





















- Breiman, L. J.; Friedman, J. H.; Olshen, R. A.; and Stone, C. J. 1984. *Classification and Regression Trees*. Belmont, Calif.: Wadsworth.
- Breiman, L. 1996. Bagging predictors. *Machine learning* 24(2):123–140.
- Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.
- Carreira-Perpiñán, M. Á., and Tavallali, P. 2018. Alternating optimization of decision trees, with application to learning sparse oblique trees. In *Advances in Neural Information Processing Systems*, 1219–1229.
- Chang, C.-C., and Lin, C.-J. 2011. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2(3):27.
- Criminisi, A., and Shotton, J. 2013. *Decision Forests for Computer Vision and Medical Image Analysis*. Advances in Computer Vision and Pattern Recognition. Springer-Verlag.
- De’Ath, G. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. *Ecology* 83(4):1105–1117.
- Dheeru, D., and Karra Taniskidou, E. 2017. UCI machine learning repository.
- Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; and Lin, C.-J. 2008. Liblinear: A library for large linear classification. *Journal of machine learning research* 9(Aug):1871–1874.
- Freund, Y., and Schapire, R. 1999. A fast algorithm for introducing boosting. *Journal-Japanese Society For Artificial Intelligence* 14(77):780–1612.
- Fromont, N., and Hinton, G. 2017. Distilling a neural network into a soft decision tree. *arXiv preprint arXiv:1711.07844*.
- Garofalakis, M.; Hyun, D.; Rastogi, R.; and Shim, K. 2003. Building decision trees with constraints. *Data Mining and Knowledge Discovery* 7(2):187–214.
- Hastie, T.; Tibshirani, R.; and Friedman, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction*. New York, NY: Springer, second edition.
- Heath, D.; Kasif, S.; and Salzberg, S. 1993. Induction of oblique decision trees. In *IJCAI*, volume 1993, 1002–1007.
- Hsia, C.-Y.; Zhu, Y.; and Lin, C.-J. 2017. A study on trust region update rules in newton methods for large-scale linear classification. In *Asian Conference on Machine Learning*, 33–48.
- Huber, P. J., et al. 1964. Robust estimation of a location parameter. *The annals of mathematical statistics* 35(1):73–101.
- Huber, P. J. 2011. Robust statistics. In *International Encyclopedia of Statistical Science*. Springer. 1248–1251.
- Ikonomovska, E.; Gama, J.; and Džeroski, S. 2011. Incremental multi-target model trees for data streams. In *Proceedings of the 2011 ACM symposium on applied computing*, 988–993. ACM.
- Jordan, M. I., and Jacobs, R. A. 1994. Hierarchical mixtures of experts and the em algorithm. *Neural computation* 6(2):181–214.
- Kocev, D.; Džeroski, S.; White, M. D.; Newell, G. R.; and Griffioen, P. 2009. Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. *Ecological Modelling* 220(8):1159–1168.
- Kocev, D.; Vens, C.; Struyf, J.; and Džeroski, S. 2013. Tree ensembles for predicting structured outputs. *Pattern Recognition* 46(3):817–833.
- Laurent, H., and Rivest, R. L. 1976. Constructing optimal binary decision trees is np-complete. *Information processing letters* 5(1):15–17.
- Levatić, J.; Ceci, M.; Kocev, D.; and Džeroski, S. 2014. Semi-supervised learning for multi-target regression. In *International Workshop on New Frontiers in Mining Complex Patterns*, 3–18. Springer.
- Li, A. H., and Martin, A. 2017. Forest-type regression with general losses and robust forest. In *International Conference on Machine Learning*, 2091–2100.
- Loh, W.-Y., and Shih, Y.-S. 1997. Split selection methods for classification trees. *Statistica sinica* 815–840.
- Maibing, S. F., and Igel, C. 2015. Computational complexity of linear large margin classification with ramp loss. In *Artificial Intelligence and Statistics*, 259–267.
- Murthy, S. K.; Kasif, S.; and Salzberg, S. 1994. A system for induction of oblique decision trees. *Journal of artificial intelligence research* 2:1–32.
- Nijssen, S., and Fromont, E. 2007. Mining optimal decision trees from a dataset. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 530–539. ACM.
- Norouzi, M.; Collins, M.; Johnson, M. A.; Fleet, D. J.; and Kohli, P. 2015a. Efficient non-greedy optimization of decision trees. In *Advances in Neural Information Processing Systems*, 1729–1737.
- Norouzi, M.; Collins, M. D.; Fleet, D. J.; and Kohli, P. 2015b. Co2 forest: Improved random forest by continuous optimization of oblique splits. *arXiv preprint arXiv:1506.06155*.
- Quinlan, J. R. 2014. *C4. 5: programs for machine learning*. Elsevier.
- Struyf, J., and Džeroski, S. 2005. Constraint based induction of multi-objective regression trees. In *International Workshop on Knowledge Discovery in Inductive Databases*, 222–233. Springer.
- Viola, P., and Jones, M. J. 2004. Robust real-time face detection. *International journal of computer vision* 57(2):137–154.