A1) How to calculate info gain

$$H(t,D) = -\sum_{l \in levels(t)} \left(P(t=l) \times \log_2 P(t=l)\right)$$

$$rem(d,D) = \sum_{l \in levels(d)} \frac{|D_{d=l}|}{D} \times H(t, D_{d=l})$$

$$IG(d,D) = H(t,D) - rem(d,D)$$

$$H(t,D) = -\sum_{l \in (true, false)} \left(P(t=l) \times \log_2 P(t=l)\right)$$

$$= -\left(P(t=true) \times \log_2 P(t=true)\right) - \left(P(t=false) \times \log_2 P(t=false)\right)$$

$$= -\left(3/6 \times \log_2(3/6)\right) - \left(3/6 \times \log_2(3/6)\right)$$
$$= -2\left(1/2 \times \log_2(1/2)\right) = -\log_2(0.5)$$
$$= 1 \text{ bit}$$

GB ≡ Good Behavior
A ≡ Age < 30
DD ≡ Drug Dependent
R ≡ Recidivist
T ≡ true
F ≡ false

$$rem(GB, D) = \left(\frac{|D_{GB=true}|}{|D|} \times H(t, D_{GB=true})\right)$$

$$+ \left(\frac{|D_{GB=false}|}{|D|} \times H(t, D_{GB=false})\right)$$

$$= (3/6(-P(t=true) \times \log_2 P(t=true))$$

$$- P(t=false) \times \log_2 P(t=false))$$

$$+ (3/6(-P(t=false) \times \log_2 P(t=false))$$

$$- P(t=true) \times \log_2 P(t=true)))$$

$$= 1/2\left(-1/3 \log_2(1/3) - 2/3 \log_2(2/3)\right)$$

$$+ 1/2\left(-2/3 \log_2(2/3) - 1/3 \log_2(1/3)\right)$$

$$= -1/6\left(2\log_2(1/3) + 4\log_2(2/3)\right)$$

$$= 0.918 \text{ bits}$$

$$\therefore \boxed{IG(GB, D) = (1-0.918) \text{ bits} = 0.082 \text{ bits}}$$

$$\mathrm{rem}(DD, D) = \left(\frac{|D_{DD} = \mathrm{true}|}{|D|} \times H(t, D_{DD} = \mathrm{true})\right)$$

$$+ \left(\frac{|D_{DD} = \mathrm{false}|}{|D|} \times H(t, D_{DD} = \mathrm{false})\right)$$

$$= \tfrac{1}{6}\left(-1 \times \log_2(1) - 0\right)$$

$$+ \tfrac{5}{6}\left(-\tfrac{2}{5}\log_2\left(\tfrac{2}{5}\right) - \tfrac{3}{5}\log_2\left(\tfrac{3}{5}\right)\right) = 0.809 \text{ bits}$$

$$\therefore \boxed{IG(DD, D) = (1 - 0.809) \text{ bits} = 0.191 \text{ bits}}$$

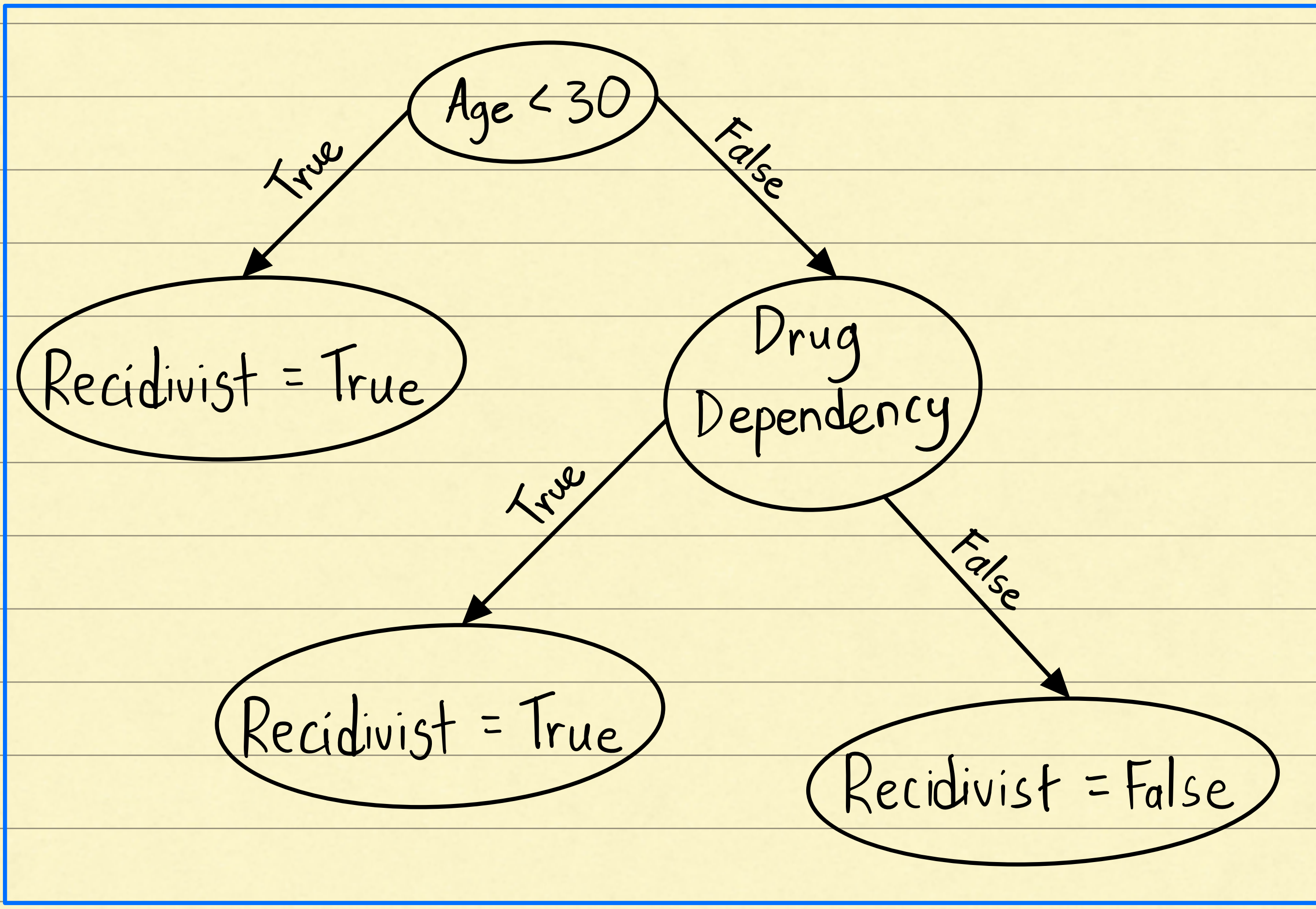
$$\mathrm{rem}(A, D) = \tfrac{2}{6}\left(-1\log_2(1) - 0\right)$$

$$- \tfrac{4}{6}\left(-\tfrac{1}{4}\log_2\left(\tfrac{1}{4}\right) - \tfrac{3}{4}\log_2\left(\tfrac{3}{4}\right)\right) = 0.541 \text{ bits}$$

$$\therefore \boxed{IG(A, D) = (1 - 0.541) \text{ bits} = 0.459 \text{ bits}}$$

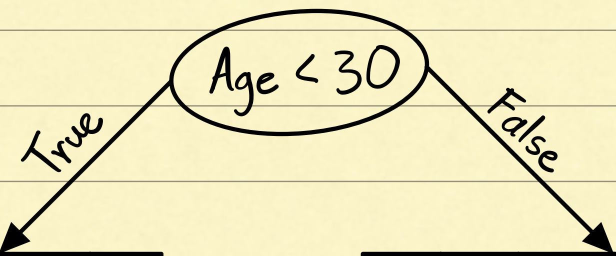
Order of IG : $A > DD > GB$

Final Decision Tree



All calculations shown on following pages...

Decision Tree:



ID	GB	DD	R
1	F	F	T
3	F	F	T

	ID	GB	DD	R
	2	F	F	F
D_2	4	T	F	F
	5	T	T	T
	6	T	F	F

After splitting along elevation, we need to split among the features in the false branch

$$H(R, D_2) = -\left(P(R = \text{true}) \times \log_2 P(R = \text{true})\right)$$

$$- \left(P(R = \text{false}) \times \log_2 P(R = \text{false})\right)$$

$$= -\tfrac{1}{4} \log_2\left(\tfrac{1}{4}\right) - \tfrac{3}{4} \log_2\left(\tfrac{3}{4}\right) = 0.811 \text{ bits}$$

$$\text{rem}(GB, D_2) = \tfrac{3}{4}\left(-\tfrac{1}{3} \log_2\left(\tfrac{1}{3}\right) - \tfrac{2}{3} \log_2\left(\tfrac{2}{3}\right)\right)$$

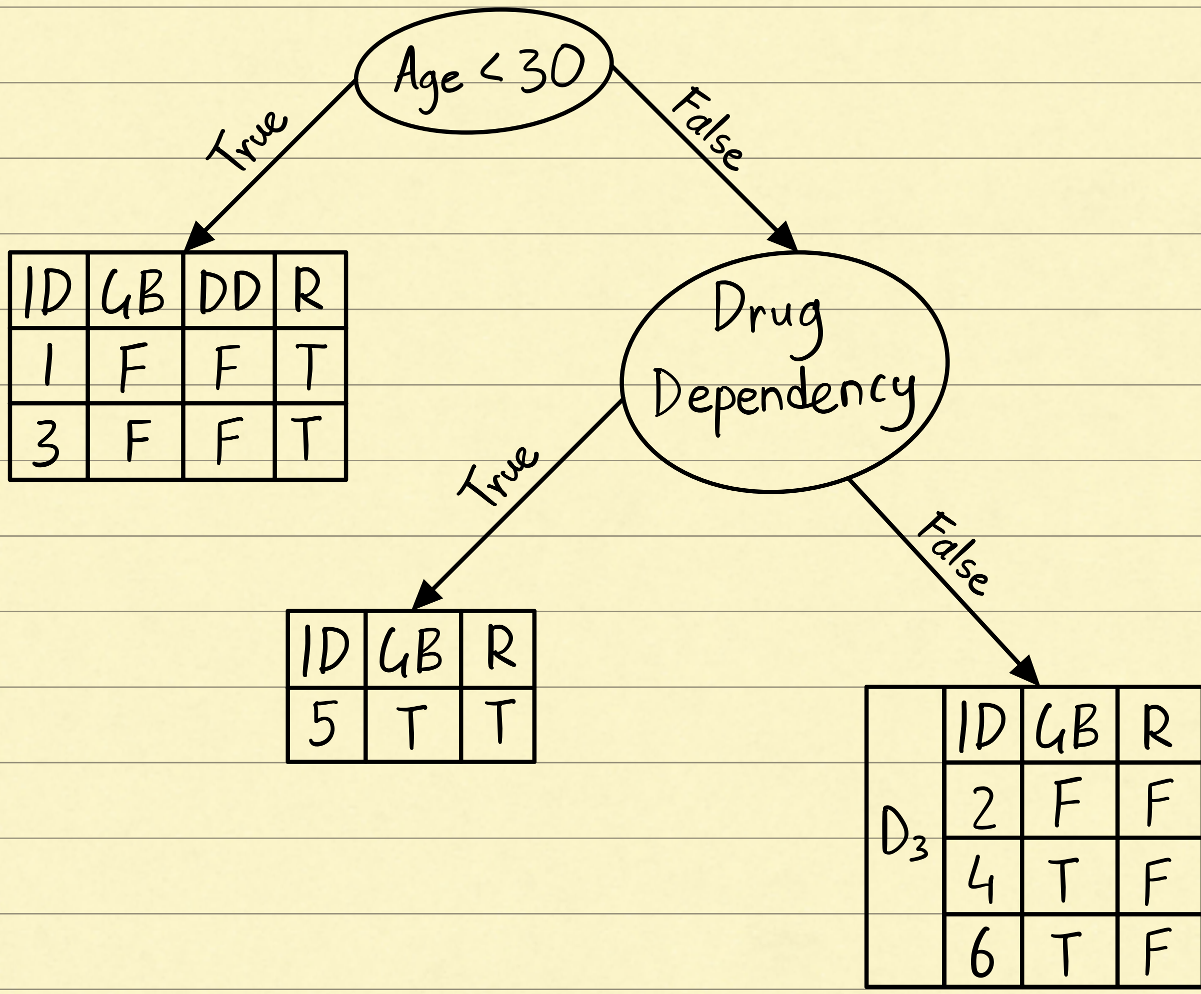
$$+ \tfrac{1}{4}\left(-1 \log_2(1) - 0\right) = 0.689 \text{ bits}$$

$$\therefore \boxed{IG(GB, D_2) = (0.811 - 0.689) \text{ bits} = 0.112 \text{ bits}}$$

$$rem(DD, D_2) = \tfrac{1}{4}\left(-1\log_2(1) - 0\right) + \tfrac{3}{4}\left(-1\log_2(1) - 0\right)$$

$$= 0 \text{ bits}$$

$$\therefore \boxed{IG(DD, D_2) = (0.811 - 0) \text{ bits} = 0.811 \text{ bits}}$$



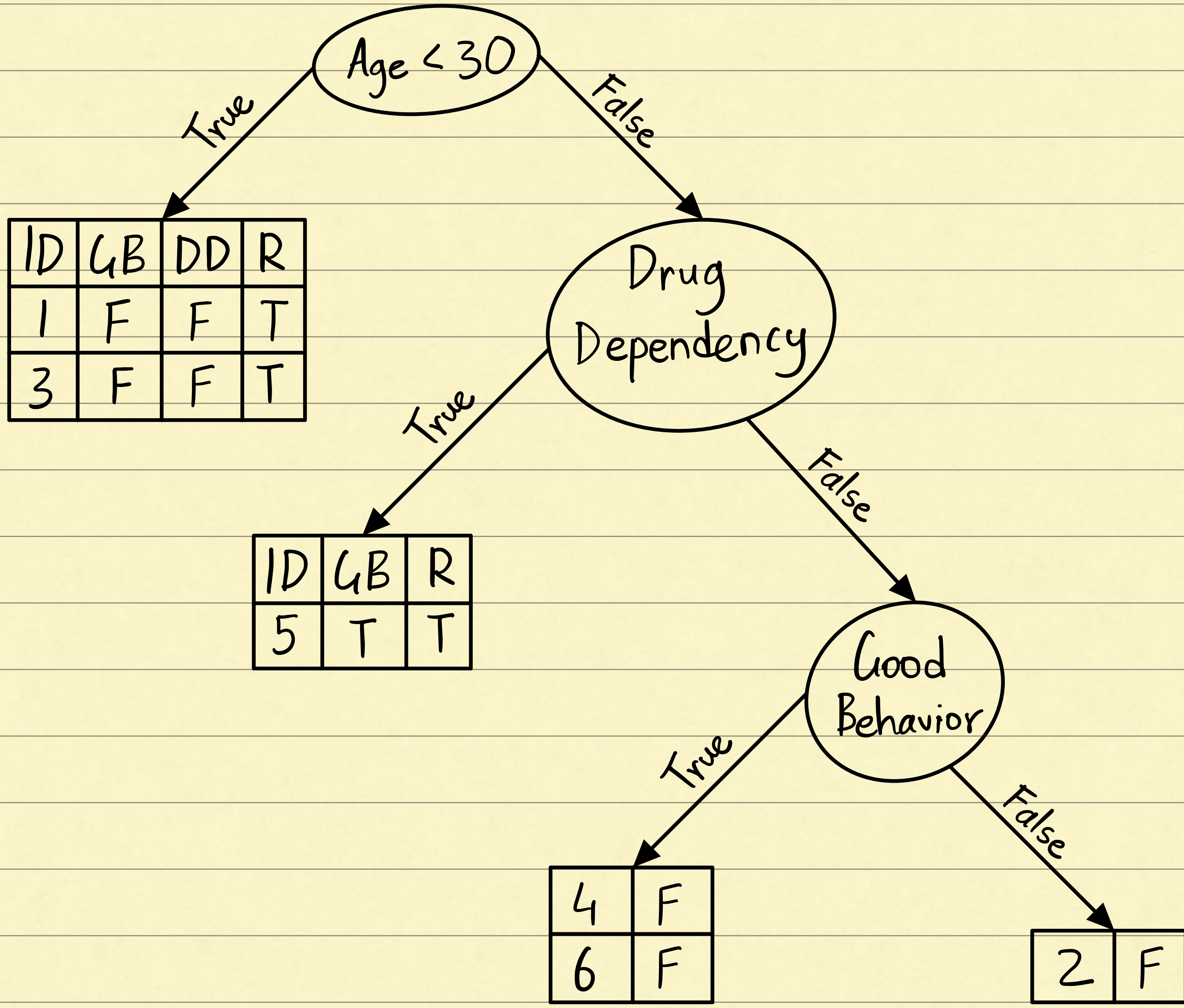
ID	GB	DD	R
1	F	F	T
3	F	F	T

ID	GB	R
5	T	T

D_3:

ID	GB	R
2	F	F
4	T	F
6	T	F

Only remaining feature to split on is GB


Age < 30
True
False
ID GB DD R
1 F F T
3 F F T
Drug Dependency
True
False
ID GB R
5 T T
Good Behavior
True
False
4 F
6 F
2 F