

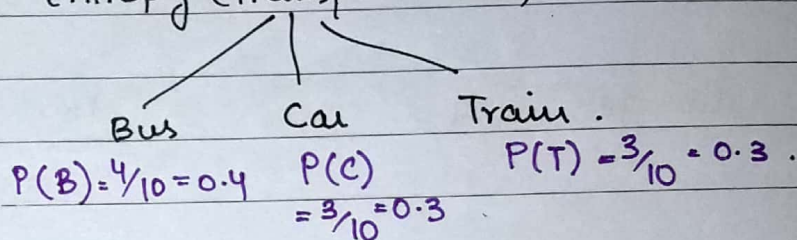
# Date: Decision Tree Class Activity

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum [P(S|A) \cdot \text{Entropy}(S|A)]$$

(A) Attribute  $\rightarrow$  Travel Cost  $\begin{cases} \text{a) Cheap} \\ \text{b) Standard} \\ \text{c) Expensive.} \end{cases}$

So we have :

$$E(S) = 1.5704 \quad \text{Entropy (Transportation)}$$



Travel cost (A)	Transportation (B)
Cheap	Bus
Cheap	Bus
Cheap	Train
Cheap	Bus
Cheap	Bus
Standard	Train
Standard	Train
Expensive	Car
Expensive	Car
Expensive	Car

$$P(S|\text{cheap}) = 5/10 = 0.5$$

$$P(S|\text{std}) = 2/10 = 0.2$$

$$P(S|\text{exp}) = 3/10 = 0.3$$

for "cheap" class :  $\begin{cases} \text{Bus} = 4/5 = 0.8 \\ \text{Train} = 1/5 = 0.2 \\ \text{Car} = 0 \end{cases}$

$$\begin{aligned} E(A_1) &= -P(B|\text{cheap}) \cdot \log_2(P(B|\text{cheap})) - P(T|\text{cheap}) \cdot \log_2(P(T|\text{cheap})) \\ &\quad - P(C|\text{cheap}) \cdot \log_2(P(C|\text{cheap})) \\ &= -0.8 \cdot \log_2(0.8) - 0.2 \cdot \log_2(0.2) \\ &= 0.257 + 0.464 \end{aligned}$$

Date:

$$E(S|A=cheap) = 0.721$$

for "standard" :

$$\begin{aligned} \text{Bus} &= 0/2 = 0 \\ \text{Train} &= 2/2 = 1 \\ \text{Car} &= 0/2 = 0 \end{aligned}$$

$$\begin{aligned} E(S|A=std) &= -P(T|A=std) \cdot \log_2(T|A=std) - P(B|A=std) \cdot \log_2(P(B|A)) \\ &\quad - P(C|A=std) \cdot (\log_2(P(C|A=std))) \\ &= -1(\log_2(1)) - 0 \end{aligned}$$

$$P(S|A=std) = 0$$

for "Expensive" :

$$\begin{aligned} \text{Bus} &= 0/3 = 0 \\ \text{Train} &= 0/3 = 0 \\ \text{Car} &= 3/3 = 1 \end{aligned}$$

$$\begin{aligned} E(S|A=Exp.) &= -P(T|A=Exp.) \cdot (\log_2(T|A=exp.)) - P(B|exp.) \cdot \log_2(B|exp) \\ &\quad - P(C|A=exp.) \cdot (\log_2(C|A=exp)) \\ &= -0-0-1(\log_2(1)) \end{aligned}$$

$$E(S|A=exp.) = 0.$$

Now,

$$\begin{aligned} \text{Gain}(S;A) &= \text{Entropy}(S) - \sum [P(S|A) \times \text{Entropy}(S|A)] \\ &= 1.5704 - [0.5 \times 0.721] - [0.2 \times 0] - [0.3 \times 0] \\ &= 1.5704 - 0.3605 \\ &= 1.2099 \end{aligned}$$

$$\boxed{\text{Gain}(S,A) = 1.21} \quad \underline{\text{Ans.}}$$