

Two-Eyed Algorithms and Problems

Leo Breiman

Department of Statistics, University of California, Berkeley
`leo@stat.Berkeley.Edu`

Two-eyed algorithms are complex prediction algorithms that give accurate predictions and also give important insights into the structure of the data the algorithm is processing. The main example I discuss is RF/tools, a collection of algorithms for classification, regression and multiple dependent outputs. The last algorithm is a preliminary version and further progress depends on solving some fascinating questions of the characterization of dependency between variables.

An important and intriguing aspect of the classification version of RF/tools is that it can be used to analyze unsupervised data—that is, data without class labels. This conversion leads to such by-products as clustering, outlier detection, and replacement of missing data for unsupervised data.

The talk will present numerous results on real data sets. The code (f77) and ample documentation for RFtools is available on the web site www.stat.berkeley.edu/RFtools.

References

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