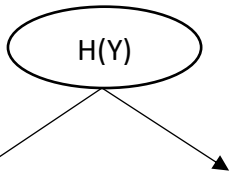
Student: Mi Yang
SID:3034392755

HW 7: Entrophy

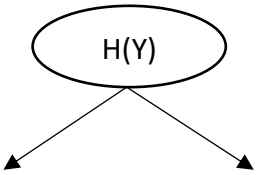
1.Let Y denote Defaulter, X1 denote Hasjob, X2 denote HasFamily, X3 denote IsAbove30years.



X1=0 W p=3/8
P(Y=1|X1=0)=2/3
H(Y=1|X1=0)
=2/3log3/2+1/3log3
=0.918

X1=1
P(Y=1|X1=1)=2/5
H(Y=1|X1=1)
=2/5log5/2+3/5log5/3
=0.971

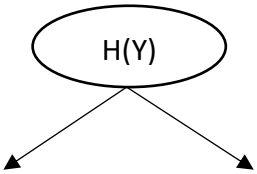
H(Y|X)=3/8x0.918+5/8x0.971=0.951
Info gained=H(Y)- H(Y|X)=1-0.951=0.049



X2=0 W p=1/2
P(Y=1|X1=0)=3/4
H(Y=1|X1=0)
=3/4log4/3+1/4log4
=0.81

X1=1
P(Y=1|X1=1)=1/4
H(Y=1|X1=1)
=1/4log4+3/4log3/4
=0.81

H(Y|X)=0.81x0.5+0.81x0.5=0.81
Info gained=H(Y)- H(Y|X)=1-0.81=0.19



X3=0 W p=1/4
P(Y=1|X1=0)=1/2
H(Y=1|X1=0)

X1=1
P(Y=1|X1=1)=1/2
H(Y=1|X1=1)

=1/2log2+1/2log2 =1/2log2+1/2log2
= 1 =1
H(Y|X)=1/4x1+3/4x1=1
Info gained=H(Y)- H(Y|X)=1-1=0

From the analysis above, it can be seen that the info gained from X2 is the biggest, so HasFamily is the best feature to do the first split in a binary tree.

2.
H(A)=$\log_2(\frac{10}{7})$ H(B)=$\log_2(5)$ H(C)=$\log_2(10)$

Info content:
H(S)= $\frac{7}{10}\log_2\left(\frac{10}{7}\right)+\frac{2}{10}\log_2(5)+\frac{1}{10}\log_2(10)$ =1.157

According to the source coding theorem, H(S) is the smallest codeword length that is theoretically possible for 'S' which is 1.157 bits per symbol.