Workshop: Regionalization of Forest Stand Variables
-TCP /IND/3505 -

Lab 04 Model building

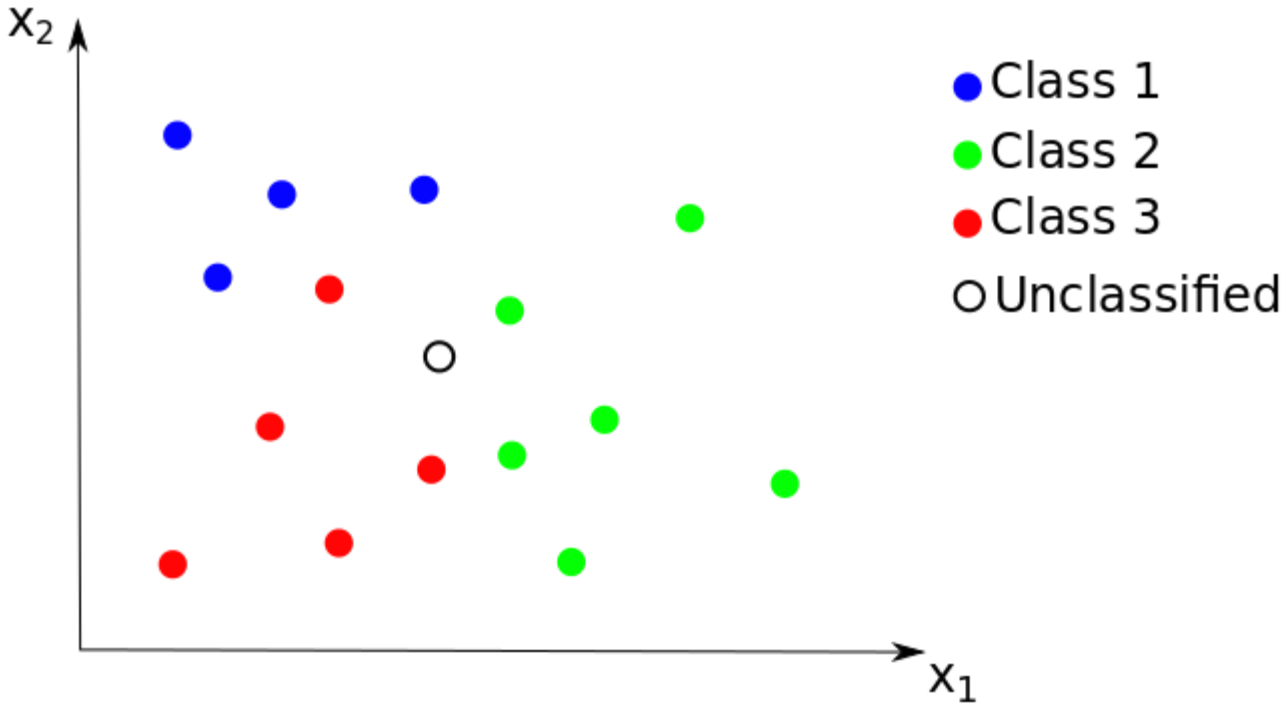
Paul Magdon

Forest Survey India, August,2017, Dehradun, India

Goals

- Select the most important image features / bands
- Create a non-parametric RandomForest Model

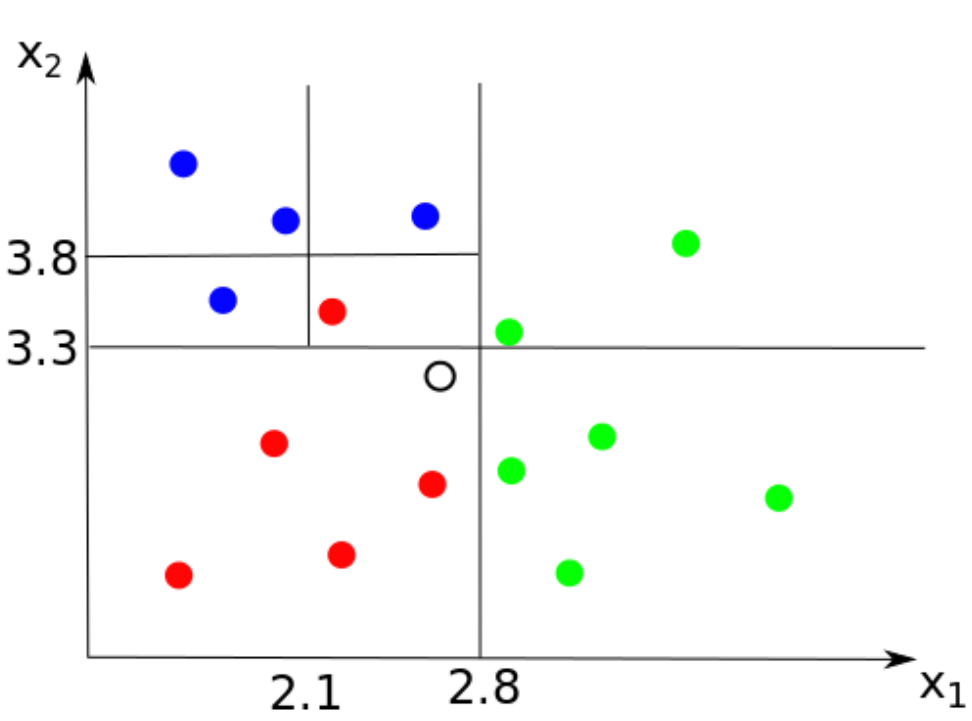
Classification

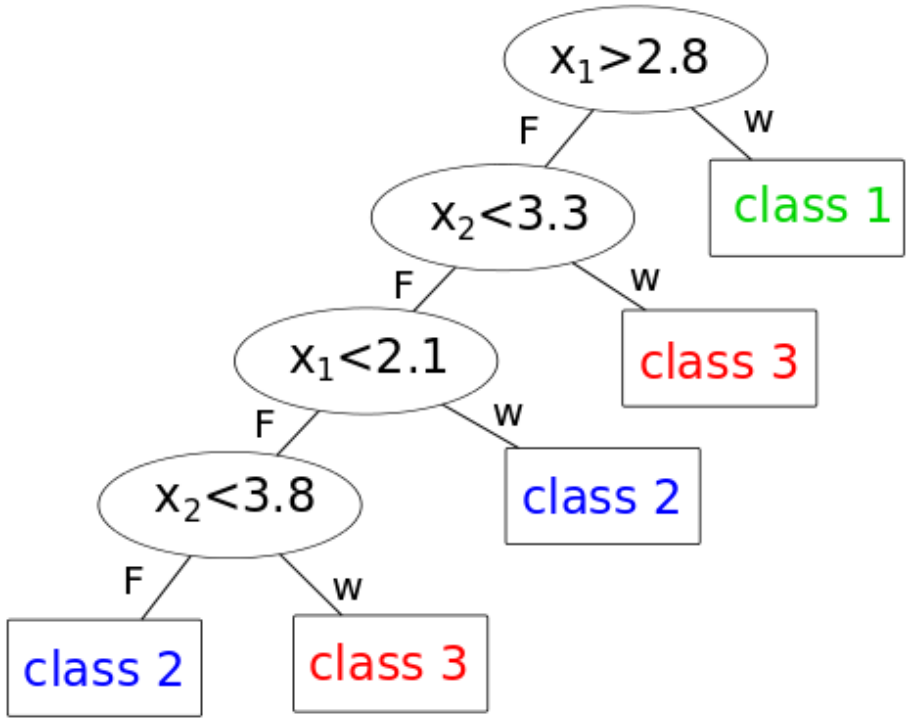


Training data or Reference data:
Data set, where both independent variables *X* (predictors) and dependent variables *Y* (response) are known.

Test data or Target data:
Data set where only the independent *X* variables are known.

Classification



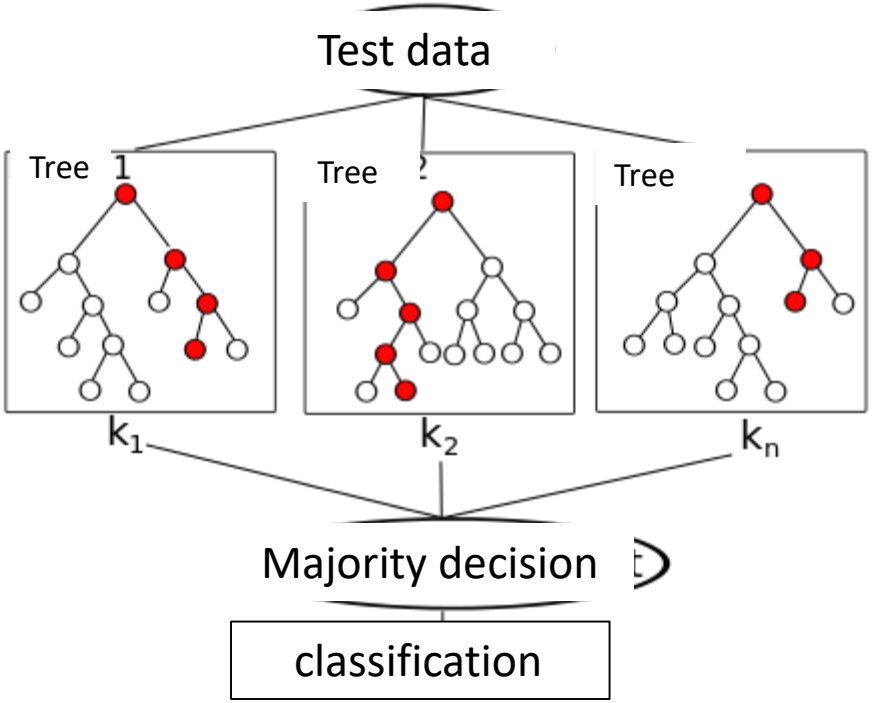


- Decision trees divide the *n*-dimensional space by thresholds.
- Binary decision rules (decision tree).

RandomForests

- Drawback of the the CART method: small changes of the training data set induce very different results.
- Therefore Breimann et al. 2001 developed the **Random Forests classifier.**
- Combining several decision trees (ensemble classifier), where the class is used for classification which was most often used for prediction (majority decision).
- During generation of the single decision tree only $n < N$ objects of the training data are randomly selected. In addition on each decison node only $m < M$ predictors are used to ensure that the decision trees are different.

Random Forests



Method:

Combination of several decison trees

Parameter:

- *n*= number of decision trees
- M= number of predictors
- m= number of predictors which are used at each node

Pros:

- Low classification error
- Transparent model is build which can be used for new targets (White Box)
- Importance of predictors can be estimated

Cons:

- High computational effort