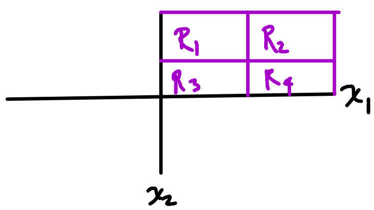
CART : Classification & Regression Trees

Suppose the feature space is given by

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_p \subseteq F$$



- Assume f is piecewise constant on $\{R_m\}_{m=1}^{M}$

and take $\hat{f}(x_i) = \frac{1}{|R_m|} \sum_{x_i \in R_m} y_i$

Choosing $\{R_m\}_1^M$

- Only consider cuts along the axes
- Find cut that most diminishes the empirical error.
- Stop when $n < n_{min}$ in some bin.

Process

$$(j,s)^* = \underset{j,s}{\arg\min} \left\{ \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right\}$$

To choose the depth of the tree penalize via

$$\sum_{m=1}^{M} N_m Q_m(T) + \alpha |T|$$

$$Q_m(T) = \frac{1}{N_m} \sum_{x_i \in R_m} (y_i - \hat{c}_m)^2$$

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m} y_i$$

The optimizing tree for penalty $C_\alpha(T)$ is achieved through this process.

<u>Boosting</u>: A collection of classifiers

<u>Ada Boost</u>: