An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking)

Rising O. Odegua

Ambrose Alli University, Computer Science Department

INTRODUCTION AND MOTIVATION

The technique of combining the predictions of multiple models have long be investigated by researchers. This area popularly known as ensemble technique has been demonstrated to outperform a single best model in most tasks. A good ensemble is one where all the individual models are both accurate and diverse in their error.

Consider a supervised learning problem where we are given training example *S* of the form $\{(x_1, y_1), (x_1, y_1), \ldots, (x_n, y_n)\}$ for some function $y = f(x)$. The *x* values is usually a multidimensional matrix called features while the *y* values typically drawn from a continuous set of numbers for regression problem and a distinct set of classes for a classification problem is called the target.

Given *S* of such training example, a learning algorithm *h* tries to learn the true function *f* that validates some hypothesis, so that given a new *x* (test examples), the algorithm can predict the corresponding *y* values. An ensemble basically looks for the best way to combine individual learning algorithms $(h_1, h_2, \ldots, h_L)$ to predict new examples typically by weighted or unweighted voting.

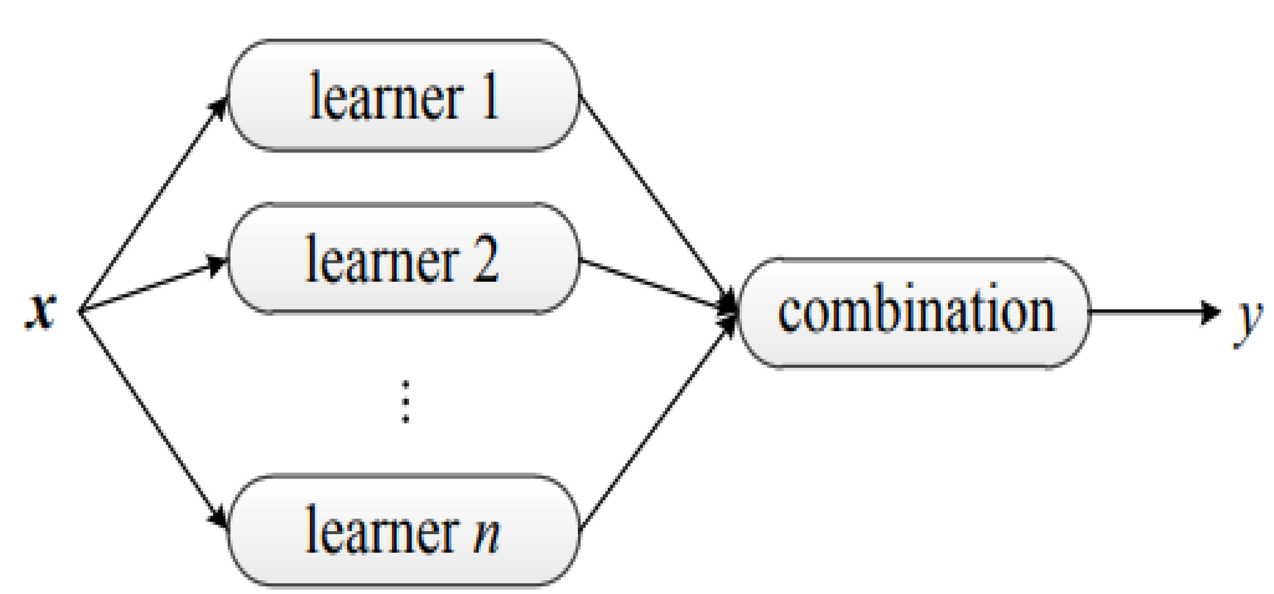

Fig 1. An ensemble architecture

METHODOLOGY

- First, we performed basic data preparation and cleaning. These included taking care of missing values, and encoding categorical features. We avoided extensive feature engineering on the data sets.
- Secondly, we trained 4 single base models (Linear Regression, Random Forest, Extra Trees, KNN) on our data sets and measured their mean absolute error. We used a 10-fold cross validation strategy.
- Thirdly, we performed bagging using each of the single base models listed above and also measured their errors.
- Fourthly, we used 3 boosting implementations; the vanilla Ada-Boost and modern techniques like GradientBoosting and LightGBM. And finally, we performed stacking using Random Forest, Extra Trees, KNN, Extra Trees, GradientBoosting and LightGBM as first level learners and LightGBM as the meta-learner.

DATASETS

S/N	Data sets	Instances	Total Features	Features Num .	Features Cat.	Features Size after encoding
1	German Bank Credit	1000	21	12	9	63
2	Automobile Pricing	195	24	14	10	67
3	Avocado	18249	11	9	2	66
4	Real Estate	414	5	5	-	-
5	Loan default DSN	4368	30	12	18	130
6	House Pricing	1458	80	26	54	221
7	Retweets pred data	89309	19	8	11	95
8	Traffic prediction kenya	51645	10	2	8	69
9	Chukwudi Supermarkets	4990	12	4	8	33

CONCLUSION

- ➢ Bagging ensemble nearly always outperforms a single classifier.
- ➢ Boosting ensemble on the average will outperform both Bagging and a single classifier.
- ➢ Stacking ensemble nearly always outperform Boosting, Bagging and a single base classifier.
- ➢For some data sets, Boosting and Stacking may show zero gain or even a decrease in performance from a single classifier.
- ➢Boosting may suffer from overfitting in the presence of noise which explains some of the decrease in performance.

RESULTS

	Single Base Models				Bagging				Boosting			Stacking
DATASETS	LR	KNN	RF	ET	LR_BG	KNN_BG	RF_BG	ET_BG	GB	ADAB	LGB	STACKED MODEL
Loan_Data	0.321	0.324	0.328	0.328	0.319	0.314	0.311	0.32	0.303	0.355	0.293	0.3
House_Pricing	2.457	0.164	0.097	0.089	0.082	0.161	0.094	0.084	0.08	0.102	0.079	0.079
Tweets_Data	6.503	5.706	6.133	6.124	6.4	5.787	6.118	6.102	5.822	20.937	5.768	5.611
Traffic_Data	15.5	4.833	4.367	4.393	15.13	4.35	4.32	4.251	4.104	4.147	4.059	4.027
Avocado_Data	0.145	0.227	0.128	0.137	0.145	0.225	0.122	0.13	0.123	0.21	0.115	0.132
Real_Estate	5.654	5.31	4.768	4.571	5.611	5.258	4.315	4.314	4.558	5.429	4.62	4.277
AutoMobile	4848.5	3649.4	2253.3	2199.3	4234.3	3641.1	2217.8	2101.1	2184.8	2209.5	2199.9	2101.0
German_Cred	7.55	10.544	7.36	8.004	7.617	10.256	7.181	7.225	7.725	8.531	7.854	7.276
Supermarket	75.732	81.908	80.974	86.29	75.627	80.93	79.418	81.093	74.80	75.16	74.47	71.458

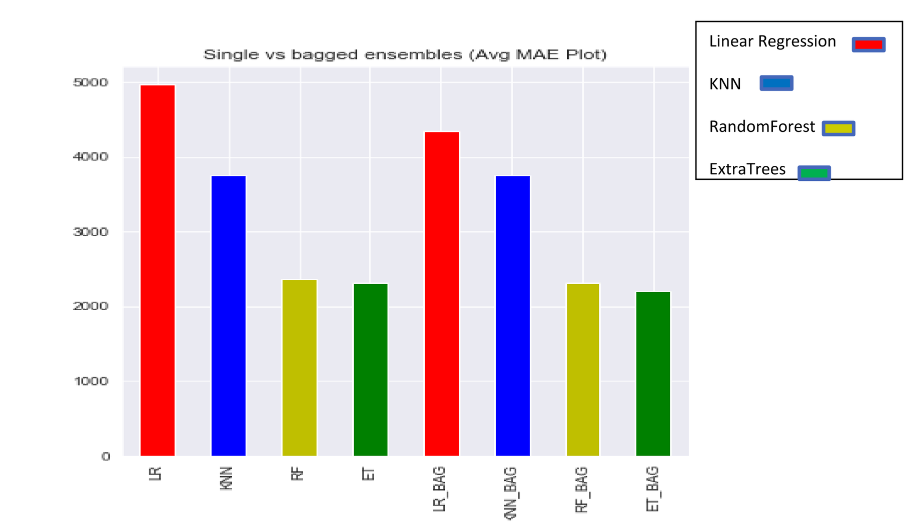

Fig.2 Comparison of Single and Bagging Ensemble

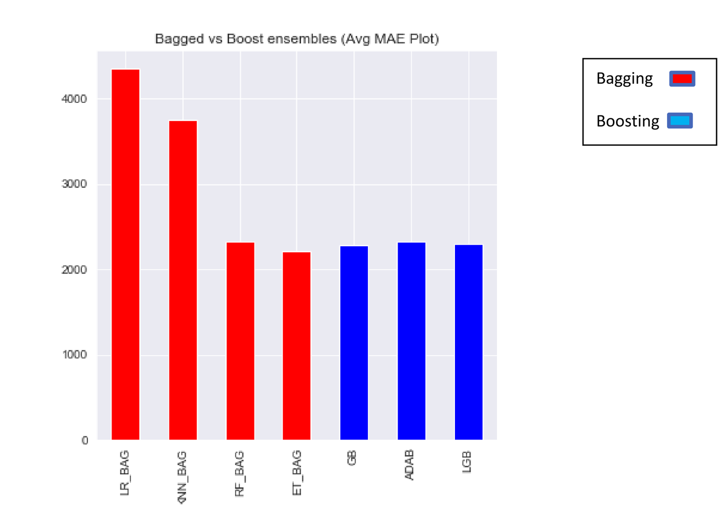

Fig.3 Comparison of Bagging and Boosting Ensembles

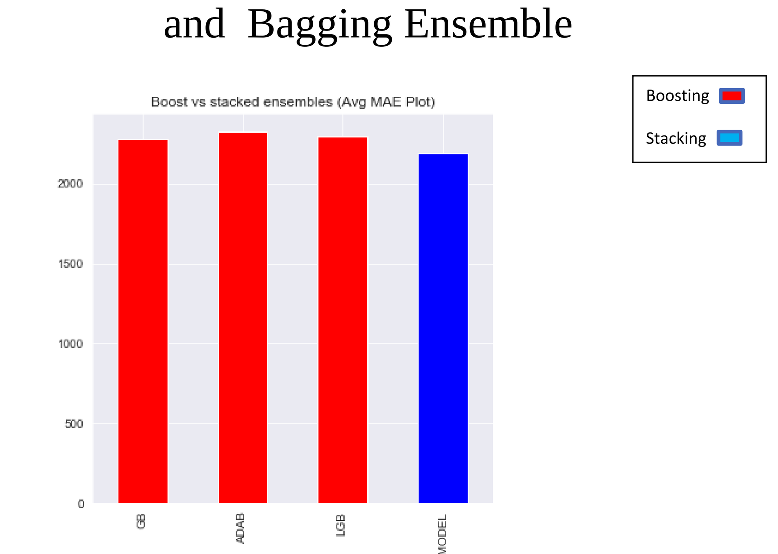

Fig.4 Comparison of Boosting and Stacking Ensemble

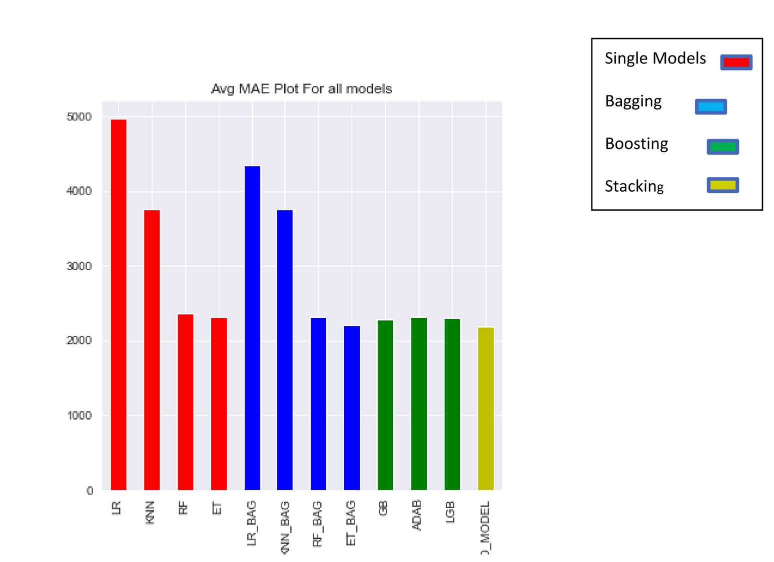

Fig.5 Comparison of all models

REFERENCES

Breiman, L. (1996a). Bagging predictors. Machine Learning, 24 (2), 123–140.

Dietterich, T. G. Ensemble methods in machine learning. In Proceedings of the 1st Workshop on Multiple Classifier Systems, pages 1–15, Sardinia, Italy, 2000a.

Efron, B., & Tibshirani, R. (1993). An Introduction to the Bootstrap. Chapman and Hall, NY

Freund, Y., & Schapire, R. (1996). Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning, pp. 148–156 Bari

Z.-H. Zhou and M. Li. Ensemble Methods. Foundations and Algorithms, -13: 978-1-4398-3005-5, 2012.