

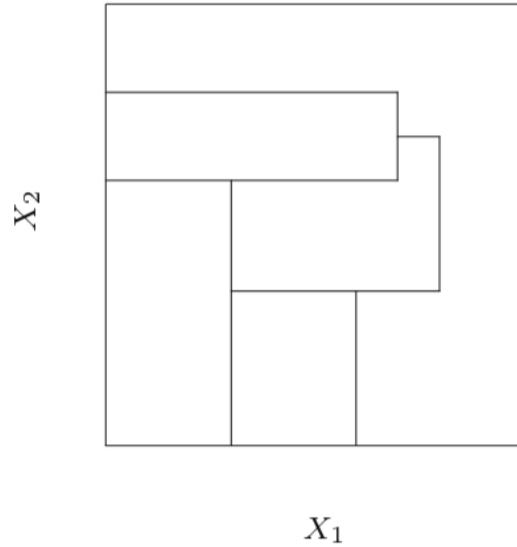
Decision Trees

Decision trees can be applied to both **regression** and **classification** problems, which makes it very flexible and interpretable.

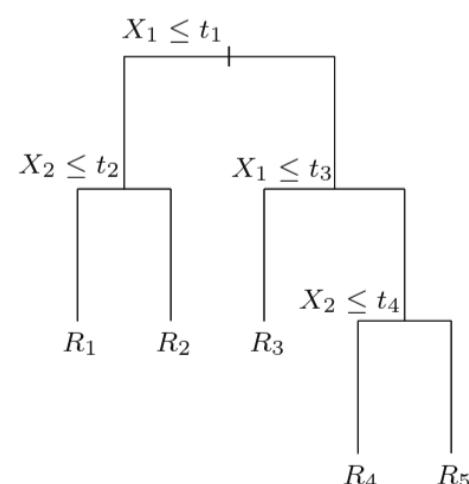
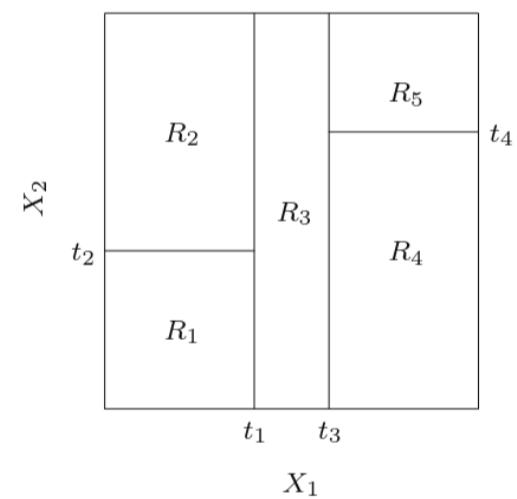
CART

Stands for **Classification and Regression Trees** which is an (algorithm/method) for tree based methods.

Considering a response Y and predictor X_1, X_2



By restricting it to **recursive binary partitions** by splitting the region into two and model the response by the mean of Y



Regression Trees

To construct a **regression tree**, the algorithm needs to automatically decide on splitting variables(feature) and splitting point(s) also what **shape** the tree should have.

$$y = \sum_{m=1}^M c_m I(x \in R_m)$$

- This models the response as a constant c_m in each region R_m .

Using the Residual Sum of Squares :

$$\sum_{i=1}^n (y_i - \sum_{m=1}^M c_m I(x \in R_m))^2$$

For one **Region** we get :

$$\mathcal{L}(c) = \sum_{i=1}^n (y_i - c)^2$$

Deriving w.r.t. c :

$$\frac{d \mathcal{L}}{d c} = \sum_{i=1}^n 2(c - y_i)$$

Setting it to zero results in:

$$\hat{c} = \frac{1}{N_m} \sum_{i=1}^n y_i \equiv \hat{c}_m = \text{ave}(y_i | x_i \in R_m)$$

Note: $\sum_{i \in R_m} c = N_m \cdot c$

To find the best binary partition in terms of minimum sum of squares is computationally infeasible. Hence **regression trees** use a greedy algorithm