DECISION TREES

- supervised ML algo
- do NOT make any assumption about the structure of data OR relationship b/w input & output
- non-parametric model; this is more flexible in modelling
- model can't be represented by some numbers

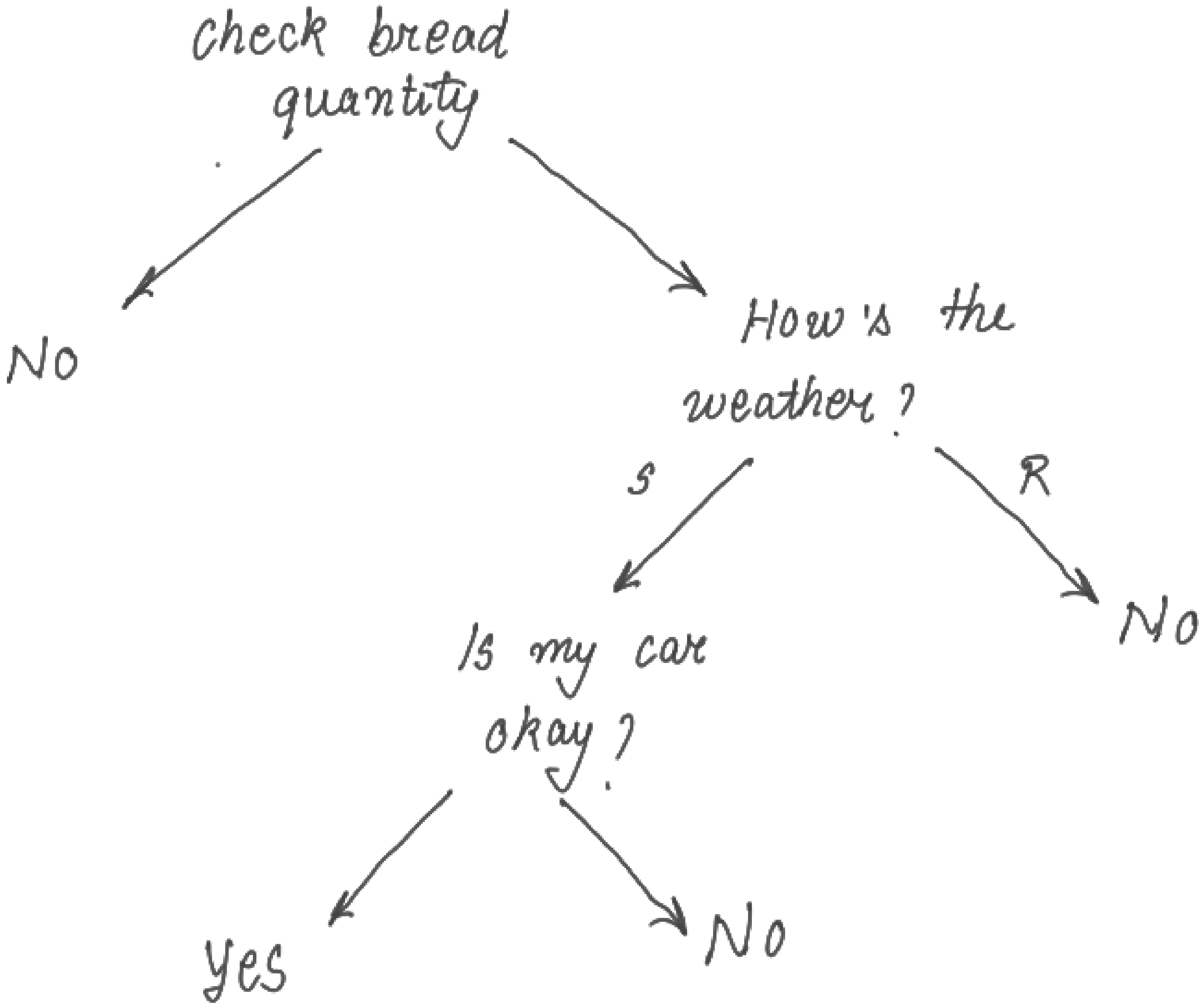
Why?

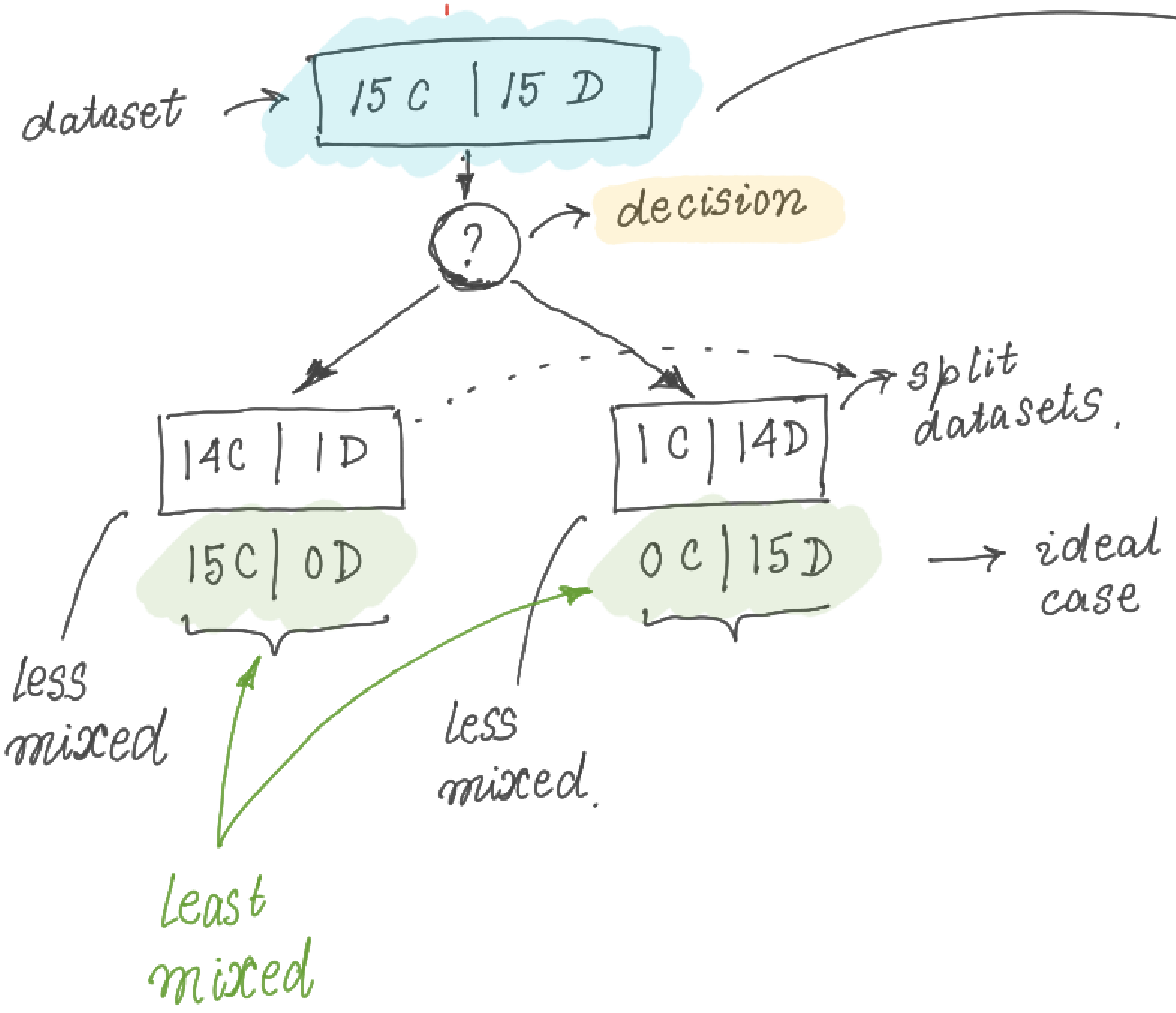
- linR & logR assume linearity b/w X & some form of Y.
- assume some structure of data that is linear
- assume some distribution of the parameters w.

$$w \sim t_{N-1}$$

- assumes some (usually Norm) distⁿ of residuals

Representative Example of a Decision Tree





Decision Trees for classification involve moving from

i. dataset that is more mixed (has higher entropy)

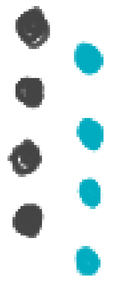
To

ii. several datasets that are lesser mixed. (have lower entropy)

Measuring homogenity.

Entropy : how uniform a particular collection of objects is.

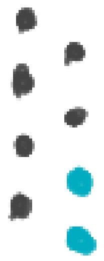
$$En = -\sum_i p_i \log_2 p_i \qquad \log_2 \frac{1}{2} = -1$$

1. 

$$p_B = \frac{1}{2} \quad p_B = \frac{1}{2}$$

$$En = -\left(p_B \log_2 p_B + p_B \log_2 p_B\right)$$

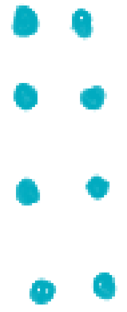
$$En = 1$$

2. 

$$p_B = \frac{6}{8} = \frac{3}{4} \quad p_B = \frac{1}{4}$$

$$En = -\left(\frac{3}{4}\log_2 \frac{3}{4} + \frac{1}{4}\log_2 \frac{1}{4}\right)$$

$$= 0.81$$

3. 

$$p_B = 1$$

$$En = -(1 \times \log_2 1) = 0$$

More homogeneous a sample; lesser is the entropy.

Choosing best decision

OD $\quad 15C \mid 15D$

1. Gender

$15C \mid 15D \xrightarrow{F} 2C \mid 8D$

$15C \mid 15D \xrightarrow{M} 13C \mid 7D$

$\underbrace{2C \mid 8D}_{En_{g,F}}$ $\qquad$ $\underbrace{13C \mid 7D}_{En_{g,M}} = 0.2811$

$$\left.\begin{array}{c} En = 1 \\ \downarrow \\ En_g = 0.2598 \end{array}\right\} \Delta = IG_{gender}$$

$$En_{g,F} = -\left(\frac{2}{10}\log_2\frac{2}{10} + \frac{8}{10}\log_2\frac{8}{10}\right)$$

$$= 0.2173$$

$$En_g = \frac{\overset{10}{n_F} \times En_{g,F} + \overset{20}{n_M} \times En_{g,M}}{n_F + n_M}$$

$$= 0.2598$$

Choosing best decision.

OD

2. Class

15C | 15D

1X: 6C | 8D

x: 9C | 7D

$En_{c,1x} = 0.2965$

$En_{c,x} = 0.2976$

$$En_c = \frac{14 \times 0.2965 + 16 \times 0.2976}{30} = 0.2970$$

$En = 1$

$\downarrow$

$En_c = 0.2970$

$\Delta = IG_{class}$

Information Gain

$$IG_f = En(X) - En(X, f)$$

ID3 Algorithm → (used only for classification DT)

- Find root node by calculating IG for all features.

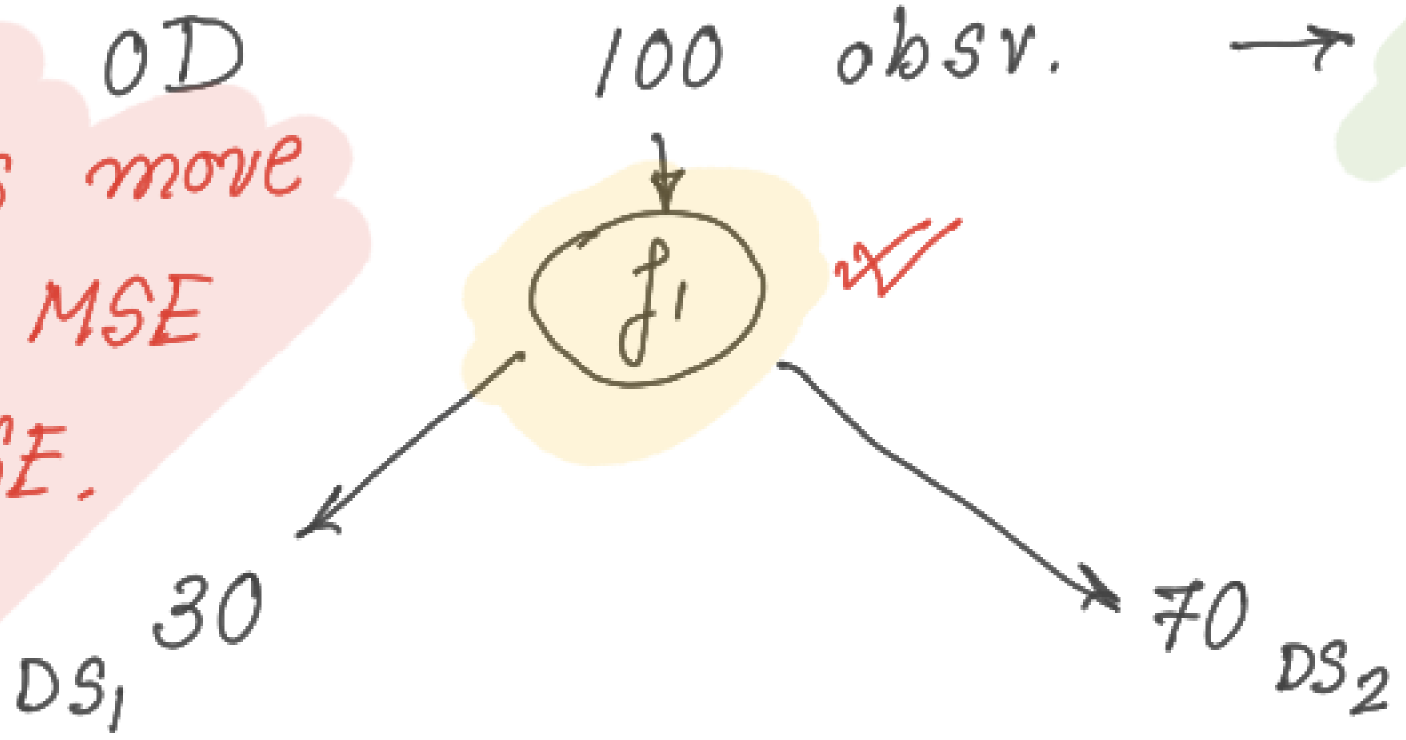
  $max(IG_f)$: → f becomes root node

- Find subsequent features based on max IG to grow the tree

- Until stopping criteria.

CART Algorithm (Regression Example)

Decision trees move from higher MSE to lower MSE.



$\rightarrow MSE_{OD} > MSE_{DS_1} + MSE_{DS_2}$

$\Delta_i = MSE_{initial} - MSE_{final, f_i}$

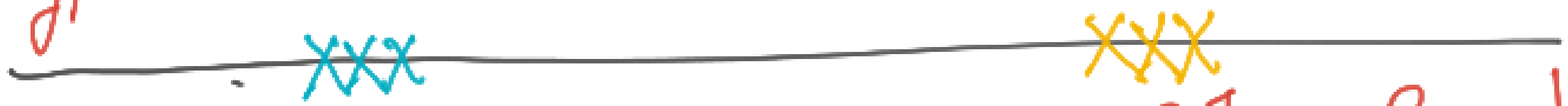
$\bar{y} = 52$

$\underbrace{1, 2, 3}_{DS_1}, \underbrace{101, 102, 103}_{DS_2}$

$\bar{y}_1 = 2$

$MSE_{OD} = \frac{1}{6}\left[\Sigma (y_i - 52)^2\right]$

$= 2500.66$

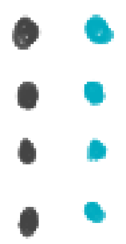


$MSE_{DS_1} = \frac{1}{3}\left[\Sigma (y_i - 2)^2\right] = \frac{2}{3}$ | $MSE_{DS_2} = \frac{2}{3}$

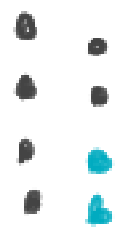
Measuring Homogenity II

Gini Index:

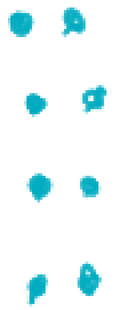
$$1-\sum_i p_i^2$$



$p_B = \frac{1}{2}$ $\quad p_B = \frac{1}{2}$ $\quad GI = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2} = \frac{4}{8}$



$p_B = \frac{3}{4}$ $\quad p_B = \frac{1}{4}$ $\quad GI = 1 - \frac{9}{16} - \frac{1}{16} = \frac{6}{16} = \frac{3}{8}$



$p_B = 0$ $\quad p_B = 1$ $\quad GI = 1 - (1)^2 = 0$

Easier to calculate than entropy.