Tree-based methods partition the feature space into a set of rectanggles and fit a simple model in each one. They are conceptually simple yet powerful tools.

Given a region $R$ which is a subset of domain $D$, define function $I_R$ on $D$

$$I_R(p) = \begin{cases} 1 & \text{if } p \in R, \\ 0 & \text{if } p \notin R \end{cases}$$

Consider a regression problem with response $y$ and feature $x \in \mathbb{R}^p$. A regression tree consists of a partition of $\mathbb{R}^p$ $R_1, R_2, \cdots, R_m$ and a prediction of $y$ for each region in the partition: $c_1, c_2, \cdots, c_m$. Formally

$$\hat{y} = \sum_{i=1}^{m} c_i I_{R_i}(x)$$

We will recursively define the process of growing a tree. Suppose we already have a partition $R_1, \cdots, R_{m-1}$. On each of the region $R_i$, we want to further split the region into two parts: $R_{i1}^{js} = \{x | x_j < s, x \in R_i\}$ and $R_{i2}^{js} = \{x | x_j >= s, x \in R_i\}$. Such a split has two unfixed parameters: which feature we are going to split over $(j)$ and where we are goint to split over $(s)$. We select this value by achieving the best local result in the target function. Suppose we have mean-square target fnction. Then

$$\begin{aligned} (j, s) &= \min \arg_{j,s}[\min_{c_1} \sum_{x_k \in R_{i1}^{js}} (y_k - c_1)^2 + \min_{c_2} \sum_{x_k \in R_{i2}^{js}} (y_k - c_2)^2] \\ &= \min \arg_{j,s}[ \sum_{x_k \in R_{i1}^{js}} (y_k - \bar{y}_{i1}^{js})^2 + \sum_{x_k \in R_{i2}^{js}} (y_k - \bar{y}_{i2}^{js})^2] \end{aligned}$$

Here $\bar{y}$ is the average of $y$ which belong to the region indicated by the subscripts of $\bar{y}$