Adaptive Random Forest - How many “experts” to ask before making a decision? Supplementary Material

Alexander G. Schwing
ETH Zurich
aschwing@inf.ethz.ch

1. Introduction and Outline

As a prove of concept we provide additional results within this document. It is structured by data set, *i.e.* we provide details for each and every data set used within the submission in a section titled accordingly. Note that for better insight we changed the scaling of the ordinate between different data sets. We further point at some practical issues.

2. Tic-Tac-Toe

The absolute error rate (leave-one-out) of the classifier trained with the settings specified in the submission is 3.4% when using 100 “experts.”

Fig. 1 shows the dependence of the error rate on the average number of classifiers $\bar{\mathrm{E}}$.

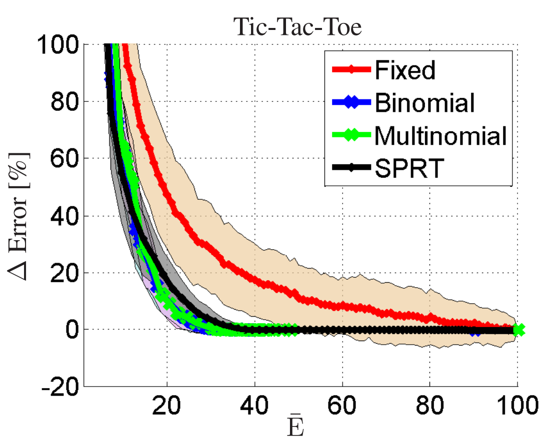

Figure 1. Dependence of the error rate on the average number of classifiers ($\bar{\mathrm{E}}$) for the four specified methods.

Next we show in Fig. 2 the standard deviation of the average number of experts ($\bar{\mathrm{E}}$) for the binomial and the multinomial formulation when choosing a particular confidence value α. When applying $\alpha = 0$ we always use all available “experts.” Note, that the next smallest value we applied is $\alpha = 10^{-4}$ as shown in the figure.

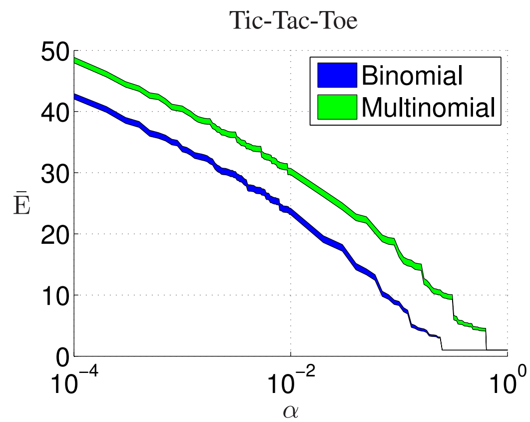

Figure 2. The standard deviation of $\bar{\mathrm{E}}$ depending on the confidence value α.

Finally we provide in Fig. 3 the standard deviation of the average number of experts for the SPRT test when choosing ϵ.

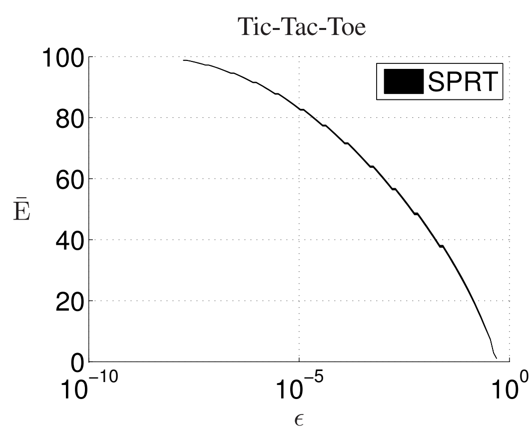

Figure 3. The standard deviation of $\bar{\mathrm{E}}$ for the SPRT test.

3. Ionosphere

As stated in the submission, the absolute error rate (leave-one-out) for 100 “experts” is 6.6%. Similar to the previous dataset considered, we show in Fig. 4 to Fig. 6 the dependence of the classification error on the average number of "experts," the standard deviation of $\bar{\mathrm{E}}$ depending on α and on ϵ for the binomial and multinomial formulation as well as the SPRT procedure.

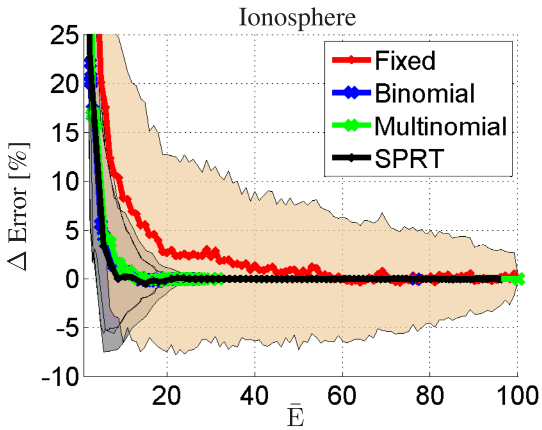

Figure 4. Dependence of the error rate on the average number of classifiers ($\bar{\mathrm{E}}$) for the four specified methods.

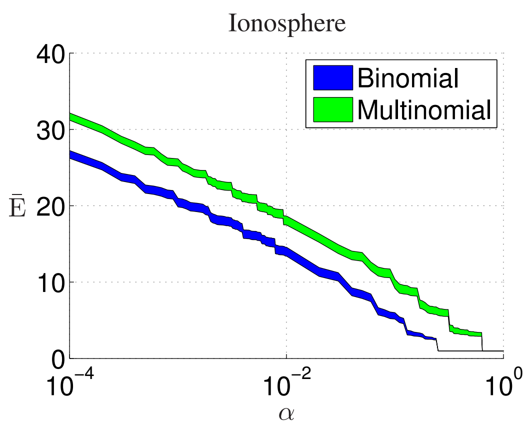

Figure 5. The standard deviation of $\bar{\mathrm{E}}$ depending on the confidence value α.

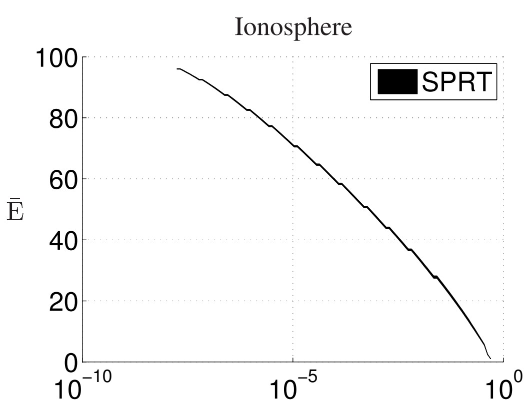

Figure 6. The standard deviation of $\bar{\mathrm{E}}$ for the SPRT test.

4. Iris

The absolute error rate for our leave-one-out experiment with the settings specified in the submission is 4.4%. As the binomial and SPRT formulation are not applicable to problems with more than two classes, the plot giving the dependence of the error rate on the average number of classifiers (Fig. 7) and the standard deviation of $\bar{\mathrm{E}}$ subject to α (Fig. 8) show the multinomial formulation only.

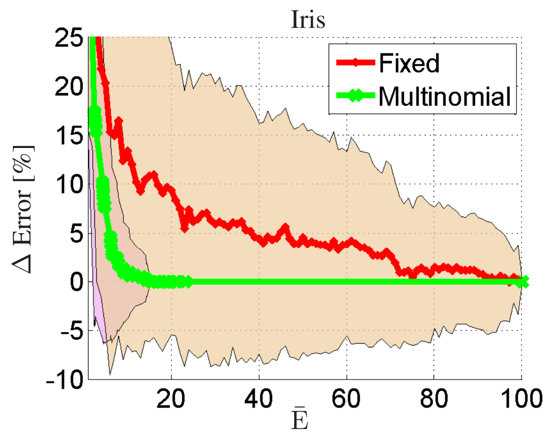

Figure 7. Dependence of the error rate on the average number of classifiers ($\bar{\mathrm{E}}$) for the two methods specified.

Figure 8. The standard deviation of $\bar{\mathrm{E}}$ depending on the confidence value α.

5. Wine

With our leave-one-out experiments we achieved an absolute error rate of 2.1% on this three class data set. Fig. 9 and Fig. 10 show the error plottet against $\bar{\mathrm{E}}$ and the average number of classifiers subject to α respectively.

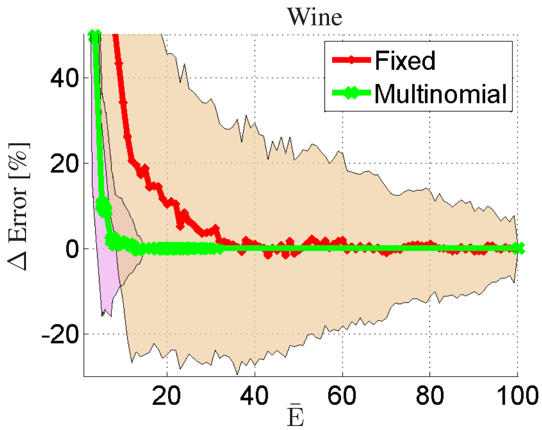

Figure 9. Dependence of the error rate on the average number of classifiers ($\bar{\text{E}}$) for the two methods specified.

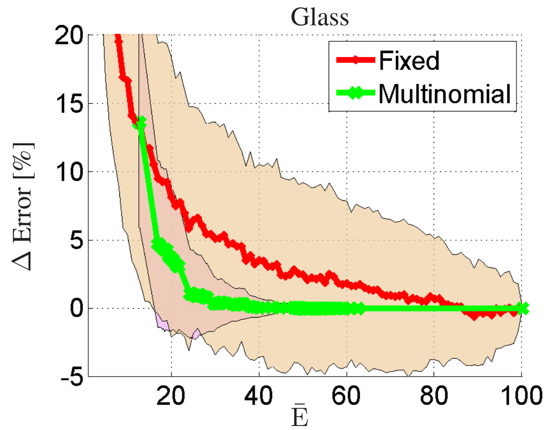

Figure 11. Dependence of the error rate on the average number of classifiers ($\bar{\text{E}}$) for the two methods specified.

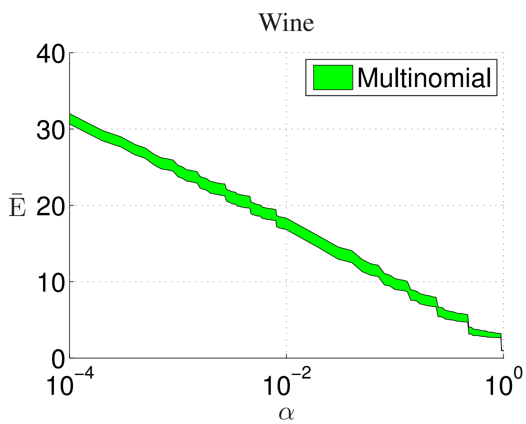

Figure 10. The standard deviation of $\bar{\text{E}}$ depending on the confidence value α.

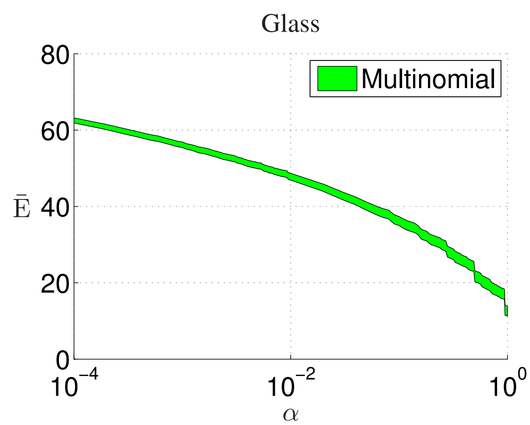

Figure 12. The standard deviation of $\bar{\text{E}}$ depending on the confidence value α.

6. Glass

The absolute error rate achieved on the six label Glass data set is 20.3%. We show the dependence of the error on the average number of classifiers and $\bar{\text{E}}$ plotted against α in Fig. 11 and Fig. 12 respectively.

7. Ecoli

On the eight class Ecoli data set our leave-one-out experiment results in an absolute error rate of 12.3%. Again, the error rate depending on the average number of classifiers (Fig. 13) and the dependence of $\bar{\text{E}}$ on the confidence α (Fig. 14) is provided in respective plots.

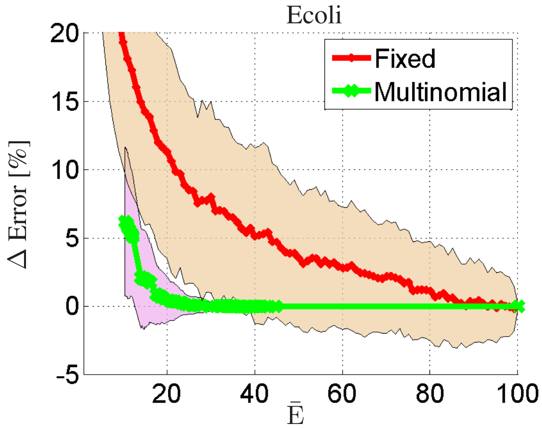

Figure 13. Dependence of the error rate on the average number of classifiers ($\bar{\mathrm{E}}$) for the two methods specified.

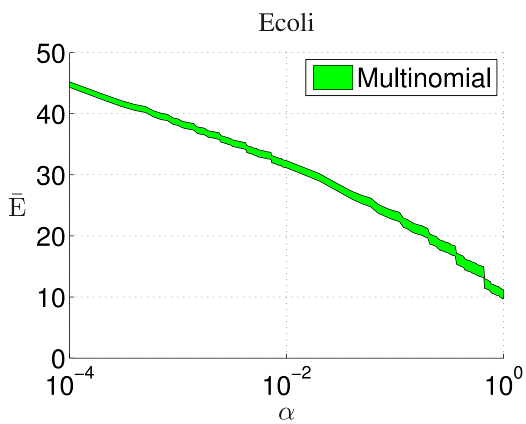

Figure 14. The standard deviation of $\bar{\mathrm{E}}$ depending on the confidence value α.

8. Yeast

We obtain an absolute error rate of 37.7% on this ten class data set. The error rate depending on the average number of classifiers (Fig. 15) and the dependence of $\bar{\mathrm{E}}$ on the confidence α (Fig. 16) is given.

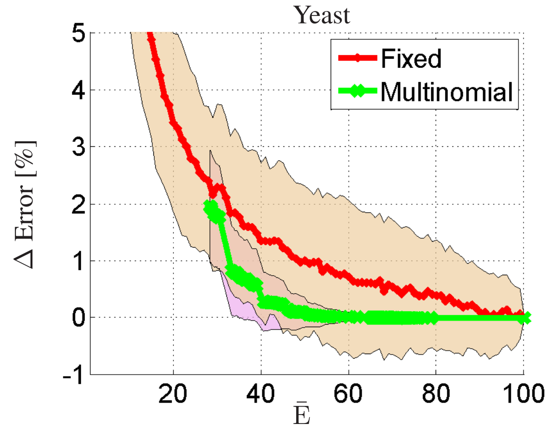

Figure 15. Dependence of the error rate on the average number of classifiers ($\bar{\mathrm{E}}$) for the two methods specified.

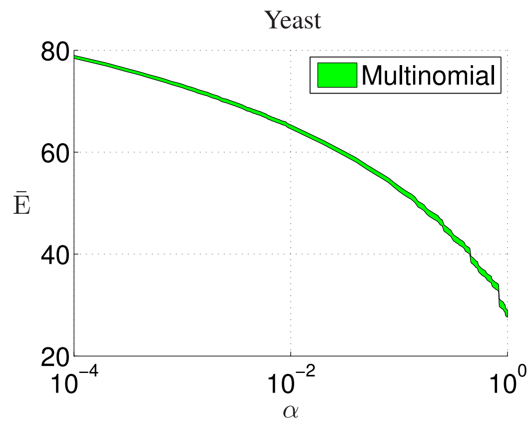

Figure 16. The standard deviation of $\bar{\mathrm{E}}$ depending on the confidence value α.

9. Large Scale Data Set

To allow navigation through the space of α, ϵ and a fixed number of experts E_X we provide ROC curves in a video. Note that the legend in the video also gives the average number of "experts" $\bar{\mathrm{E}}$ consulted during classification of the test set, if applicable.

10. Practical Notes

For all our data sets we show the impact of adjusting the confidence alpha through a wide ranging interval. Nevertheless, we note that it is certainly hard to estimate, directly during the application stage of the classifier, that particular value of alpha, that does not degrade performance w.r.t. all available classifiers. There are two possible solutions: (1) Estimating its value during the training stage as suggested by one of the Reviewers. In case of a Random Forest classifier we can use the out-of-bag samples for obtaining the appropriate confidence alpha. (2) Before shipping an application to the end-user we perform leave-one-out or crossvalidation tests in any case. From those tests we obtain the curves illustrated in e.g. Fig. 1 and Fig. 2. Those plots are used to adjust the operating point, i.e. to obtain an appropriate confidence value alpha. We'd like to highlight that Fig. 2 shows the mean and its standard deviation. With the standard deviation being rather small (for all the evaluated data sets), we obtain a rather tight bound for the confidence alpha given a specific number of average "experts" to be evaluated. Considering the second solution, we have a practical solution for inference of the free parameter alpha.

An additional note on the compatibility of different samples/locations. Classifier outputs of different samples/locations are according to our opinion not incompatible. Unless a sample requires usage of all available classifiers, we already know its confidence beforehand. They all have approximately the confidence α of belonging to the respective class. Hence, the confidence measure has been applied such that the individual samples are actually "equal" w.r.t. this measure. We acknowledge that the equal confidence does not allow a comparison of the different samples. Note, that the number of trees necessary for achieving this confidence can alternatively be used as a measure. To summarize: rather than having an equal number of trees and different confidence for every sample we now have equal confidence and different number of trees.