

Tree - Based Model

Ideas: Split feature space into simpler regions

- Use a sequence of rules
- Sequence of rules can be visualized as tree

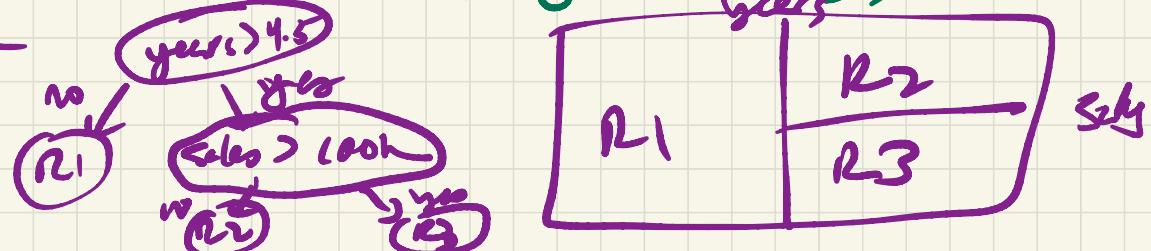
+ easy to interpret

x very high variance (sensitive to individual points in training set)

x not that accurate.

(solution: bagging, boosting)

Example: Regression tree



Regression Tree \rightarrow Decision Tree for continuous values

- Predict continuous values
- Split node: Each split minimizes mean squared error.

Greedy Approach

$$L_1(j, s) = \{x \mid x_j < s\}$$

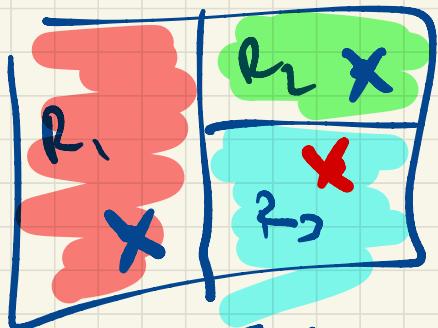
$$R_2(j, s) = \{x \mid x_j \geq s\}$$

Minimize w.r.t. j, s

$$\sum_{i \in R_1} (y_i - \hat{y}_{R_1})^2 - \sum_{i \in R_2} (y_i - \hat{y}_{R_2})^2$$

Classification Tree → Decision Tree for classes

- Predict categorical values
- Split criterion: Minimize classification error rate



→ fraction of training observations that
don't belong to most common class in
a region.

Gini Index

$$G = \sum_{k=1}^n \hat{p}_{mk} (1 - \hat{p}_{mk})$$

Entropy

$$D = - \sum_{k=1}^n \hat{p}_{mk} \log \hat{p}_{mk}$$

Ensemble Methods

- Bagging (bootstrap aggregation) [Random Forest]
- Boosting [Adaboost, Gradient Boosted trees]

① Bagging (Random Forest)

→ N separate models w/ repeated sample from the data
(with replacement)

→ Average predictions of models

• Regression: Average

• Classification: Vote



Boosting

- Sequentially add predictions ("weak" predictors)
- Example: Gradient Boosting (GBRT)

