Classification and Regression Trees

Mihaela van der Schaar

Department of Engineering Science
University of Oxford

March 1, 2017

Overview - Regression Tree

- **Regression Tree:**
  - Given the dataset $\mathcal{D} = (x_1, y_1), ..., (x_n, y_n)$ where $x_i \in \mathbb{R}, y_i \in \mathcal{Y} = \mathbb{R}$
  - minimize the error (e.g. square loss error) in each leaf
  - the parameterized function is

$$\hat{f}(x) = \sum_{j=1}^{K} \beta_j \cdot \mathbb{I}(x \in \mathcal{R}_j)$$

  - Using squared loss, optimal parameters are:

$$\hat{\beta}_j = \frac{\sum_i y_i \cdot \mathbb{I}(x_i \in \mathcal{R}_j)}{\sum_i \mathbb{I}(x_i \in \mathcal{R}_j)}$$

    which is sample mean.

Assign the prediction for each leaf

For each leaf, we need to assign the prediction y which minimizes the loss for the regression problem.

- **Regression Tree:**

$$\hat{y} = \arg\min_{y \in \mathbb{R}} \sum_{i \in \mathcal{R}_k} (y - y_i)^2 = \frac{1}{\sum \mathbb{I}(x_i \in \mathcal{R}_k)} \cdot \sum \mathbb{I}(x_i \in \mathcal{R}_k) \cdot y_i$$

Growing the Tree: Overview

- Ideally, would like to find partition that achieves minimal risk: lowest mean-squared error for regression problem.
- Number of potential partitions is too large to search exhaustively.
- **Greedy** search heuristics for a good partition:
  - Start at root.
  - Determine the best feature and value to split.
  - Recurse on children of node.
  - Stop at some point (with heuristic pruning rules).

Algorithm for Regression Trees

- Start with $\mathcal{R}_1 = \mathbb{R}^d$
- For each feature $j = 1, ..., d$, for each value $v \in \mathbb{R}$ that we can split on:
  - Split the data set:

$$I_< = \{i : x_{ij} < v\} \text{ and } I_> = \{i : x_{ij} \geq v\}$$

  - Estimate parameters:

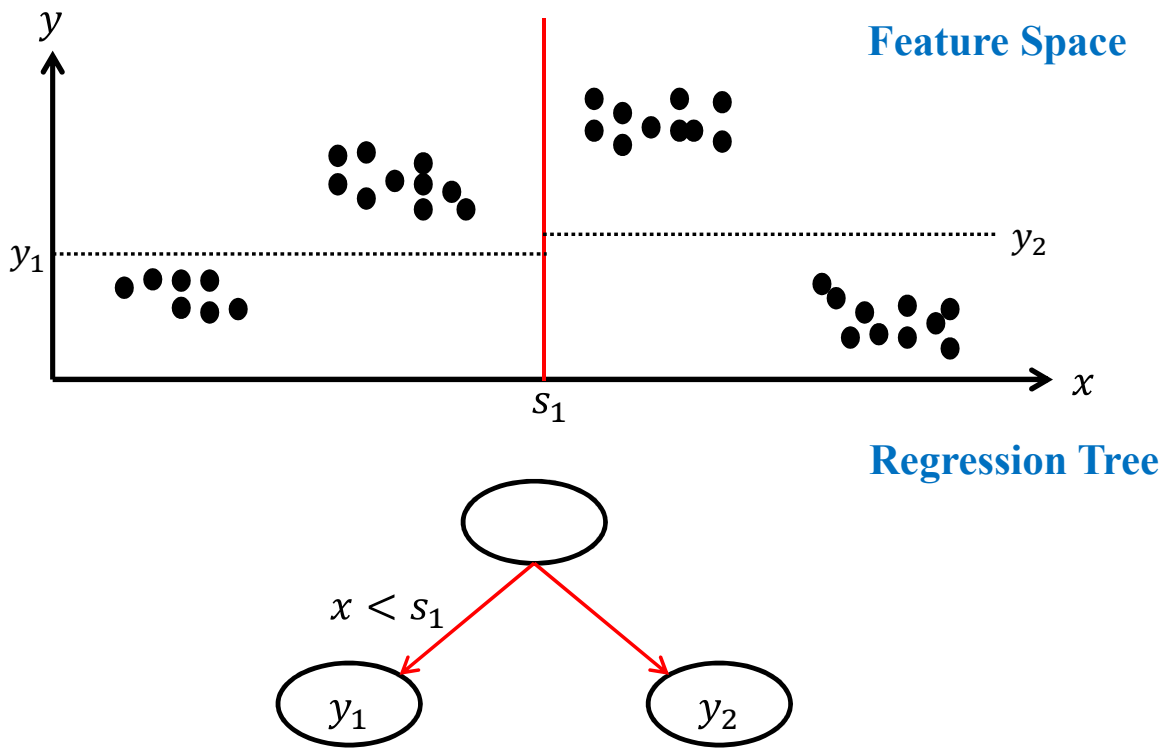
$$\beta_< = \frac{\sum_{i \in I_<} y_i}{|I_<|} \text{ and } \beta_> = \frac{\sum_{i \in I_>} y_i}{|I_>|}$$

  - Quality of split is measured by the squared loss:

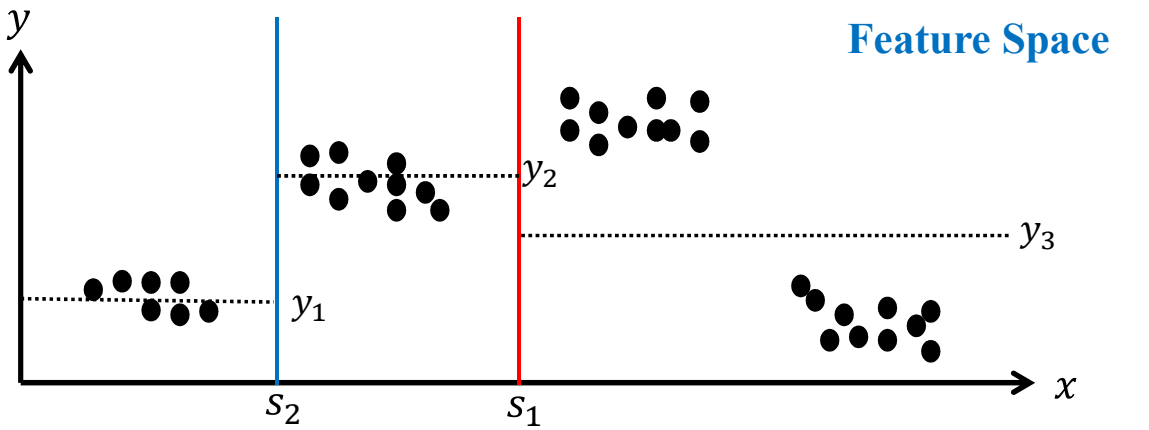
$$\sum_{i \in I_<} (y_i - \beta_<)^2 + \sum_{i \in I_>} (y_i - \beta_>)^2$$

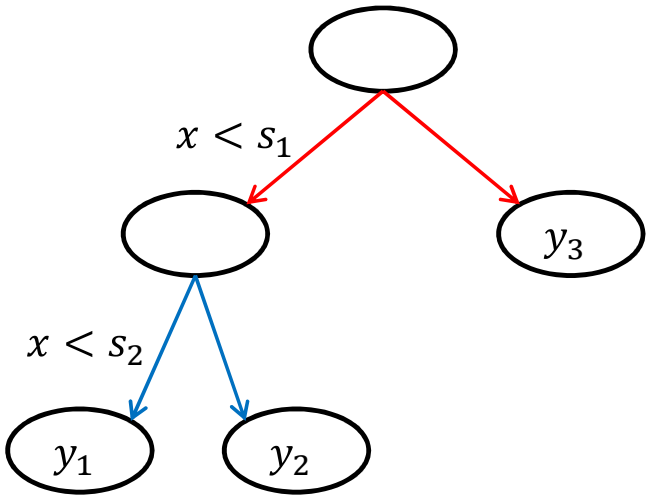
- Choose split with minimal loss.
- Recurse on both children, with $(x_i, y_i)_{i \in I_<}$ and $(x_i, y_i)_{i \in I_>}$.

Example of Regression Trees



Example of Regression Trees





Example of Regression Trees