CREWES Data Science

Lecture 10: Tree Based Models for Regression

Marcelo Guarido

www.crewes.org

CREWES DSI

Today's Agenda

Part 1:	Decision Tree
Part 2:	Ensemble Models
Part 3:	Random Forest
Part 4:	Gradient Boosting
Part 5:	Live Coding

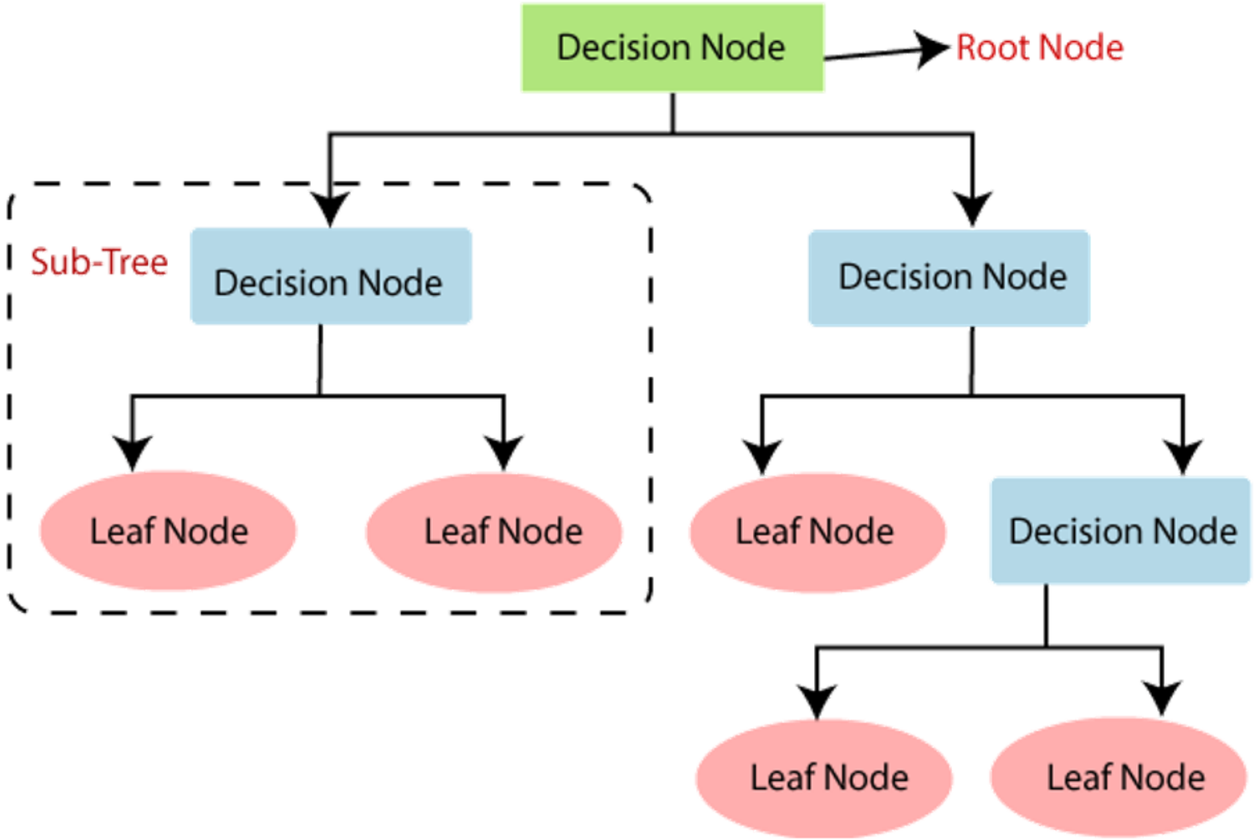
CREWES DSI

Part 1:

DECISION TREE

Decision Tree

- Used for regression or classification
- Simple algorithm
- Basically, a sequence of “if” conditions
- Select the branch separation that minimizes the error
- Can handle continuous and categorical features
- Usually, considered a weak model

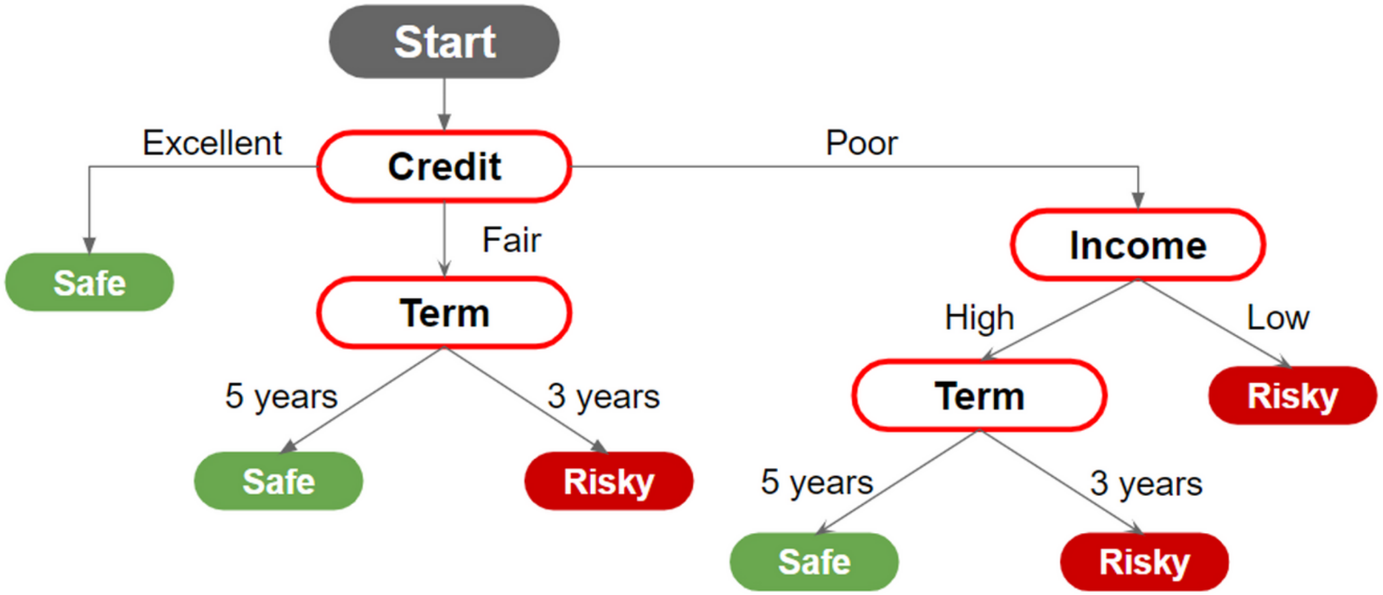


Decision Tree

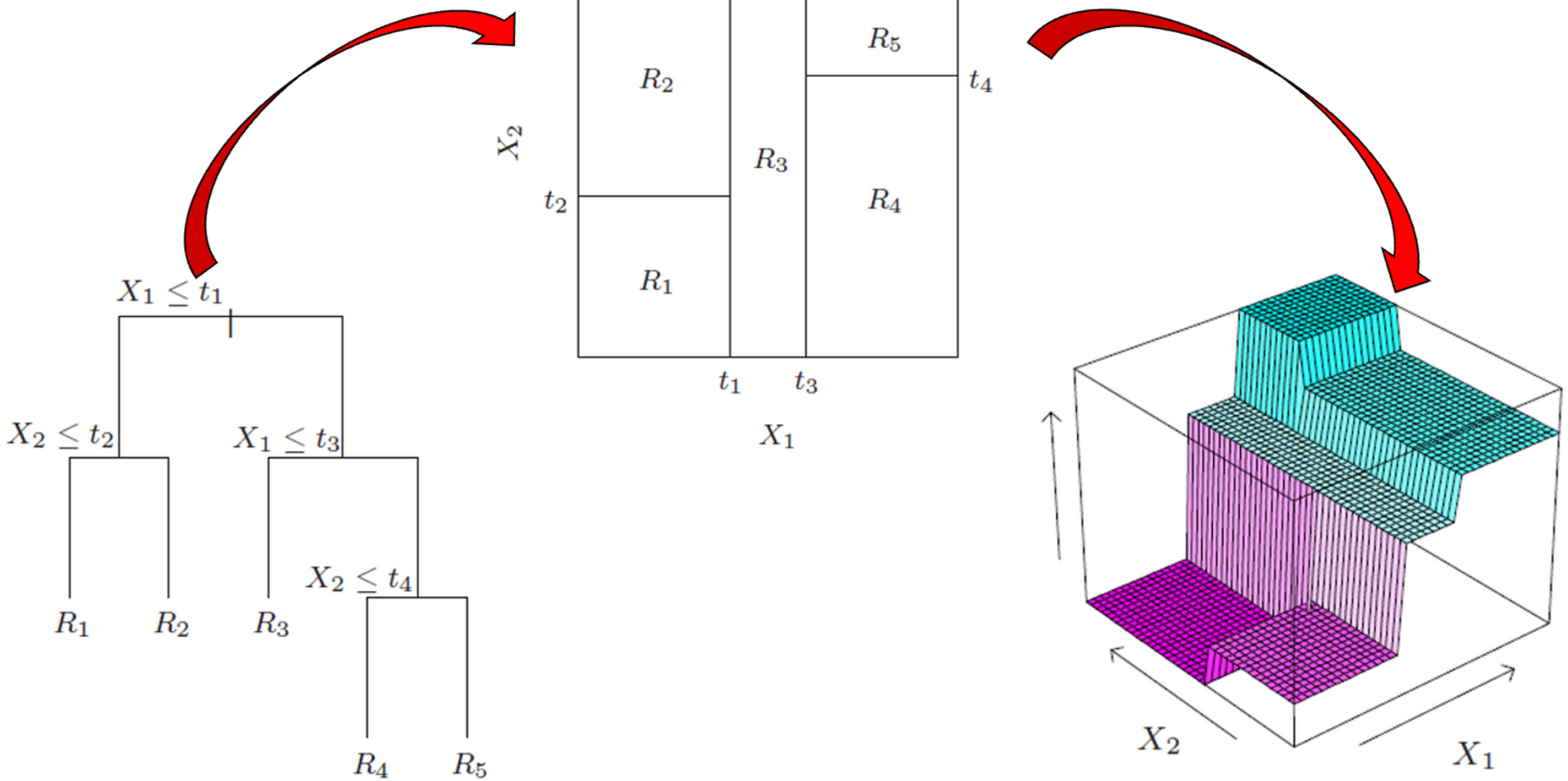
Credit
- Excellent
- Fair
- Poor

Income
- High
- Low

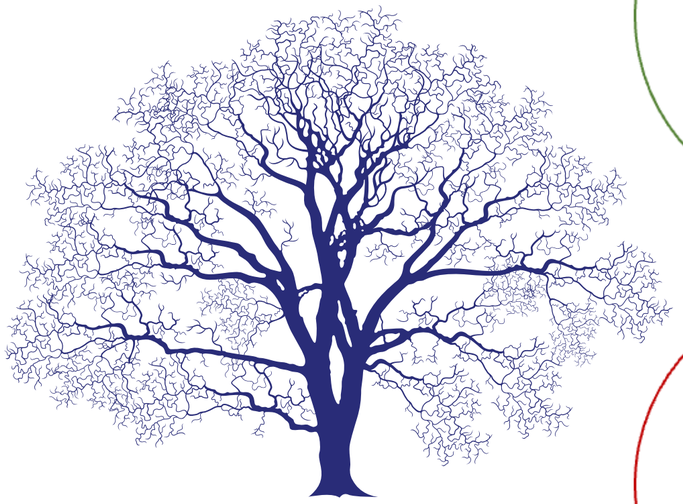
Term
- 3 years
- 5 years



Decision Tree



Decision Tree



Pros

- Requires less data preparation
- Numerical and categorical data
- No scaling or normalization
- Not affected by missing data
- Easy and intuitive to interpret and explain

Cons

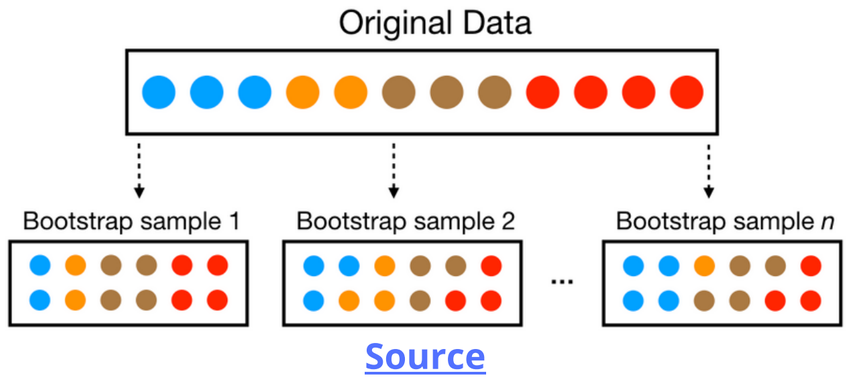
- Unstable
- Overfitting
- Tricky to apply for regression
- Relatively slow
- Limited on complex data

Part 2:

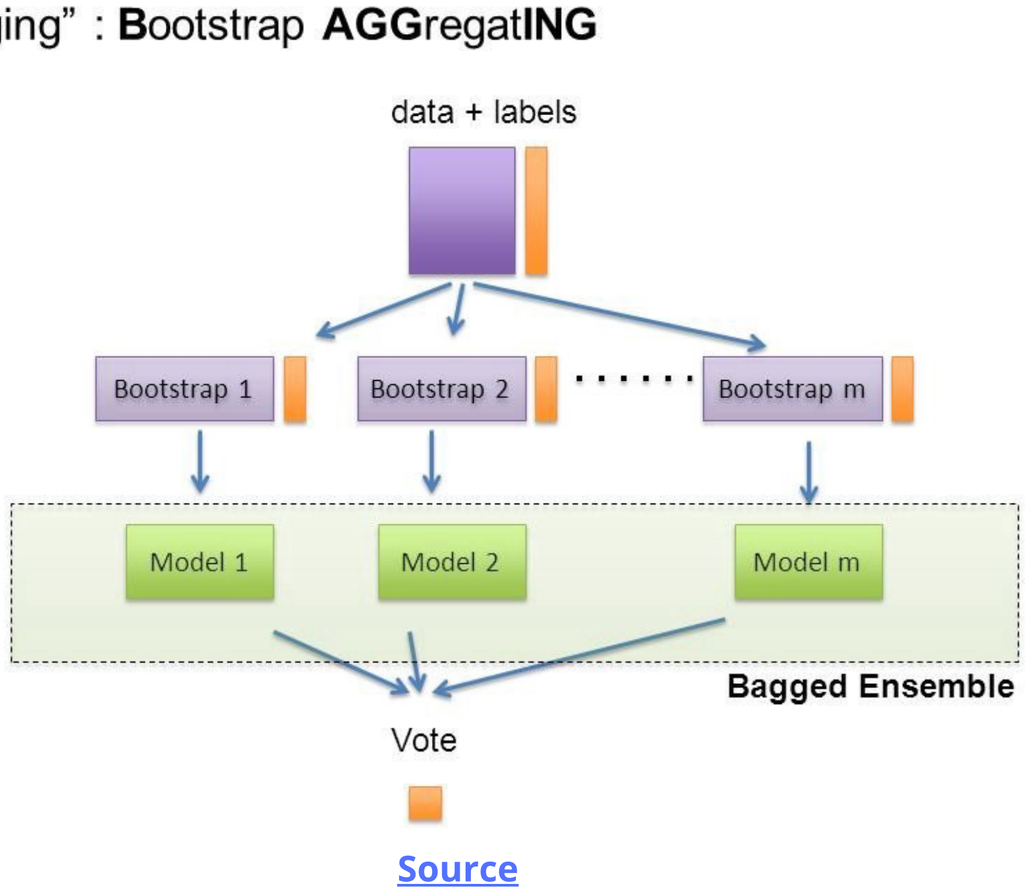
ENSEMBLE MODELS

Ensemble Models

- Combination of "weak" models
- Divide the train data into different subsets (Bootstrap)
- Train on different subsets of the training data (Bagging)
- Vote system (hard of soft)
- Regression instead of vote



Source



Source

CREWES DSI

Part 3:

RANDOM FOREST

Random Forest

- Ensemble of decision trees
- Each tree is trained over a subset of the features randomly selected
- Can also train over a random selected subsample of the data
- Vote system (classification)
- Average predictions (regression)

Random Forest

Bootstrap sampling

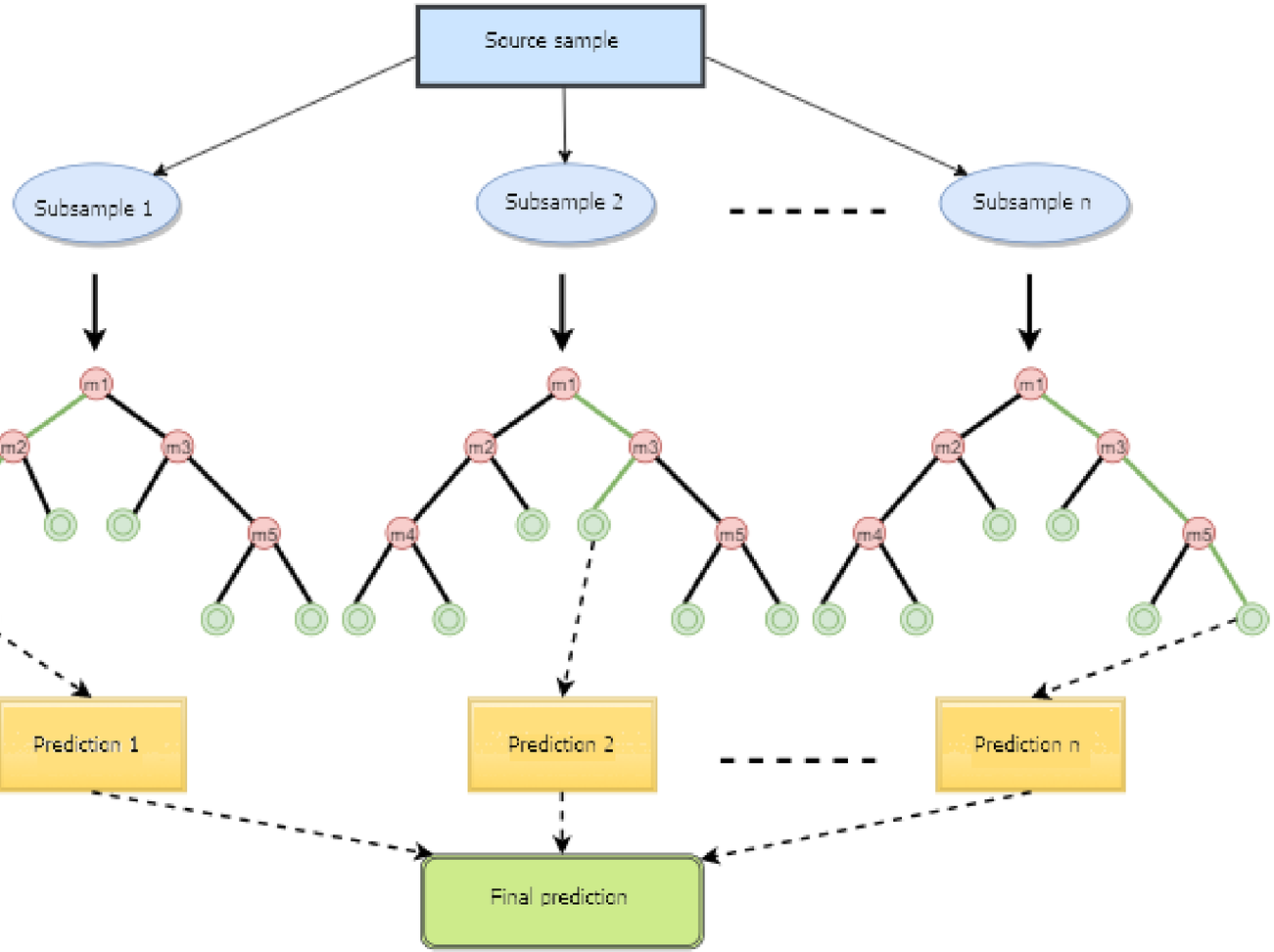
r (percentage) examples are selected (0.63 in classical implementation) in n random subsamples

Building the models

for each subsample, a decision tree is constructed based on a random set of m features (covariants), the results fall into leaves

Bootstrap aggregating

results from all constructed trees are gathered and averaged



Source

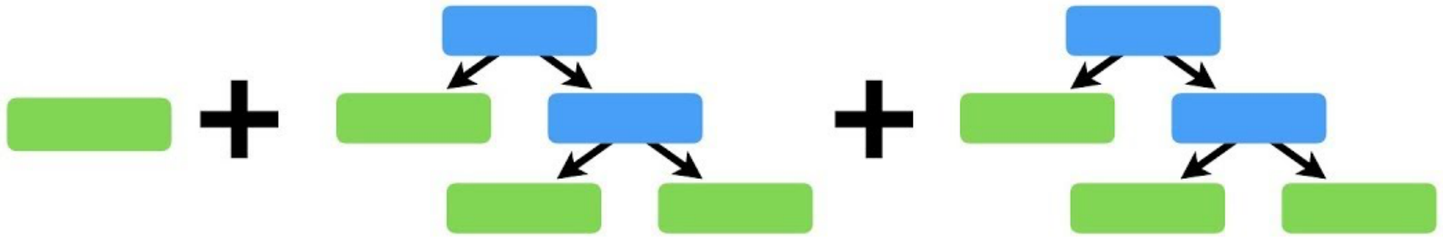
Random Forest

Pros

- Requires less data preparation
- Numerical and categorical data
- Handle outliers
- Less overfitting
- Still relatively easy to interpret

Cons

- Biased with categorical features
- Bad with sparse features
- Slow training



CREWES DSI

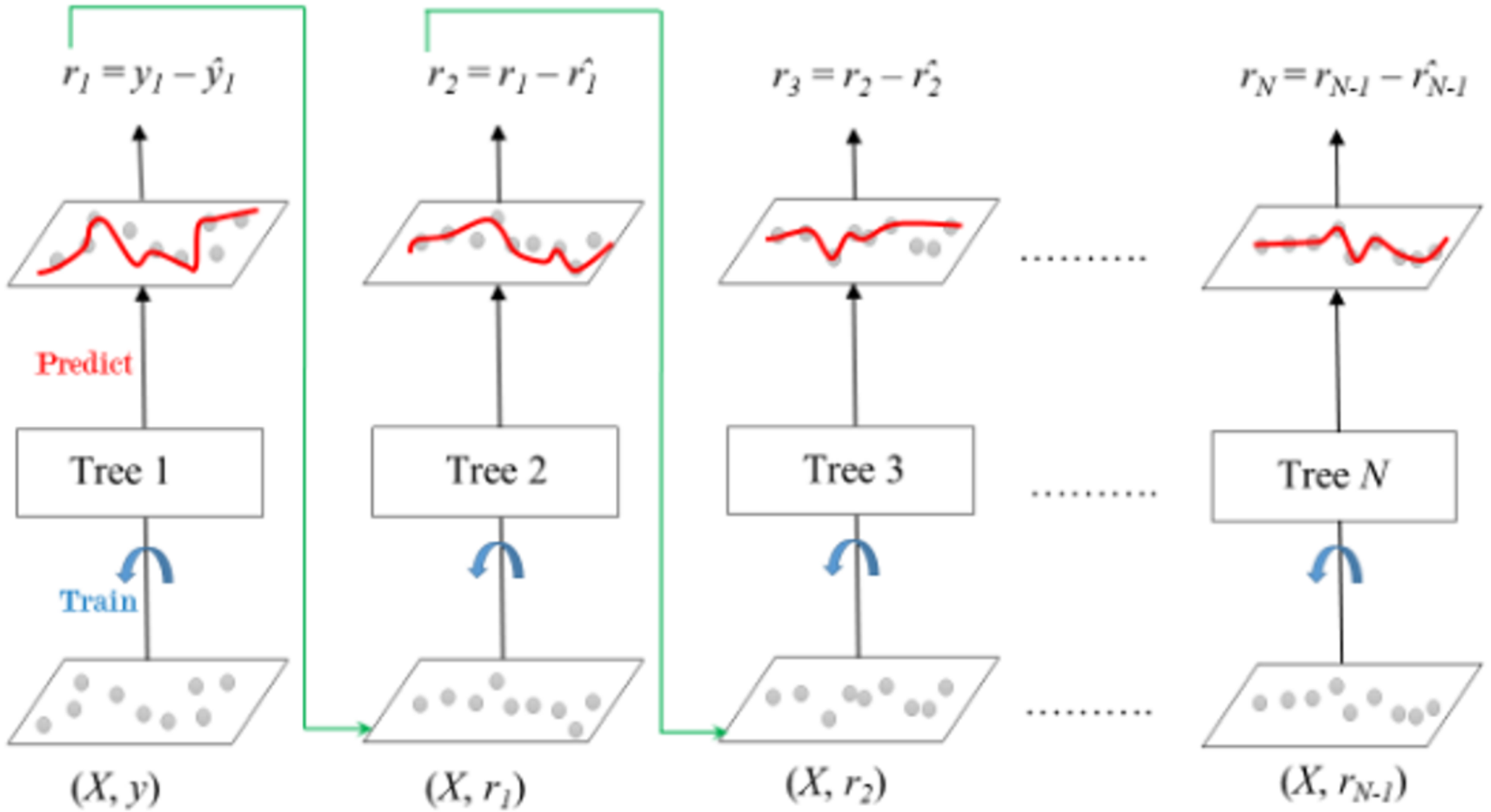
Part 4:

GRADIENT BOOSTING

Gradient Boosting

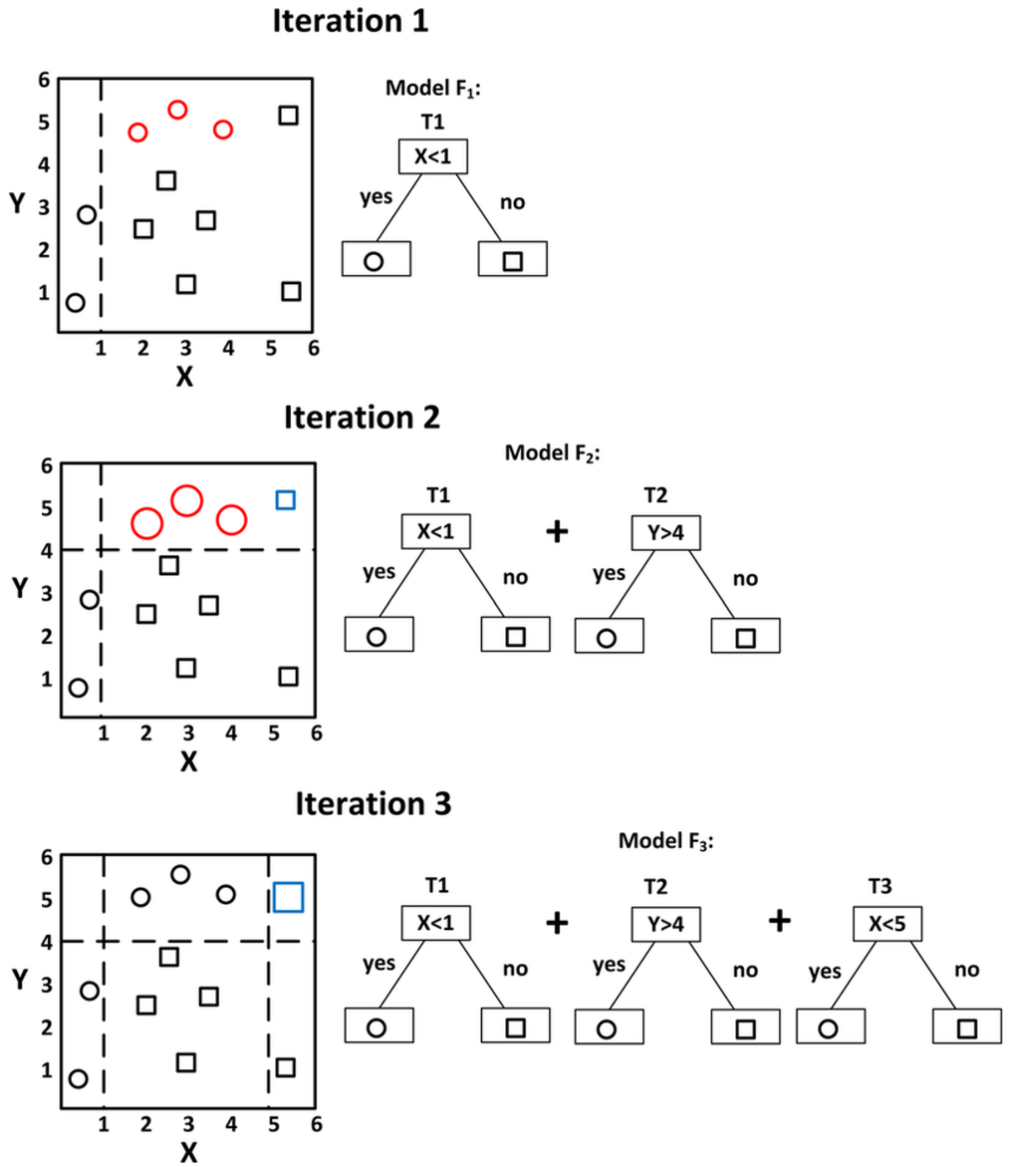
- Ensemble of decision trees
- Sequential trees
- Trained over residuals of previous tree
- Gradient descent

Gradient Boosting



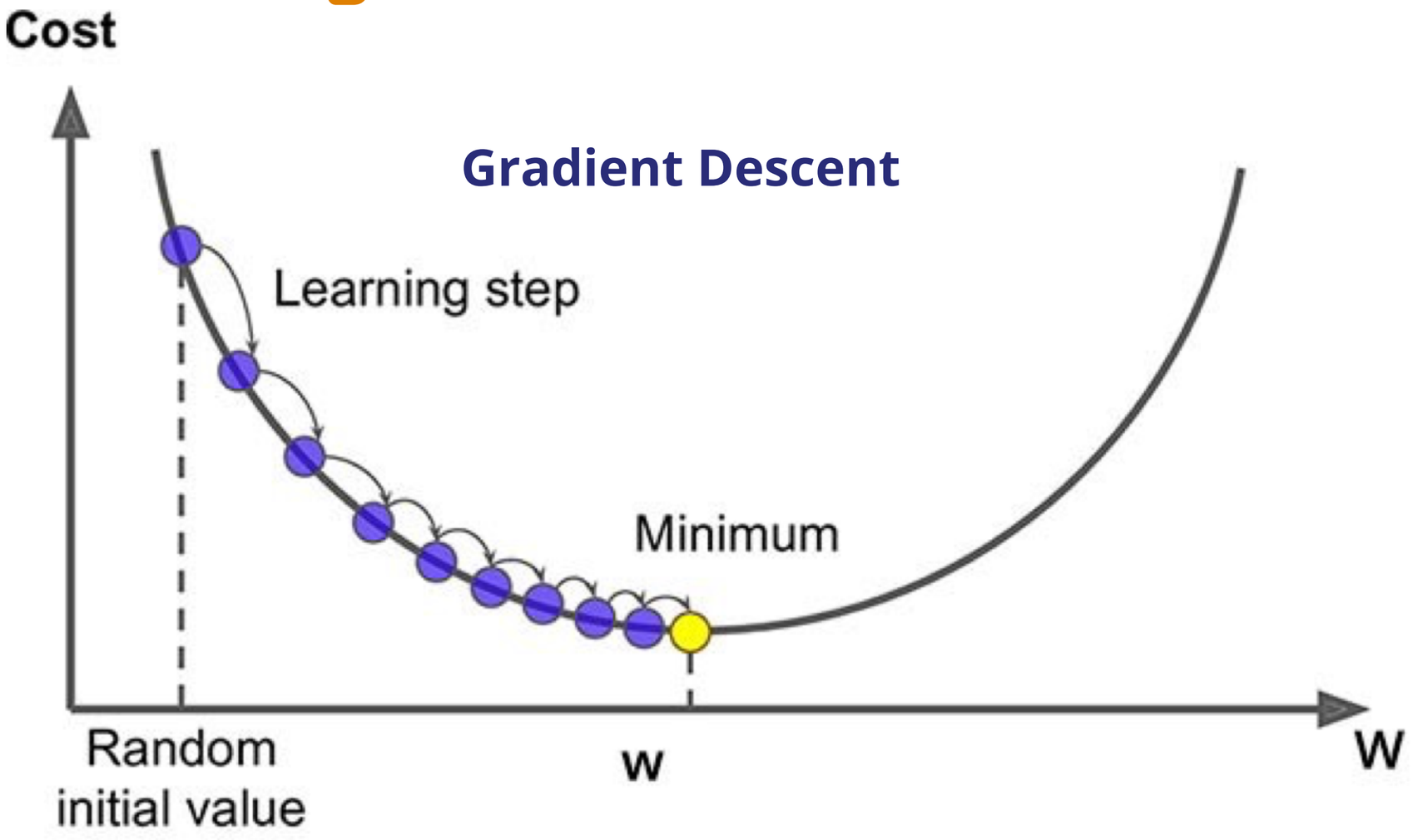
Source

Gradient Boosting



Source

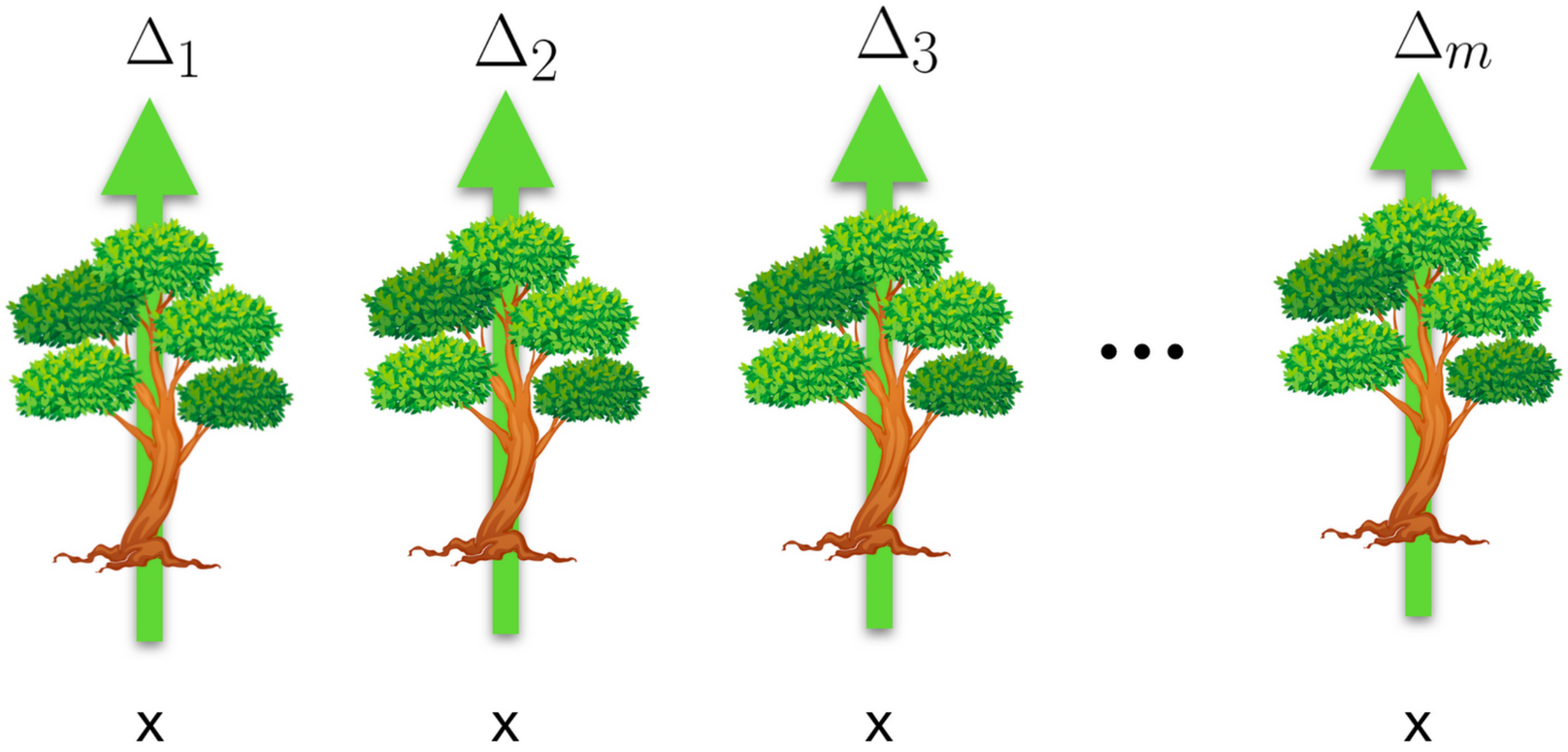
Gradient Boosting



Source

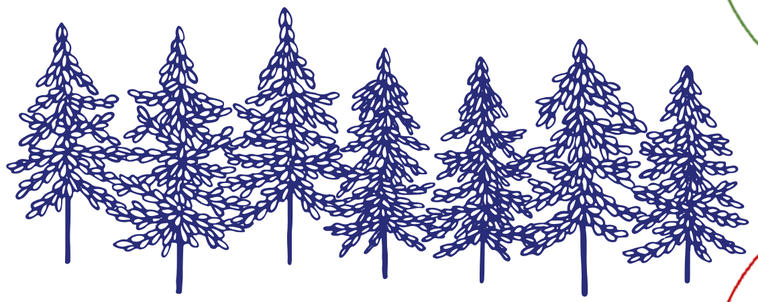
Gradient Boosting

$$\hat{y}_1 = \Delta_1 \qquad \hat{y}_2 = \hat{y}_1 + \eta\Delta_2 \qquad \hat{y}_3 = \hat{y}_2 + \eta\Delta_3 \qquad \hat{y}_m = \hat{y}_{m-1} + \eta\Delta_m$$



Source

Gradient Boosting



Pros

- Requires less data preparation
- Numerical and categorical data
- No scaling or normalization
- Less overfitting
- Still relatively easy to interpret

Cons

- Sensitive to outliers
- Hard for parallel processes
- Can get quite slow

CREWES DSI

Part 5:

LIVE CODING (PYTHON)