Machine Learning

Abdelhak Mahmoudi
abdelhak.mahmoudi@um5.ac.ma

INPT- 2020

Content

1. The Big Picture
2. Supervised Learning
   - Linear Regression, Logistic Regression, Support Vector Machines, Trees, Random Forests, Boosting, Artificial Neural Networks
3. Unsupervised Learning
   - Principal Component Analysis, K-means, Mean Shift

Supervised Learning

- Linear Regression
- Logistic Regression
- Support Vector Machines
- Trees (Decision and Regression)
- **Random Forests**
- Boosting
- Artificial Neural Networks

Random Forest

Bagging

Bootstrap Aggregating

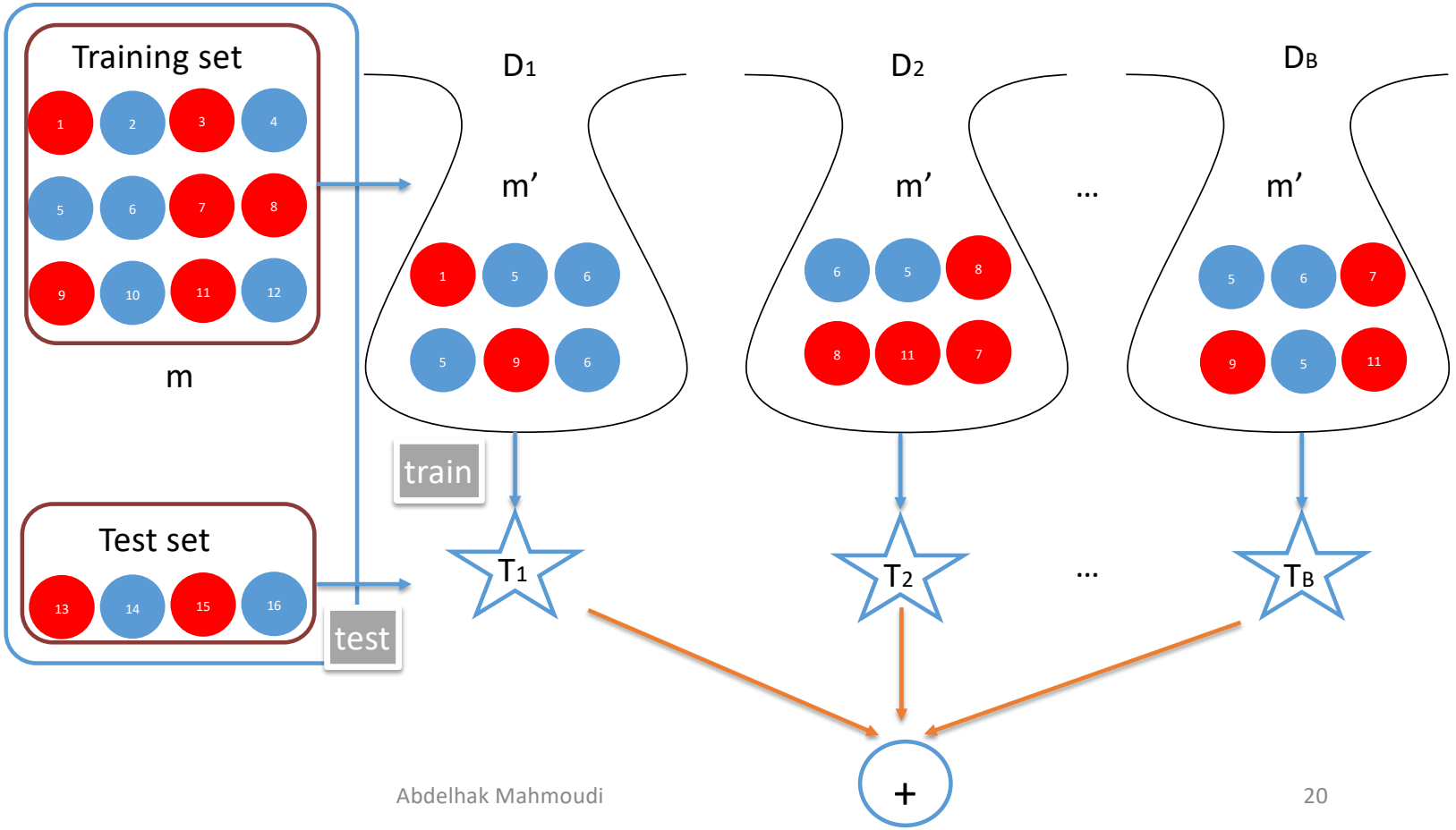
Training

Pick m' examples with replacement and train B trees

Testing

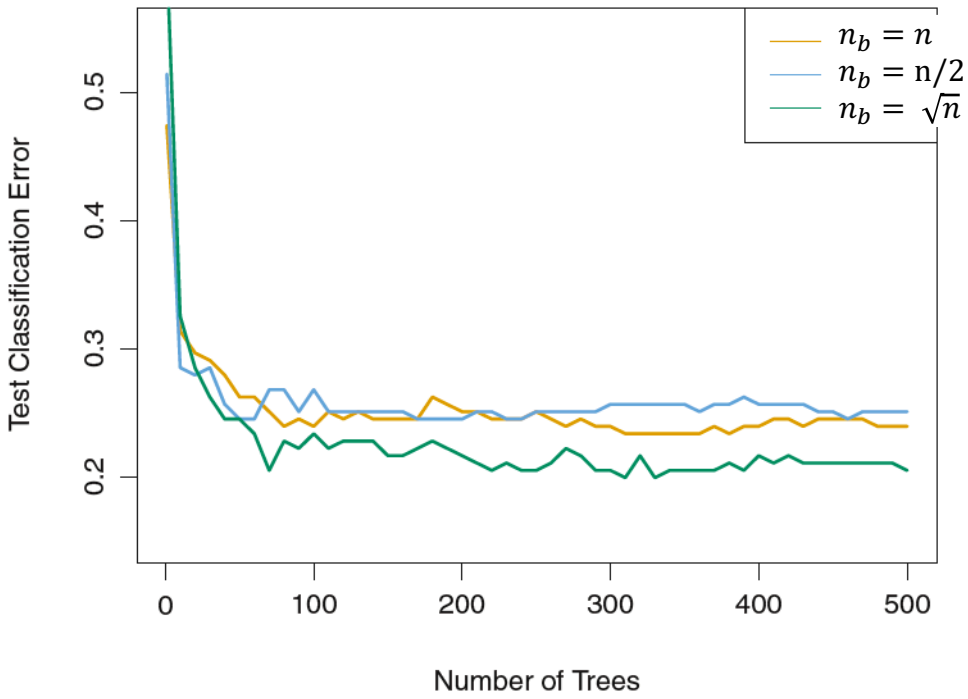
Regression: mean errors of all the B trees

Classification: vote



Random Forest

- **Problem:** Bagged trees will look quite similar to each other, so averaging them will not led to much reduction of variance!
- **Solution:** Random Forest constructs multiple trees where each tree uses n_b random features from the n initial features (generally $n_b = \sqrt{n}$)
- $n_b = n$-> Bagging case



Random Forest

- Both training and prediction are very fast, because of the simplicity of the underlying decision trees.
- Tasks can be straightforwardly parallelized, because the individual trees are entirely independent entities.
- The multiple trees allow for a probabilistic classification: a majority vote among estimators gives an estimate of the probability
- RF is a Nonparametric model, extremely flexible, and can thus perform well on tasks that are under-fit by other models.

Random Forest

- Hyper-Parameters Tuning
  - d: Depth of the trees
  - B: number of Bags