

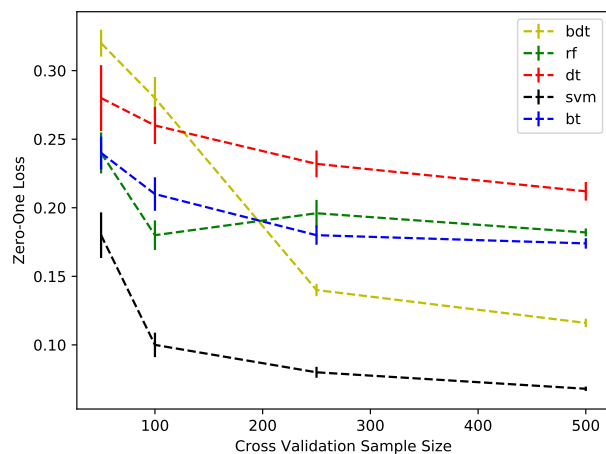
CS 573: Homework 4

Kent Gauen

April 15, 2017

Problem 1: Do Ensembles Improve Performance?

(a)



Here are comments on the learning curves

Figure 1. Learning Curves for three tree models and Support Vector Machine (SVM).

(b)

Formulate hypothesis and perform one-tailed t-test

Problem 2: Do the Number of Features Affect Performance?

(a)

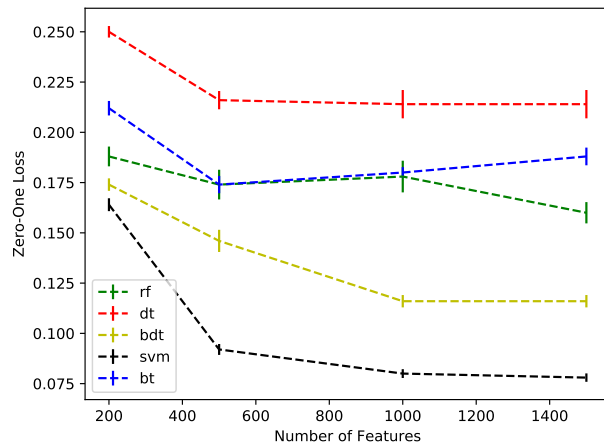


Figure 2. Learning Curves for three models with three-valued feature values: 0, 1, or 2.

Notice that when compared to Figure 1, the standard error bars are much smaller for each model in this figures. This is related to the fact that the data representation has increased amount of information in the features, which has lead to greater consistency in performance on the zero-one loss. The consistency of performance is possibly because of a greater consistency of information in all samples of cross-validation. Notice how the difference in model performance is less pronounced than with fewer features too. This is likely because of the large model space for each model. Finally, we also note the continued trend down, which went up in Figure 1. The models do not seem to overfit the data, and the error seems to still be trending downward.

(b)

For LR, we predict the zero-one loss would decrease overall as more data is introduced. This is because the data is now encoded with more information, which would imply the model can make a more informative decision. However, with smaller amounts of data it might be challenging for the model to make accurate prediction, because more discrepancies in features might take longer to train. With more complicated features, perhaps the number of samples is larger for the error to start to decrease when the features are non-binary. This is analogous to a “complex things take more time to learn” mantra.

(c)

Our initial hypothesis was correct with respect to our guess the overall error would decrease for LR when the 2 was introduced into the data. We were also correct that the LR model did not perform as well with the smaller sample size on the three-valued, when compared to the same sample size with binary features. Another change which we did not predict, is that the standard error smaller for every data sample size. Visually, the difference is significant between Figures 1 and 2. This can be explained by the greater amount of information in the data when the 2 is added. Therefore even in a small sample, there is a greater opportunity for consistency of information. It seems fair to conclude that samples are less likely to be biased when the feature space grows.