Random Forest

WeizhouS

3/18/2019

Random Forest

How to build a random forest:

- Sample with replacement
- Choose $m << M$ features
- No pruning

The performance depends on:

- correlation
- strength

Around 1/3 left out for each tree ensures:

- unbiased error
- variable importance

Proximities:

- missing imputation. proximity weighted
- outlier. small proximity with ones in same class

Variable importance:

- permute features in oob sample. permutation importance measure
- gini importance

Interaction:

- a split on m systematically make split on k more/less possible

Mislabled case: using outlier measure

Balance prediction error by setting weights

Why sample with replacement? To reduce overfit. In the end, your data is only a sample.