Ensemble Learning for 5-fold Cross-Validation Performance Report

- Breast Cancer Classification Task -

Ji-seong Yang

I. The Average Accuracy and Standard Deviation Report for 5-fold Cross-Validation[1]

	DT	SVM	LR
ACC	0.917368	0.945564	0.950904
STD	0.010774	0.018714	0.015935

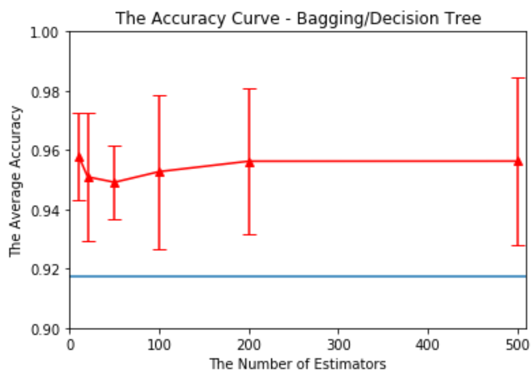
II. The Average Accuracy and Standard Deviation Report for 5-fold Cross-Validation with Bagging

a. Decision Tree

i. Report

	10	20	50	100	200	500
ACC	0.957922	0.950966	0.949134	0.952736	0.956245	0.956275
STD	0.014824	0.021493	0.012463	0.025941	0.024422	0.028409

ii. Visualization



\

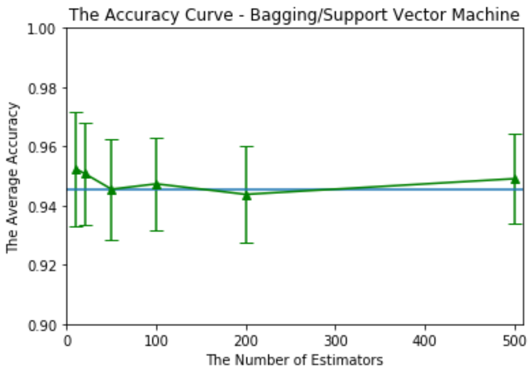
b. Support Vector Machine (SVM)

◎ [1] **DT**: Decision Tree, **SVM**: Support Vector Machine, **LR**: Logistic Regression, **ACC**: Accuracy, **STD**: Standard Deviation

i. Report

	10	20	50	100	200	500
ACC	0.952459	0.950812	0.945533	0.947334	0.943794	0.949073
STD	0.019224	0.017272	0.017097	0.015562	0.016249	0.015061

ii. Visualization

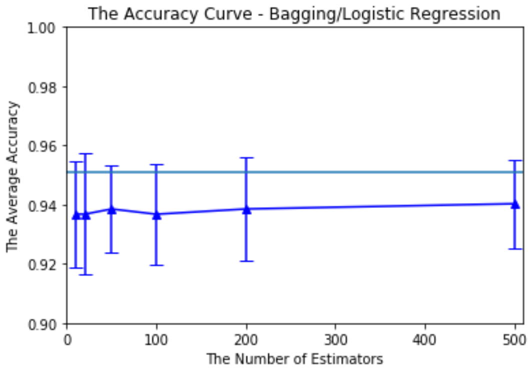


c. Logistic Regression

i. Report

	10	20	50	100	200	500
ACC	0.936776	0.936745	0.938515	0.936745	0.938515	0.940285
STD	0.017866	0.020430	0.014671	0.017090	0.017585	0.015038

ii. Visualization



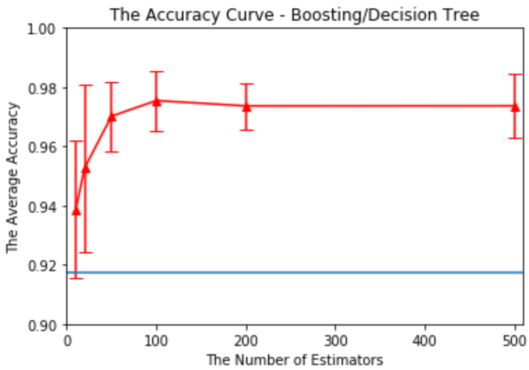
III. The Average Accuracy and Standard Deviation Report for 5-fold Cross-Validation with Boosting

a. Decision Tree

i. Report

	10	20	50	100	200	500
ACC	0.938638	0.952736	0.970189	0.975467	0.973667	0.973698
STD	0.023153	0.028253	0.011771	0.010068	0.007781	0.011002

ii. Visualization

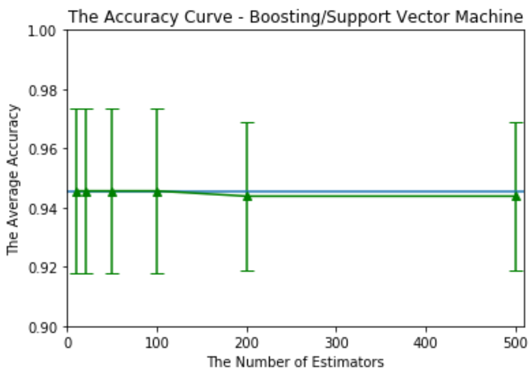


b. Support Vector Machine (SVM)

i. Report

	10	20	50	100	200	500
ACC	0.945625	0.945625	0.945625	0.945625	0.943855	0.943855
STD	0.027857	0.027857	0.027857	0.027857	0.025048	0.025048

ii. Visualization

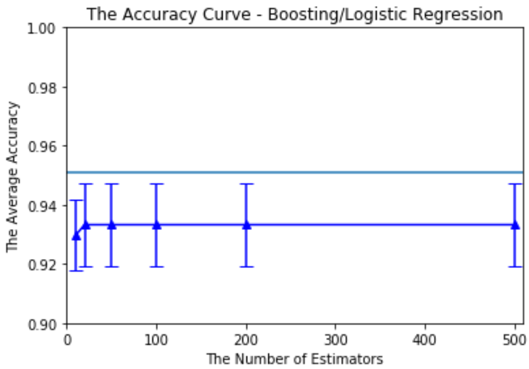


c. Logistic Regression

i. Report

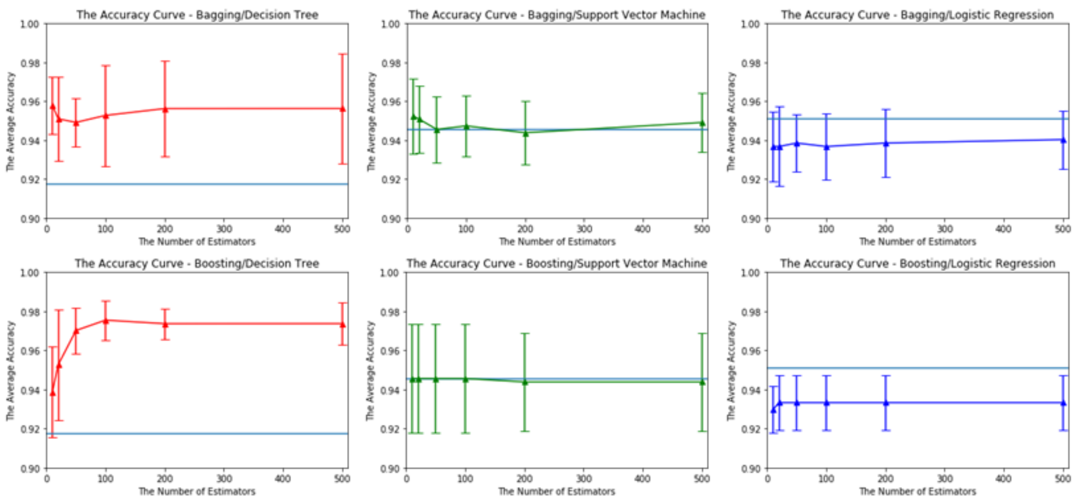
	10	20	50	100	200	500
ACC	0.929758	0.933267	0.933267	0.933267	0.933267	0.933267
STD	0.012091	0.014014	0.014014	0.014014	0.014014	0.014014

ii. Visualization



IV. Visualization Summary

The following six graphs are the aggregation of previously shown accuracy curve visualizations. They will be referred in the next chapter, denoted by their location (upper or lower) and color (red, green, and blue) for the readers' convenience. For example, the accuracy curve for Support Vector Machine by Boosting will be indexed as 'the lower green chart'. Horizontal lines (cyan) in each chart indicate the average accuracy calculated before the performance of bagging and boosting.



V. Findings and Conclusion

a. Will bagging or boosting effectively improve the classification performance for each classifier?

The following tables show an accuracy (ACC) of 5-fold cross-validation and classification using Decision Tree (DT) Support Vector Machine (SVM), and Logistic Regression (LR), before and after applying bagging and boosting. Out of 6 different results by the varying size of estimators, the score of the *highest* accuracy has been chosen to represent each classifier performance.

i. Accuracy Report for Bagging

ACC	**DT**	**SVM**	**LR**
Before	0.917368	0.945564	0.950904
After (Best)	0.957922	0.952459	0.940285
Diff	0.040554	0.006895	-0.010619

The table shows that when compared with the accuracy calculated before *bagging*, DT performance has increased by 040554, SVM by 0.006895, and LR by minus 0.010619. In this particular trial, only DT and SVM has resulted in the improvement in the accuracy as one may also notice in three upper charts in chapter 4.

ii. Accuracy Report for Boosting

ACC	**DT**	**SVM**	**LR**
Before	0.917368	0.945564	0.950904
After (Best)	0.975467	0.945625	0.933267
Diff	0.058099	0.000061	-0.017637

The table suggests the accuracy scores in comparison to their counterparts calculated before *boosting*, DT performance has improved by 0.058099, SVM by 0.000061, whereas LR decreased by 0.017637. In this trial, only DT and SVM has resulted in the enhancement in the accuracy as the three lower charts in chapter 4 demonstrate.

iii. Conclusion

Without the consideration of errors, it can be concluded that bagging and boosting will improve the classification performance of DT and SVM, whereas it is not the case for LR.

b. Will bagging or boosting effectively improve the classification stability in terms of the standard deviation of the 5-fold cross validation classification accuracy?

The following tables show standard deviation (STD) of 5-fold cross-validation and classification using Decision Tree (DT) Support Vector Machine (SVM), and Logistic Regression (LR), before and after adopting bagging and boosting. Out of 6 different results by the varying size of estimators, the *lowest* standard deviation has been chosen to represent each classifier performance.

i. Standard Deviation Report for Bagging

STD	**DT**	**SVM**	**LR**
Before	0.010774	0.018714	0.015935
After (Least)	0.012463	0.015061	0.014671
Diff	0.001689	-0.001264	-0.001264

The table indicates that when compared with the standard deviation calculated before *bagging*, DT performance stability has decreased by 0.001689 while SVM and LR improved their stability by 0.001264. In this particular trial, only SVM and LR have resulted in the stability improvement.

It can be noted in the upper red chart, however, intervals of errors of DT does not overlap the average accuracy score before *bagging*. This means that the performance of bagging will always better than that without it, regardless of the trial.

ii. Standard Deviation Report for Boosting

STD	**DT**	**SVM**	**LR**
Before	0.010774	0.018714	0.015935
After (Least)	0.007781	0.025048	0.012091
Diff	-0.002993	0.006334	-0.003844

The table reports that there has been an instability raise in SVM by 0.006334 while DT and LR have reduced the variance by 0.002993 and 0.003844, respectively. However, it is difficult to elicit a conclusion from this information, so further exploration on data of the visualized from would be required. Most of the intervals of errors of DT, except only one, does not overlap the average accuracy score before *boosting*. This means that bagging would allow the model to outperform the other models without it.

iii. Conclusion

All in all, it seems that it is DT that would greatly be boosted with the aid of bagging and boosting while SVM does not do much for the performance. LR is not recommended to use with either bagging or boosting for it would rather degrade the level of performance, in terms of the accuracy.

c. With the increasing of the number of bagging/boosting estimators, will the performance converge (convergence means the cross-validation performance will not increase with more estimators)?

i. Bagging

Each classifier performance reaches its peak at a certain number of the estimator. DT (upper red chart) performs its best with 10 estimators and continues to drop until the number of estimators is 50. Though it slowly increases from that point, the rate of the change is as minimal as 0.00003. It is very likely that the accuracy would converge afterward.

SVM (upper green chart) also scores it best at 10 estimators, however, declines until the number of estimators is 200. It then steadily goes up and scores 0.95 (rounded to the tenth place) at 500 estimators. Further observation is needed in order to figure out if it would converge or not, however, it is expected to converge soon.

LR (upper blue chart), after several ups and downs in the initial stage, reaches its peak at 500 estimators, which is the last label of the x-axis. Though it may continue to increase for a while, due to the low increment rate, it would also meet its eventual convergence.

ii. Boosting

Each classifier performance also reaches its peak at a certain number of the estimator as it was the case in bagging. DT (lower red chart) performs its best with 100 but declines at 200. There is a slight boost in 500 in comparison with 200, but the difference insignificant. It is a sign of convergence.

As for SVM (lower green chart), the performance worsens as the value of x-axis get greater, converging to the value of 0.94. The same applies to the LR (lower blue chart) in that the value since 20 estimators remains unchanged until 500 with the accuracy of approximately 0.93. (rounded to the tenth place)

d. Do you observe overfitting with the increase in the number of estimators for bagging/boosting?

i. Bagging

In terms of accuracy, the is no sign of drastic decline in all of three classifiers. On the other hand, the standard deviation which is the square root of the variance, decrease or stay the same. Signs of overfitting have not been spotted. Chances are the model optimization is still underway, therefore, additional observation is required with different optional parameters trial on machine learning functions.

ii. Boosting

In terms of accuracy, the is no sign of drastic decline in all of three classifiers. On the other hand, as for the standard deviation which is the square root of the variance, DT (lower red chart) greatly reduces the size of error at the initial stage of learning, however, it starts to re-enlarge at 500, which could imply overfitting.

As for SVM and LR, standard deviation remains the same all across the learning progress. Rather they seem to bear underfitting problem than overfitting because the accuracy of the both got worse than they were before boosting.