Lect 19

Algorithm: Regression Tree Alg

① Begin with all data

② For every possible split of data

$\langle X_\ell, \vec{Y}_\ell \rangle, \langle X_r, \vec{Y}_r \rangle$

③ Calculate

$$SSE_\ell = \sum (Y_{\ell i} - \bar{Y}_\ell)^2$$

$$SSE_r = \sum (Y_{r_{1\ell}} - \bar{Y}_\ell)$$

④ Find split which

$SSE_{TOT} = SSE_\ell + SSE_r$ → take the minimum.

⑤ Create the split

⑥ ~~$\langle X_e, \vec{Y}_e \rangle$ Recurse data~~ in step 1 and do 2 ste
2~5

$\langle X_r, Y_r \rangle$ " " " " "

Recurse until "STOP"

STOP # of observation in a node $\leq N_0$.

(N_0 is a hyperparameter).

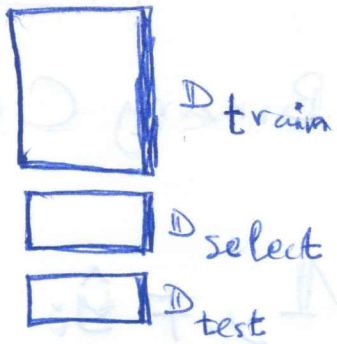
~~Default: N_0~~

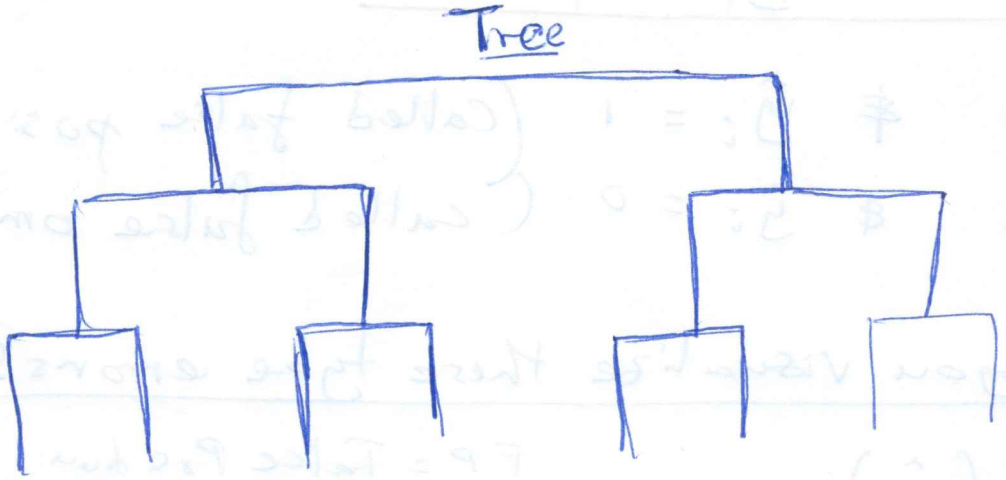
Default: $N_0 = 5$

⑥ For all leaf nodes, assign $\hat{y} = \bar{Y}_0$ where $\bar{Y}_0$ is sample avg of all Y_e's is node.

If default No = 5

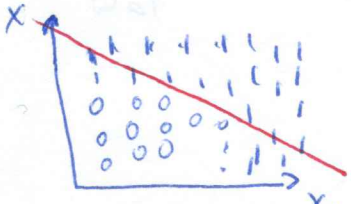
How to pick No?





$Y = \{0, 1\}$ binary classification.

$Y = \{1, 2, \ldots, k\}$ classification k-group / labels.



$$\mathcal{H} = \left\{ b_0 + w_1 x_1 + w_2 x - \vec{w} \in \mathbb{R}^3 \right\}$$

$$A = \operatorname{argm} \left\{ \frac{1}{n} \sum \max \left\{ 0, \frac{1}{2} - \left(y_i - \frac{1}{2}\right)\left(\vec{w}\vec{x} - b\right) + \lambda \| \vec{w} \right.$$

Now feed a tree