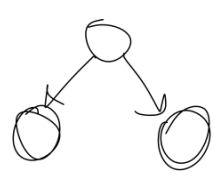


Classification and Regression Trees (Breiman et al. 1984)



x	0	0
0	0	0
x	x	x
x	0	0

$$S = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

d-dim. \mathbb{R}
For all possible features (d of them)
For all splits

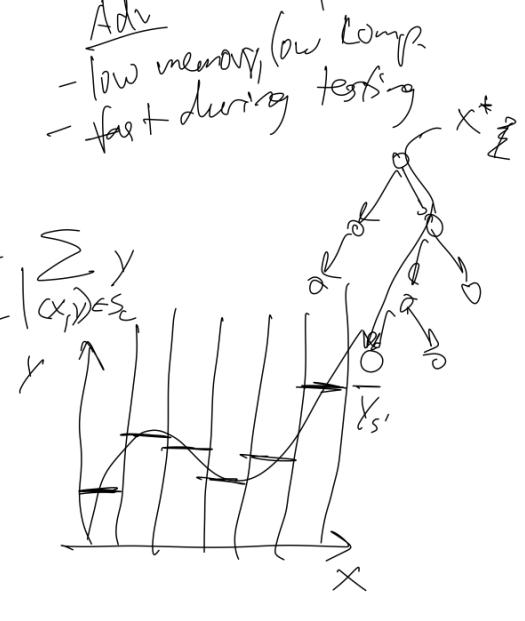
$$L(S_L) = \frac{1}{|S_L|} \sum_{(x,y) \in S_L} (y - \bar{y}_{S_L})^2$$

$$L = L(S_L) + L(S_R)$$

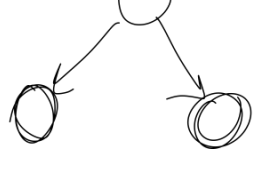
Disadv
- Generalization

Adv
- low memory, low comp
- fast during testing

what is complexity?



Classification and Regression Trees Breiman et al. 1984



x	x	x	x
0	x	0	x
0	x	0	0

$$S = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

2-dim. \mathbb{R}

Alg.

1. For all possible splits t

$$L(S_L^t) + L(S_R^t) = \frac{1}{|S_L^t|} \sum_{(x,y) \in S_L^t} (y - \bar{y}_{S_L^t})^2 + \frac{1}{|S_R^t|} \sum_{(x,y) \in S_R^t} (y - \bar{y}_{S_R^t})^2$$

$$\frac{1}{|S_L^t|} \sum_{(x,y) \in S_L^t} y$$

avg. spr. dif from avg. label y

2. Pick split t w/ lowest L^t

3. Recurse into child

4. Continue until stop \rightarrow min # of points

Disadv
- Generalization acf. not competitive

Adv
- low memory
- low computation

quiz: what is complexity?

- fast during training

Disadv
- Generalization acf. not competitive

Adv
- low memory
- low computation

bagging/boosting

