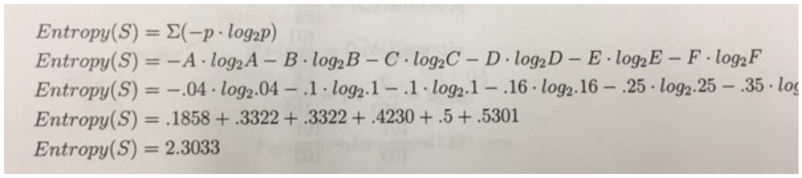
1 Q1

Just plug in the numbers in then entropy formula: $\sum_{color=A}^{F} -(p_{color} . \log_2 p_{color})$



$Entropy(S) = \Sigma(-p \cdot log_2 p)$
$Entropy(S) = -A \cdot log_2 A - B \cdot log_2 B - C \cdot log_2 C - D \cdot log_2 D - E \cdot log_2 E - F \cdot log_2 F$
$Entropy(S) = -.04 \cdot log_2 .04 - .1 \cdot log_2 .1 - .1 \cdot log_2 .1 - .16 \cdot log_2 .16 - .25 \cdot log_2 .25 - .35 \cdot log$
$Entropy(S) = .1858 + .3322 + .3322 + .4230 + .5 + .5301$
$Entropy(S) = 2.3033$

Figure 1:

2 Q2

- Is the classifier produced by Naive Bayes identical when it learns on D1 and D2?

  No The classifier is not the same, keeping everything else including the prior and conditional probabilities the same. The classifier would compute $P(class|att_1, att_2, ...att_n) = \frac{P(att_1|class)...P(att_n|class)P(class)}{P(att_1)...P(att_n)}$

  As can be seen in the formula above duplicating an attribute would result in the duplication of its conditional probability give class on the right side of the equation (numerator), which will give a different result.

- If you answer Yes please explain your answer clearly. If you answer No explain and give an example where the prediction of the two classifiers will be different.

  See the example below:

Table 1: D1

a_1	a_2	a_3	Class
1	1	1	0
1	1	0	0
0	0	0	0
1	1	1	1
0	0	1	1
0	0	0	1

Table 2: D2

a_1	a_2	a_{3^1}	a_{3^2}	a_{3^3}	Class
1	1	1	1	1	0
1	1	0	0	0	0
0	0	0	0	0	0
1	1	1	1	1	1
0	0	1	1	1	1
0	0	0	0	0	1

Observation: $b_1 = 0, b_2 = 0, b_3 = 0$
D1 trained classifier: $P(b_1, b_2, b_3 \mid class = 0)P(class = 0) = (\frac{1}{3})^2 \cdot \frac{2}{3} \cdot \frac{1}{2}$
$P(b_1, b_2, b_3 \mid class = 1)P(class = 1) = (\frac{2}{3})^2 \cdot \frac{1}{3} \cdot \frac{1}{2}$
Classify the observation as class 1.

D2 trained clasifier: $P(b_1, b_2, b_{3^1}, b_{3^2}, b_{3^3} \mid class = 0)P(class = 0) = (\frac{1}{3})^2 \cdot (\frac{2}{3})^3 \cdot \frac{1}{2}$
$P(b_1, b_2, b_{3^1}, b_{3^2}, b_{3^3} \mid class = 1)P(class = 1) = (\frac{2}{3})^2 \cdot (\frac{1}{3})^3 \cdot \frac{1}{2}$
Classify the observation as class 0.

Figure 2:

3 Q3

- How deep is the non-pruned ID3 tree for this data? (show your work)

  $1 \leq depth \leq 4$! The maximum depth is 4 which is equal to the number of attributes and the minimum depth is 1 which means that there is one attribute which classifies the entire data correctly with error=0 and entropy=0.

Level	Description	pos	neg	total	pos(frac)	neg(frac)	E	gain
0	ROOT	6	6	12	0.5	0.5	1	
	prev Morn							
	light	5	2	7	0.71	0.29	0.86	
	medium	1	1	2	0.50	0.50	1.00	
	heavy	0	3	3	0.00	1.00	0.00	0.33
	prev day							
	light	4	2	6	0.67	0.33	0.92	
	medium	1	3	4	0.25	0.75	0.81	
1	heavy	1	1	2	0.50	0.50	1.00	0.10
	prev night							
	light	2	3	5	0.40	0.60	0.97	
	medium	1	3	4	0.25	0.75	0.81	
	heavy	3	0	3	1.00	0.00	0.00	0.33
	early morn							
	light	1	6	7	0.14	0.86	0.59	
	medium	3	0	3	1.00	0.00	0.00	
	heavy	2	0	2	1.00	0.00	0.00	0.65
	LIGHT BRANCH							
	prev Morn							
	light	1	2	3	0.33	0.67	0.92	
	medium	0	1	1	0.00	1.00	0.00	
	heavy	0	3	3	0.00	1.00	0.00	0.20
	prev day							
2	light	1	2	3	0.33	0.67	0.92	
	medium	0	3	3	0.00	1.00	0.00	
	heavy	0	1	1	0.00	1.00	0.00	0.20
	prev night							
	light	0	3	3	0.00	1.00	0.00	
	medium	0	3	1	0.00	1.00	0.00	
	heavy	1	0	3	1.00	0.00	0.00	0.59

Figure 3:

- What is the non-pruned ID3 tree for this data? (show your work) See below.

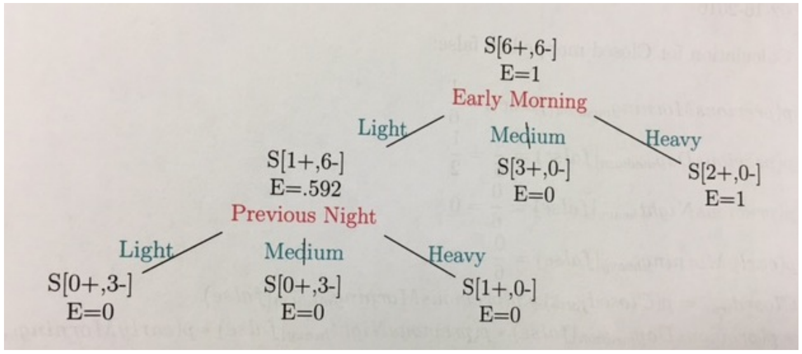


Figure 4:

4 Q4

This question can be better understood in relation to the next question. In the next question we are going to computed the posterior probability of closed =T/F give an observation. In order to be able to do that, we need the prior probabilities of closed=T and closed=F which can be computed based on the training data observed in the Q3.
p(closed=T) = $\frac{3}{6} = 1/2$,
p(closed=F) = $\frac{3}{6} = 1/2$

5 Q5

We need to compute the posterior probability of closed being true or false (hypothesis) given our observation (data), under the assumption of naive bayes which indicates that features are independent of each other. Therefore we are looking for:

$P(H|D) = \frac{P(D|H)P(H)}{P(D)} = \frac{P(D|H)P(H)}{\sum_{H_i} P(D|H)P(H)}$

Given the naive bayes assumption, the likelihood term in the numerator of the equation above can be computed in the following way for closed = True (hypothesis).

$P(D|H = T) = P(PM = m|closed = T)P(PD = M|closed = T)P(PN = H|closed = T)P(EM = H|closed = T)$

The denominator can be computed as:
$P(D) =$
$P(PM = m|closed = T)P(PD = M|closed = T)P(PN = H|closed = T)P(EM = H|closed = T)P(closed = T)+$

$P(PM = m|closed = F)P(PD = M|closed = F)P(PN = H|closed = F)P(EM = H|closed = F)P(closed = F)$

Plug in the priors computed in Q4 as well as the conditional probabilities (see below) and compute the posteriors.

$$p(previousMorning_{medium}|true) = \frac{1}{6}$$
$$p(previousDay_{medium}|true) = \frac{1}{6}$$
$$p(previousNight_{heavy}|true) = \frac{3}{6} = \frac{1}{2}$$
$$p(earlyMorning_{heavy}|true) = \frac{2}{6} = \frac{1}{3}$$

Figure 5:

$$p(previousMorning_{medium}|false) = \frac{1}{6}$$
$$p(previousDay_{medium}|false) = \frac{3}{6} = \frac{1}{2}$$
$$p(previousNight_{heavy}|false) = \frac{0}{6} = 0$$
$$p(earlyMorning_{heavy}|false) = \frac{0}{6} = 0$$

Figure 6: