and standard error across 10 folds, respectively. DT and SVM are not affected by the number of trees parameter.

5 Loss Decomposition

Let t denote the true label of a sample and y denote the predicted label.

$$
\begin{aligned}
&E[(t-y)^2] = E[t^2] + E[y^2] - 2E[ty] \\
&= E[t^2] + E[y^2] - 2E[t]E[y] \text{ (since Y}\perp\!\!\!\perp\text{T)} \\
&= (E[t^2] - E[t]^2) + (E[y^2] - E[y]^2) + (E[t]^2 + E[y]^2 - 2E[t]E[y]) \\
&= (E[t^2] + E[t]^2 - 2E[t]E[t]) + (E[y^2] + E[y]^2 - 2E[y]E[y]) + (E[t]^2 + E[y]^2 - 2E[t]E[y]) \\
&= (E[t^2] + E[E[t]^2] - 2E[tE[t]]) + (E[y^2] + E[E[y]^2] - 2E[yE[y]]) + (E[t]^2 + E[y]^2 - 2E[t]E[y]) \\
&= E[(t-E[t])^2] + E[(y-E[y])^2] + (E[t] - E[y])^2 \\
&= \text{noise} + \text{variance} + \text{bias}.
\end{aligned}
$$

6 Bonus

The learning curve for Boosted Decision Trees across different training set sizes can be found in Figure 1.

(1) As usual, we first let μ_{BODT} refer to the mean zero-one-loss of Boosted Decision Trees and μ_{SVM} refer to the mean zero-one-loss of Support Vector Machine. Then, the following null and alternative hypotheses are formed:

$$H_0 : \mu_{BODT} = \mu_{SVM},$$

$$H_0 : \mu_{BODT} < \mu_{SVM}.$$

In order to test the hypothesis, we perform a paired one-tailed t-test for each training set size, with the significance value $\alpha = 0.05$. Moreover, Bonferroni's correction is used and therefore, the null hypothesis can only be rejected if the p-value is smaller than $\frac{\alpha}{4} = 0.0125$. In my experiment, the following p-values are obtained across the training set sizes: 3.206e-06, 1.683e-05, 1.506e-05, 2.081e-05. Since all the p-values appear to be less than 0.0125, we can reject the null hypothesis on every training set size.

(2) We first let μ_{BODT} refer to the mean zero-one-loss of Boosted Decision Trees and μ_{BDT} refer to the mean zero-one-loss of Bagged Decision Trees. Then, the following null and alternative hypotheses are formed:

$$H_0 : \mu_{BODT} = \mu_{BDT},$$

$$H_0 : \mu_{BODT} > \mu_{BDT}.$$

In order to test the hypothesis, we perform a paired one-tailed t-test for each training set size, with the significance value $\alpha = 0.05$. Moreover, Bonferroni's correction is used and therefore, the null hypothesis can only be rejected if the p-value is smaller than $\frac{\alpha}{4} = 0.0125$. In my experiment, the following p-values are obtained across the training set sizes: 1.052e-03, 6.486e-04, 2.264e-04, 7.831e-06. Since all the p-values appear to be less than 0.0125, we can reject the null hypothesis on every training set size.