BREIMAN RANDOM FORESTS

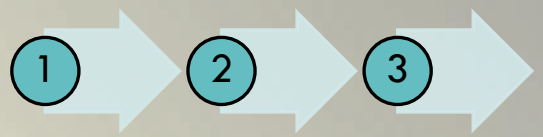
MEGHANA KILLI, NAM TRAN

Introduction

Machine Learning

Deterministic

Eg: AdaBoost

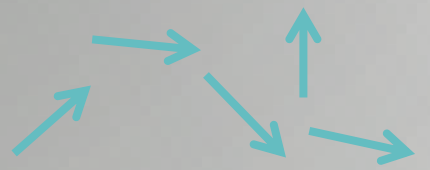


Strong Law of Large Numbers

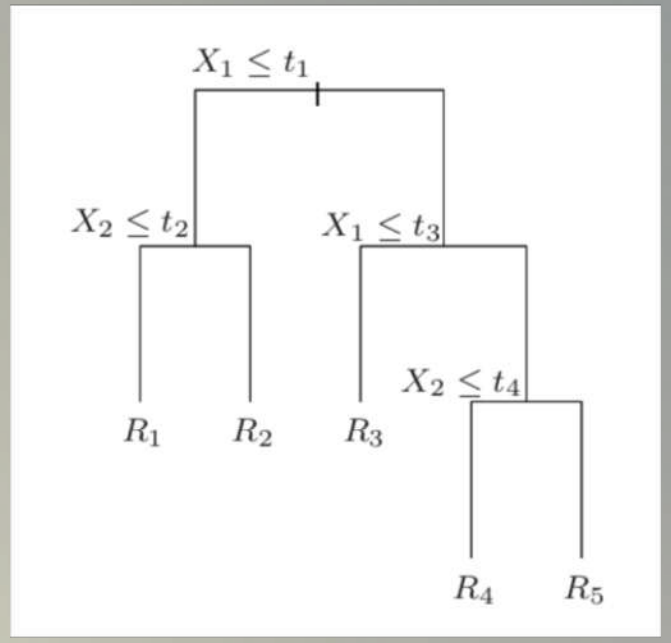
~

Random

Eg: Random Forests



Decision Trees



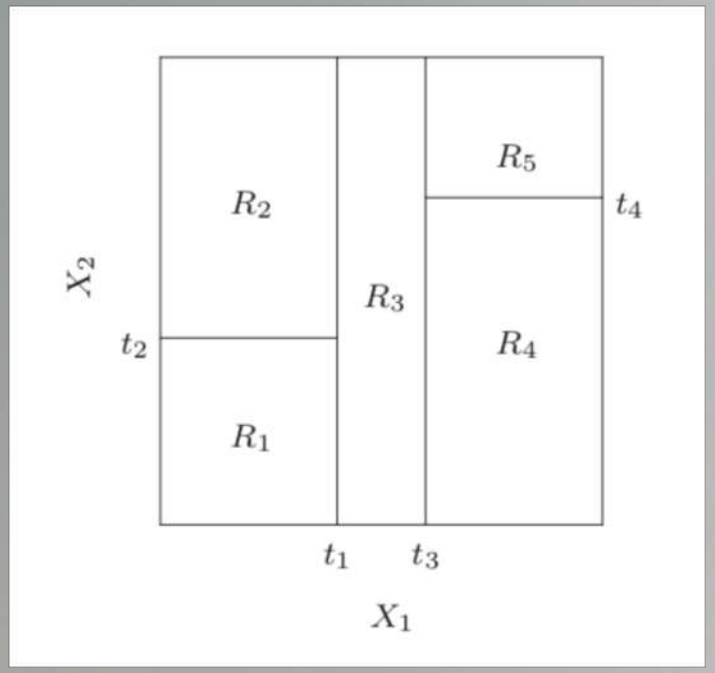


Image source: Elements of Statistical Learning by Hastie et.al.

Random Forests

Bootstrapping - Random Sampling Technique with Replacement

{1 2 3 4 5 6}

{1 1 2 2 3 3}, {1 1 1 2 2 2}, {1 2 3 4 5 5}, {5 5 6 6 6 6}

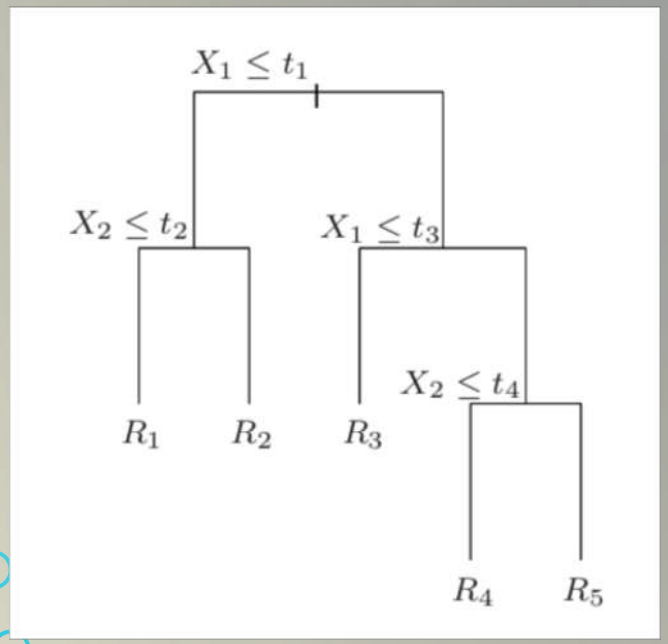
Bagging – (B)ootstrap (Agg)regat(ing)

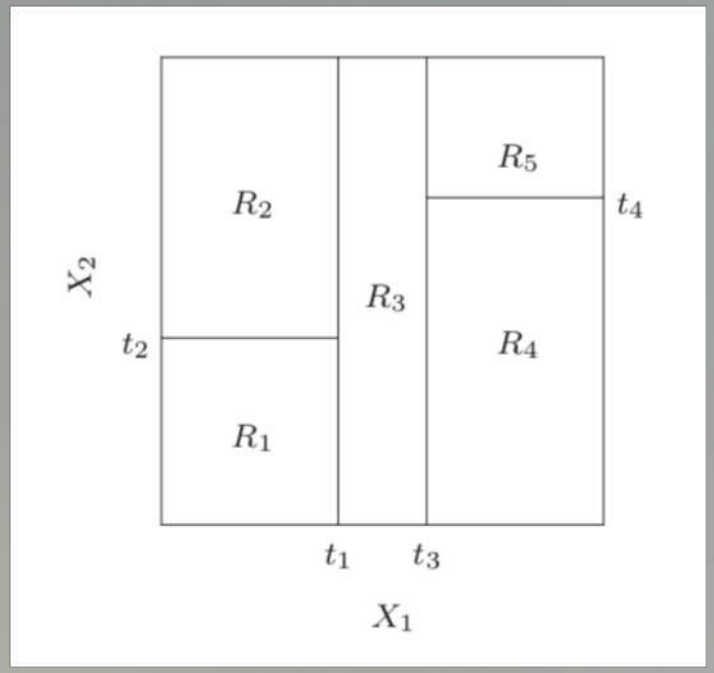
Decision Trees  Random Forests

Random Subspace Method

- Total number of features = N
- Generate a subspace at random with m (<N) features
- Select features from subspace to best split data

Decision Trees in a Subspace





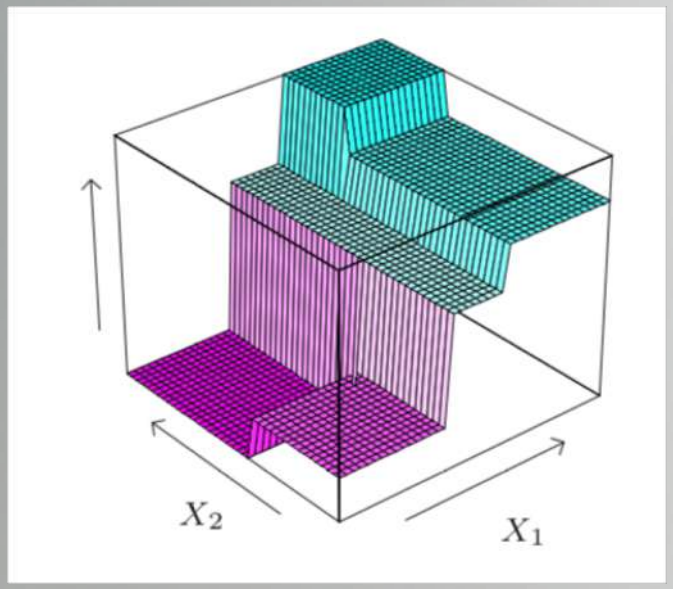
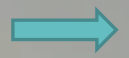
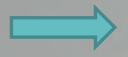


Image source: Elements of Statistical Learning by Hastie et.al.

Breiman Random Forests

Bootstrapping - Random Sampling Technique with Replacement

Bagging – (B)ootstrap (Agg)regat(ing)

Random Subspace Method  Decision Trees  Breiman Random Forests

Features of BRFs

- Classification and Regression
- Avoid overfitting
- Insensitive to number of features sampled
- Robust to outliers and noise - more than Adaboost
- Faster than bagging and boosting

THANK YOU!