**HW7**
**Name: Shubei Wang**
**Student ID: 3034358656**

## 1.

Denote “HasJob”, “HasFamily”, “IsAbove30years”, “Defaulter” as X1, X2, X3, Y, respectively.

Y = 1 with p = 0.5.
H(Y) = 0.5log(1/0.5)+0.5log(1/0.5) = 1.

Split X1:
P(X1=1) = 5/8, P(X1=0) = 3/8
P(Y=1|X1=1) = 2/5 —> H(Y|X1=1) = 2/5*log(5/2)+3/5*log(5/3) = 0.971
P(Y=1|X1=0) = 2/3 —> H(Y|X1=0) = 2/3*log(3/2)+1/3*log(3/1) = 0.918
H(Y|X1) = 5/8*0.971+3/8*0.918 = 0.951
Info Gained = H(Y)-H(Y|X1) = 0.049

Split X2:
P(X2=1) = 0.5, P(X2=0) = 0.5
P(Y=1|X2=1) = 0.25 —> H(Y|X2=1) = 0.25*log(1/0.25)+0.75*log(1/0.75) = 0.811
P(Y=1|X2=0) = 0.75 —> H(Y|X2=0) = 0.25*log(1/0.25)+0.75*log(1/0.75) = 0.811
H(Y|X2) = 0.5*0.811+0.5*0.811 = 0.811
Info Gained = H(Y)-H(Y|X2) = 0.189

Split X3:
P(X3=1) = 0.75, P(X3=0) = 0.25
P(Y=1|X3=1) = 0.5 —> H(Y|X3=1) = 0.5*log(1/0.5)+0.5*log(1/0.5) = 1
P(Y=1|X3=0) = 0.5 —> H(Y|X3=0) = 0.5*log(1/0.5)+0.5*log(1/0.5) = 1
H(Y|X3) = 0.75*1+0.25*1 = 1
Info Gained = H(Y)-H(Y|X3) = 0

From the calculation above we know that X2(“HasFamily”) is the best feature because it leads to the most information gained.

## 2.

H(S) = 0.7*log(1/0.7)+0.2*log(1/0.2)+0.1*log(1/0.1) = 1.157 bits
The H(S) is the smallest codeword length that is theoretically possible for signal ’S’. So theoretically the smallest code is 1.157 bits per symbol for S.