Random Forests

Random Forest is a tree-based ensemble with each tree depending on a collection of random variables. Fits many classification trees to a data set, and then combines the predictions from all the trees. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them (1). Classification trees build the rule by recursive binary partitioning into regions that are increasingly homogeneous with respect to the class variable. The homogeneous regions are called nodes. At each step in fitting a classification tree, an optimization is carried out to select a node, a predictor variable, and a cut‐off or group of codes (for numeric and categorical variables respectively) that result in the most homogeneous subgroups for the data, as measured by the Gini index (2).

# Bibliografía

1. *RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY.* **Cutler, D. Richard y Thomas C. Edwards Jr., Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, Joshua J. Lawler.** 11, s.l. : Ecological Society of America, 2007, Vol. 88.

2. **Segal, Mark R.** *Machine Learning Benchmarks and Random Forest Regression.* s.l. : Open Access Publications from the University of California, 2004.

3. **Adele Cutler, D. Richard Cutler, John R. Stevens.** *Ensemble Machine Learning: Methods and Applications.* s.l. : Springer, 2011.