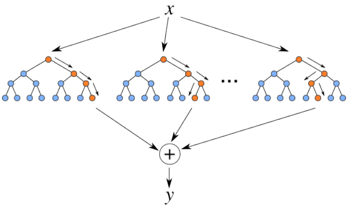
Introduction to Machine Learning

Random Forest: Introduction



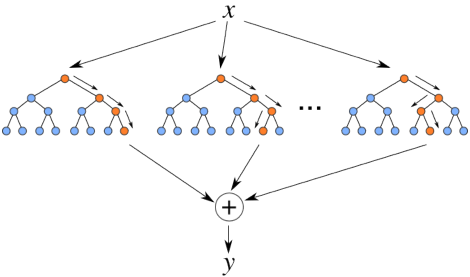
Learning goals

- Know how random forests are defined by extending the idea of bagging
- Understand that the goal is to decorrelate the trees
- Understand that the out-of-bag error is a way to obtain unbiased estimates of the generalization error during training

RANDOM FORESTS

Modification of bagging for trees proposed by Breiman (2001):

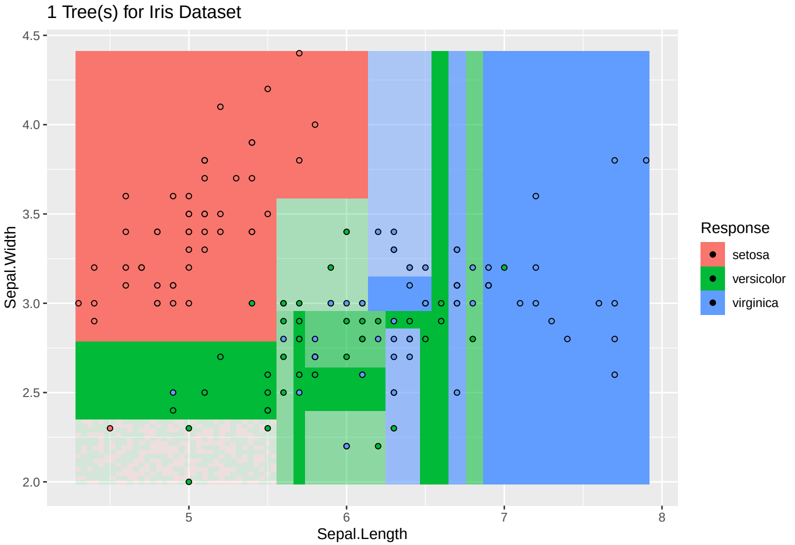
- Tree base learners on bootstrap samples of the data
- Uses **decorrelated** trees by randomizing splits (see below)
- Tree base learners are usually fully expanded, without aggressive early stopping or pruning, to **increase variance of the ensemble**



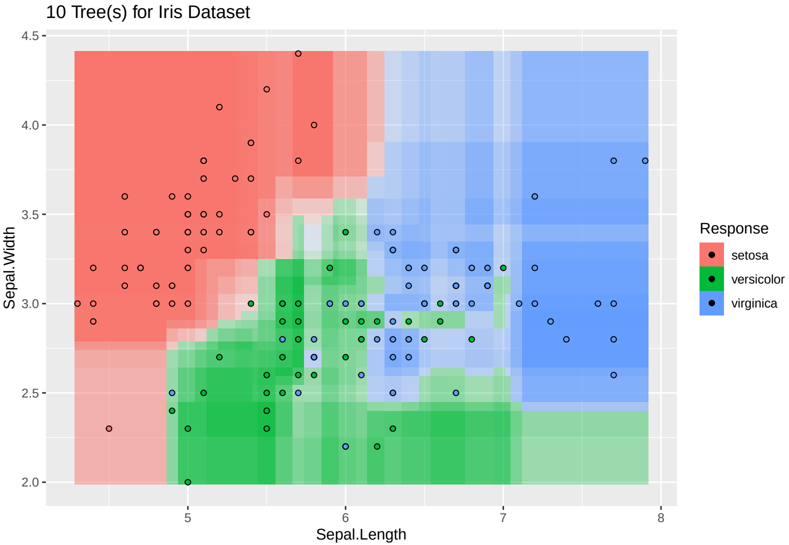
RANDOM FEATURE SAMPLING

- From our analysis of bagging risk we can see that decorrelating trees improves the ensemble
- Simple randomized approach:
  At each node of each tree, randomly draw mtry $\leq p$ candidate features to consider for splitting. Recommended values:
  - Classification: mtry $= \lfloor \sqrt{p} \rfloor$
  - Regression: mtry $= \lfloor p/3 \rfloor$

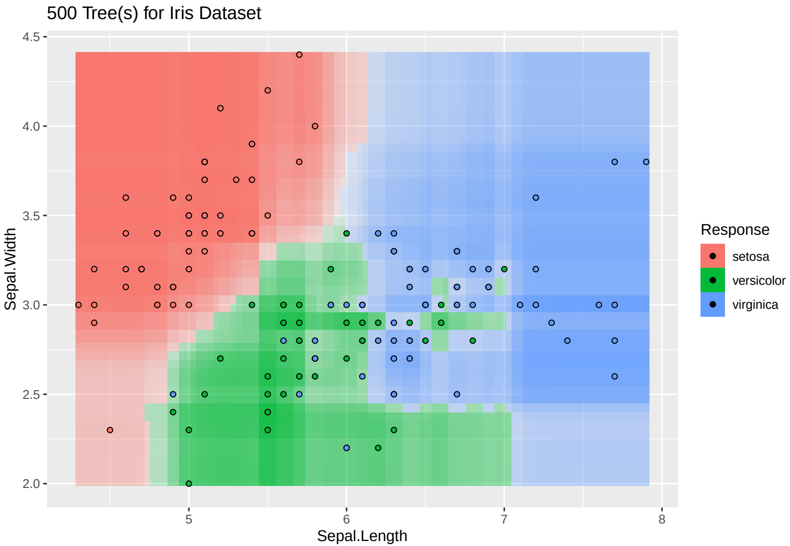
EFFECT OF ENSEMBLE SIZE



EFFECT OF ENSEMBLE SIZE

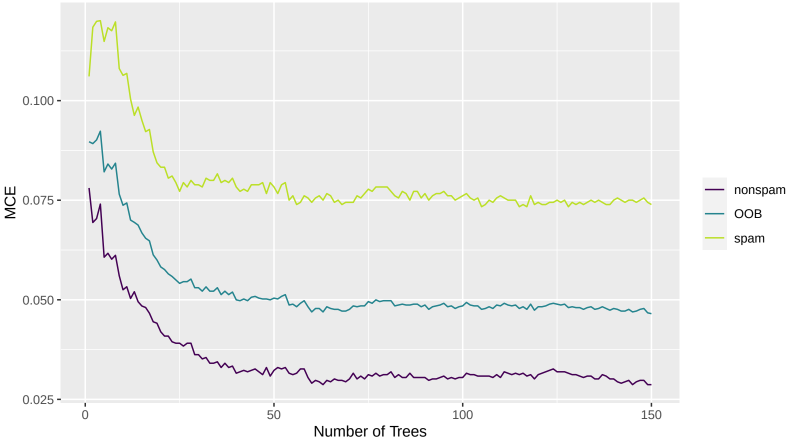


EFFECT OF ENSEMBLE SIZE

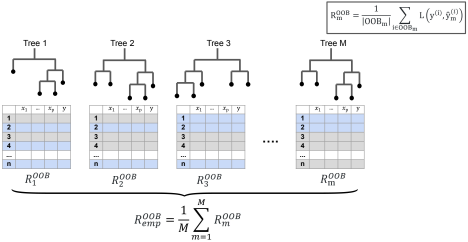


OUT-OF-BAG ERROR ESTIMATE

With the RF it is possible to obtain unbiased estimates of the generalization error directly during training, based on the out-of-bag observations for each tree:



OUT-OF-BAG ERROR ESTIMATE



$$\mathrm{R}_{\mathrm{m}}^{\mathrm{OOB}} = \frac{1}{|\mathrm{OOB}_{\mathrm{m}}|} \sum_{\mathrm{i} \in \mathrm{OOB}_{\mathrm{m}}} \mathrm{L}\left(\mathrm{y}^{(\mathrm{i})}, \hat{\mathrm{y}}_{\mathrm{m}}^{(i)}\right)$$

$$R_{emp}^{OOB} = \frac{1}{M} \sum_{m=1}^{M} R_m^{OOB}$$

▭ In-bag observations, used to build the trees {Remember: the same observation can enter the in-bag sample more than once}

▨ out-of-bag observations(OOB_m), used to evaluate prediction performance (R_{m}^{OOB})

- OOB size: $P(\text{not drawn}) = \left(1 - \frac{1}{n}\right)^n \overset{n \to \infty}{\longrightarrow} \frac{1}{e} \approx 0.37$
- Predict all observations with trees that didn't use them for training and compute average loss of these predictions
- Similar to 3-CV, can be used for a quick model selection