## Acknowledgements

We are most grateful to the anonymous reviewers who pointed a number of typos and inaccuracies and whose remarks significantly improved this paper and its full version.

This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61361136003, 61350110536.

## References

Bauer, E., and Kohavi, R. 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning* 36(1-2):105–139.

Biau, G.; Cérou, F.; and Guyader, A. 2010. On the rate of convergence of the bagged nearest neighbor estimate. *The Journal of Machine Learning Research* 11:687–712.

Biau, G.; Devroye, L.; and Lugosi, G. 2008. Consistency of random forests and other averaging classifiers. *The Journal of Machine Learning Research* 9:2015–2033.

Breiman, L. 1996. Bagging predictors. *Machine learning* 24(2):123–140.

Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.

Bühlmann, P., and Yu, B. 2002. Analyzing bagging. *The Annals of Statistics* 30(4):927–961.

Colbourn, C. J., and Dinitz, J. H. 2010. *Handbook of combinatorial designs*. CRC press.

Denil, M.; Matheson, D.; and de Freitas, N. 2013. Narrowing the gap: Random forests in theory and in practice. *arXiv preprint arXiv:1310.1415*.

Dietterich, T. G. 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning* 40(2):139–157.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The annals of Statistics* 1–26.

Freund, Y., and Schapire, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1):119–139.

Friedman, J. H., and Hall, P. 2007. On bagging and nonlinear estimation. *Journal of statistical planning and inference* 137(3):669–683.

Hartman, T., and Raz, R. 2003. On the distribution of the number of roots of polynomials and explicit weak designs. *Random Structures & Algorithms* 23(3):235–263.

Kuncheva, L. I., and Whitaker, C. J. 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning* 51(2):181–207.

Long, P., and Servedio, R. 2013. Algorithms and hardness results for parallel large margin learning. *Journal of Machine Learning Research* 14:3073–3096. (also NIPS'11).

Melville, P.; Shah, N.; Mihalkova, L.; and Mooney, R. J. 2004. Experiments on ensembles with missing and noisy data. In *Multiple Classifier Systems*. Springer. 293–302.

Tang, E. K.; Suganthan, P. N.; and Yao, X. 2006. An analysis of diversity measures. *Machine Learning* 65(1):247–271.

Ting, K. M., and Witten, I. H. 1997. Stacking bagged and dagged models. In *ICML*, 367–375.

Valiant, L. G. 1984. A theory of the learnable. *Communications of the ACM* 27(11):1134–1142.