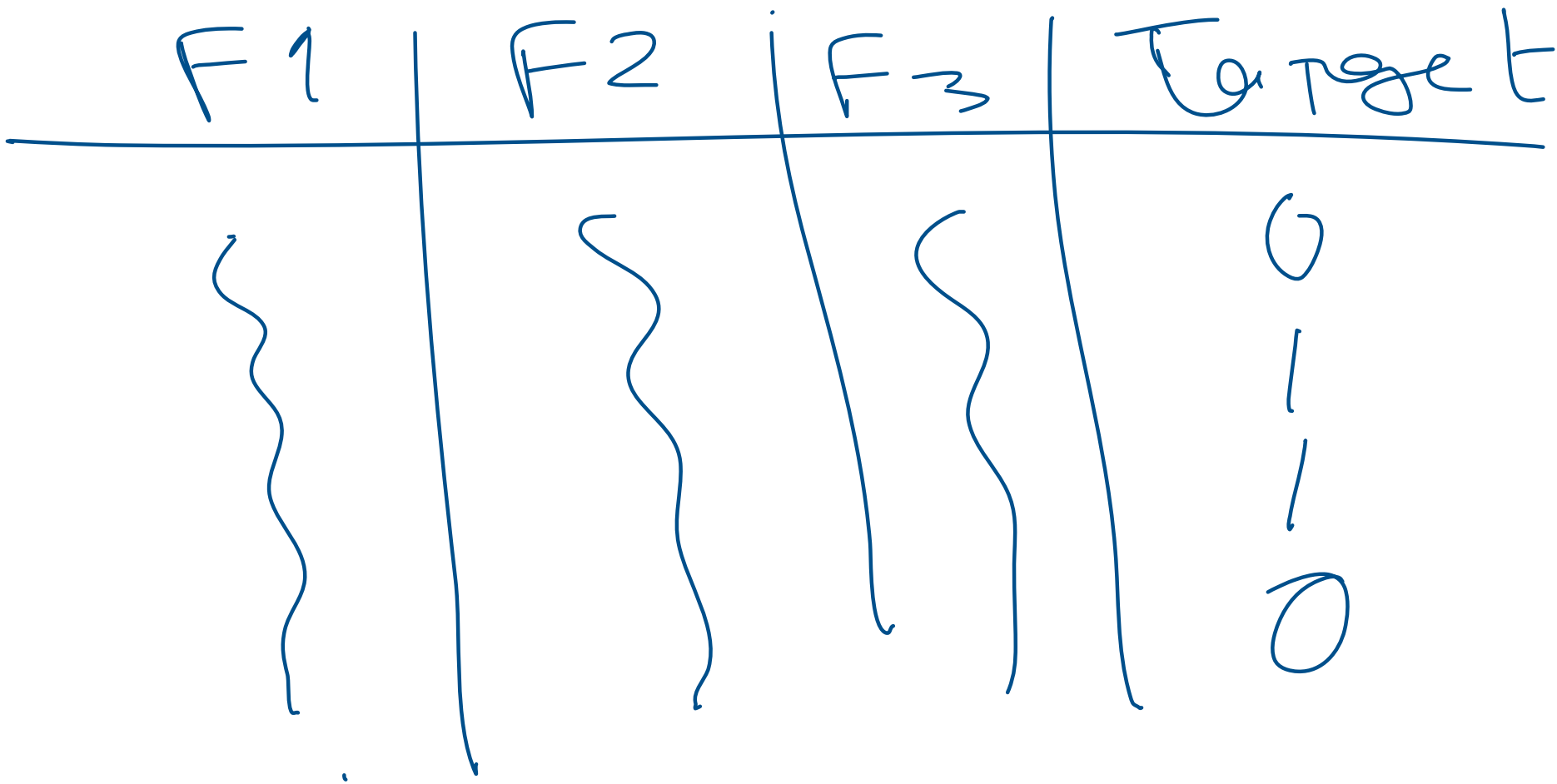
Decision Tree

- Root Node
- Node purity
- Decision Nodes

- Leaf Nodes


F1
F2
F3
Target
0
1
1
0

1. Entropy & IG (ID3)

2. Gini Index

Entropy

- variable

- avg. level of "informati

"surprise" / uncertainty

$$E = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

$$P(BC=Y) = \frac{12}{20} = 0.6$$

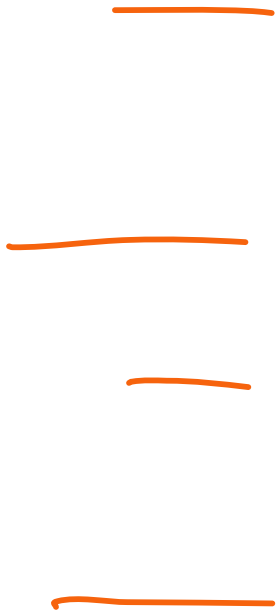
$$P(BC=N) = \frac{8}{20} = 0.4$$

$$\therefore E(BC) = -\sum_{i=1}^{2} P_i \log_2(P_i)$$

$$= -0.6\log_2(0.6) + 0.4\log_2(0.4$$

$$= 0.971$$

Information Gain



Age

$P(age < 18) = \frac{3}{20} = 0.15$

$P(age\ 18\text{-}35) = \frac{4}{20} = 0.2$

$P(age\ 36\text{-}55) = \frac{8}{20} = 0.4$

$P(age > 55) = \frac{5}{20} = 0.25$

$E(BC|age<18)$

$= -\left[\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\right.$

$= 0.913$

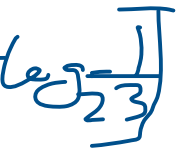
$E(BC|18\text{-}35)$

$= -\left[\frac{0}{4}\log_2\frac{0}{4} + \frac{4}{4}\log_2\right.$

$= 0$

$E(BC|36\text{-}55) = 0$

$E(BC|>55) = 0$



$$E(BC|Age) = 0.15 \times 0.913 + 0 +$$

$$0.4 \times 0.9544 + 0$$

$$= 0.5195$$

Education

$P(edu = M) = \frac{7}{20}$ → $E(BC|M) = 0.59$

$P(edu = B) = \frac{6}{20}$ → $E(BC|B) = 0.918$

$P(edu = H) = \frac{7}{20}$ → $E(BC|H) = 0.98$

$\therefore E(BC|edu) = 0.8267$

Income

$$P(inc = H) = \frac{10}{20} = 0.5$$

$$P(inc = L) = \frac{10}{20} = 0.5$$

$$E(BC \mid H) = 1$$

$$E(BC \mid L) = 0.88$$

$$\therefore E(BC(inc) = 0.9407$$

Marrital

$P(mar = S) =$

$P(mar = M) =$

$\therefore E(BC \mid mar) = 0.8685$

9Q4

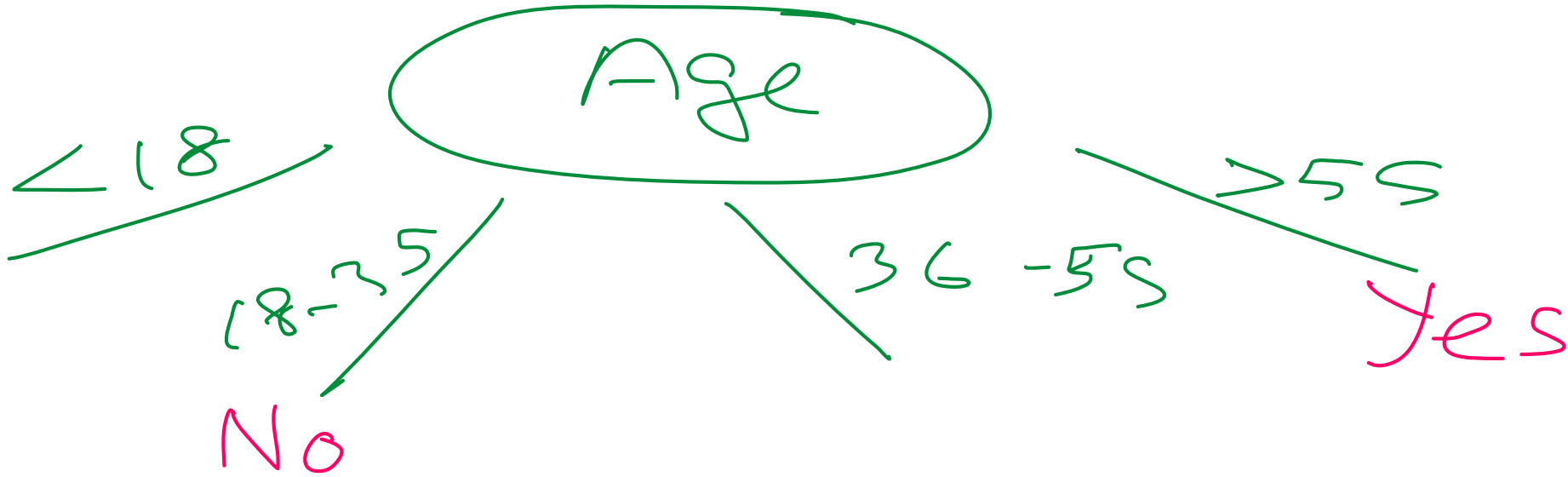
$IG(BC|age) = 0.971 - 0.5195 = 0.45$

$IG(BC|edu) = 0.971 - 0.8267 = 0.14$

$IG(BC|inc) = 0.971 - 0.9407 = 0.0$

$IG(BC|mar) = 0.971 - 0.868 5 = 0.10$

Root Node is Age



Age > 18

$P(edu = H) = 1$

$E(BC \mid edu) = 0$

$E(BC) = -[\frac{1}{2}$

$+$

$= 0.$

$P(inc = H) = \frac{1}{3}$

$P(inc = L) = \frac{2}{3}$

$E(BC|inc)$

$= -(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log$

$= 0.9183$

$= \log_{\frac{1}{2}} \frac{1}{3}$

$= \frac{2}{3} \log_2 \frac{1}{3}$)

$= 0.183$

$\log_2 \frac{2}{3}$)

$$P(\text{man} = s) = \frac{2}{3}$$

$$P(\text{man} = m) = \frac{1}{3}$$

$$E(BC \mid \text{man}) = 0.9183$$

$IG(BC|edu) = 0.9183 - 0$

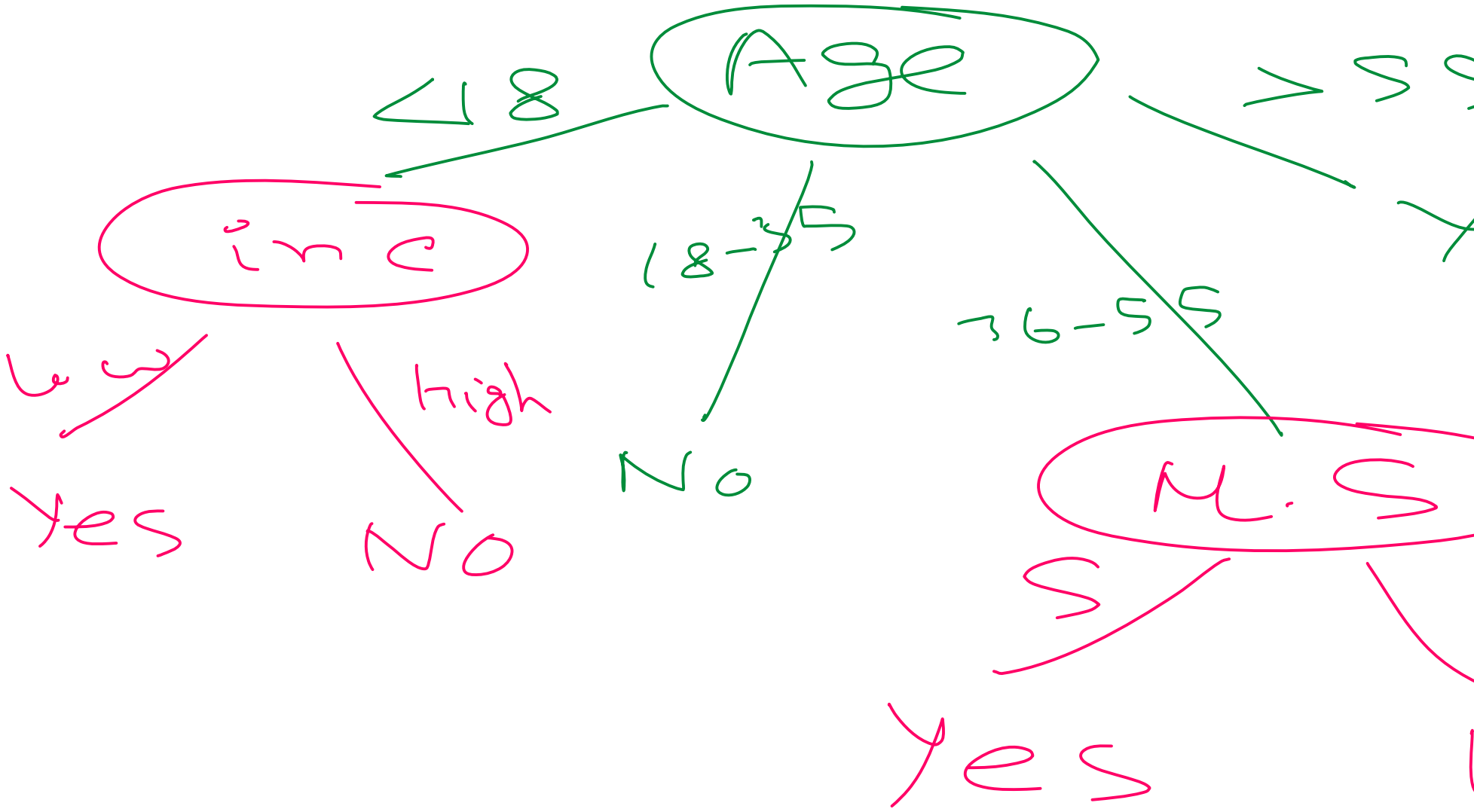
$IG(BC|inc) = 0.9183 - 0.9$

$IG(BC|mon) = 0$

$= 0.9183$

$3 = 0$

C


Age
< 18
> 55
inc
18-35
36-55
low
high
No
Yes
No
M.S
S
Yes

es

M

No

Disadv

Decision Tree classifier

— Hyperparameters

Pruning

Pre

Post

CPP