Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

1. Short Answers (35 points total)

(a) **(5 pts)** Compute the entropy of this set S: 1 1 1 2 2 3 3 3. Write **one number** for each.

P(1|S) = 3/8

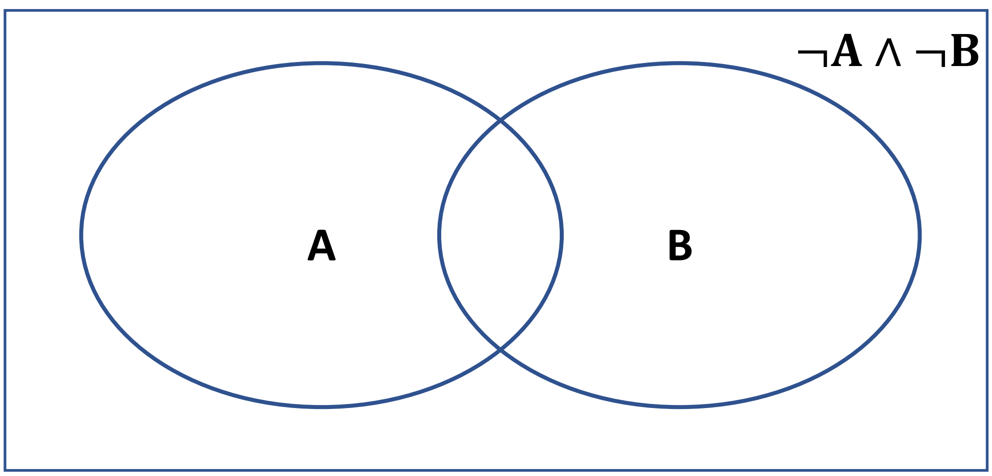
P(2|S) = 2/8

P(3|S) = 3/8

Entropy(S) = -3/8*log2(3/8) - 2/8*log2(2/8) - 3/8*log2(3/8) = 1.56127812446

My Comment Box
This question is undeniably easy enough for warming up. Timing: 2 minutes should be more than enough to read the whole of the exam sheet, and solve this problem.

(b) **(5 pts) True or False:** Let A and B be two independent events. Draw a Venn diagram to depict the relationship between A and B.



My Comment Box
This question has been covered in class on Bayesian learning. Furthermore, the time it takes to open a different app for drawing may take them a bit of time. Timing: 5 minutes

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

(c) **(5 pts)** Let A, B, and C be events. Are A and B always conditionally independent given C if P(B)=1? If yes, show your proof. If no, provide a counterexample.

Approach 1:

Show if P(B) = 1: P(A|BC) = P(A|C)

> LS: P(A|BC) = P(ABC) / P(BC)

= P(B|AC)*P(A|C)*P(C) / P(B|C)*P(C)

= P(A|C)

> RS: P(A|C)

Q.E.D

Approach 2:

Show if P(B) = 1:

P(AB|C) = P(A|C)*P(B|C)

> LS: P(AB|C) = P(ABC) / P(C)

= P(B|AC)*P(A|C)*P(C) / P(C)

= P(A|C)

> RS: P(A|C)*P(B|C) = P(A|C)

Q.E.D

My Comment Box
This is a wonderful test of base knowledge in conditional probability. <u>Timing:</u> 4 minutes for either approach

(d) **(20 pts)** Let A and B be two independent events with P(A) = 0.1 and P(B) = 0.4. Let C denote the event that at least one of A and B occurs, i.e., C = A OR B, and let D be the event that exactly one of A and B occurs, i.e., D = A XOR B.

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

i. **(3 pts)** Compute P(D). Show your work.

P(D) = P(A xor B)

= P(A) + P(B) - 2P(A and B)

= 0.42

My Comment Box
This is a very easy problem, but requires some analysis. Timing: 2 minutes

ii. **(3 pts)** Suppose A occurred. What is the probability that D occurred? Show your work.

P(D|A) = P(AD)/P(A)

= [P(A) – P(A)*P(B)] / P(A)

= (0.1 – 0.1*0.4)/0.1

= 0.6

My Comment Box
This is a very easy problem, but requires some analysis. People would struggle for a bit of time to find P(A|D) out of P(AD) before trying a different approach. Timing: 1-5 minutes (depending on how much students struggle with P(AD))

iii. **(3 pts)** Suppose C occurred. What is the probability that A occurred? Show your work.

P(C) = P(A or B) = P(A) + P(B) - P(A)*P(B) = 0.46

P(A|C) = P(AC)/P(C)

= [P(C) – P(notA)*P(B)] / P(C)

= 0.1/0.46

= 0.217

My Comment Box
This is a very easy problem, but requires some analysis. People would struggle for a bit of time to find P(C|A) out of P(AC) before trying a different approach. Timing: 1-5 minutes (depending on how much students struggle with P(AC))

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

iv. **(3 pts)** Suppose B and C both occurred. What is the probability that A occurred? Show your work.

P(A|BC) = P(ABC)/P(BC)

= [P(C) – P(A)*P(not B) – P(not A)*P(B)] / [P(C) – P(A)*P(not B)]

= 1 – [P(not A)*P(B)] / [P(C) – P(A)*P(not B)]

= 1 – [0.9*0.4] / [0.46 – 0.1*0.6]

= 0.1

My Comment Box
This is a very easy problem, but requires some analysis. People would struggle for a bit of time to find how to calculate P(ABC) and P(BC) like the above problems, before trying a different approach. Timing: 2-8 minutes (depending on how much students struggle with P(ABC) and P(BC))

v. **(3 pts)** Is A independent of D? Show your work.

P(DA) = P(A) – P(A)*P(B) = 0.1 – 0.1*0.4 = 0.06

P(D)*P(A) = 0.42*0.1 = 0.042

So they are not independent.

My Comment Box
This is a very easy problem. Timing: 2 minute

vi. **(5 pts)** Suppose C is true. Is whether A occurred independent of whether B occurred? Show your work.

We check if

P(AB|C) = P(A|C)*P(B|C)

> LS: P(AB|C) = P(ABC) /P(C) = P(A)*P(B) / P(C)

> RS: P(A|C)*P(B|C) = P(AC)*P(BC) / P(C) / P(C) = P(A)*P(B) / P(C) / P(C)

Since LS != RS, if C is true, A did not occur independently of B.

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

My Comment Box
This is a very wonderful problem about no general assumption about all independences leading to conditional independences. It's also not too challenging!
Timing: 8 minutes

Warming up and giving away free scores!
Maximum Total for section 1: 40 minutes

2. Decision Trees & Naive Bayes (45 points total)

The dataset below contains three input attributes, **A**, **B**, **C**, and one class attribute, **Y**.

A	B	C	Y
1	1	0	True
1	0	1	True
0	1	1	False
1	1	1	False
1	1	0	True
0	0	0	True
1	1	0	False
0	1	1	False
1	0	0	True
1	1	1	True

a) **(15 pts)** From the given data, if we build a decision tree using only one attribute, what attribute should we choose? Show your work by calculating IG(Y|A), IG(Y|B), and IG(Y|C), where IG is the information gain of Y when we split by A, B, and C respectively. Furthermore, show your tree.

P(Y=True) = 6/10;   P(Y=False) = 4/10

Entropy(Y) = 0.9710

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

What is IG(Y|A)?

P(Y=True|A=0) = 1/3;   P(Y=False|A=0) = 2/3

Entropy(Y|A=0) = 0.9183

P(Y=True|A=1) = 5/7;   P(Y=False|A=1) = 2/7

Entropy(Y|A=0) = 0.8631

IG(Y|A) = 0.0913

Similarly, for IG(Y|B):

P(Y=True|B=0) = 3/3;   P(Y=False|B=0) = 0/3

Entropy(Y|B=0) = 0

P(Y=True|B=1) = 3/7;   P(Y=False|B=1) = 4/7

Entropy(Y|B=1) = 0.9852

IG(Y|B) = 0.2814

Similarly, for IG(Y|C):

P(Y=True|C=0) = 4/5;   P(Y=False|C=0) = 1/5

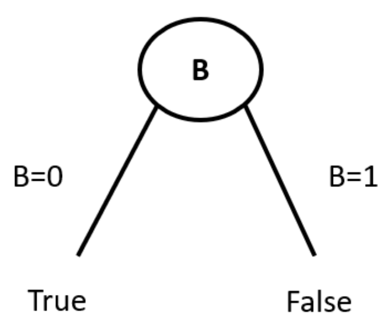
Entropy(Y|C=0) = 0.7219

P(Y=True|C=1) = 2/5;   P(Y=False|C=1) = 3/5

Entropy(Y|C=1) = 0.9710

IG(Y|C) = 0.1246

The tree ought to be as follows:



Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

My Comment Box
This is a short test of base knowledge in decision tree building. <u>Timing:</u> 15 minutes on current version (12 on calculating all 3 information gains, 3 minutes on drawing)

(b) **(10 pts)** According to the Naive Bayes classifier trained on the above dataset, how will the instance (A=1,B=0,C=1) be classified? Show your work.

Building the Naïve Bayes

Concept	**Result**
P(Y=True)	6/10
P(A=1|Y=True)	5/6
P(B=1|Y=True)	3/6
P(C=1|Y=True)	2/6
P(Y=False)	4/10
P(A=1|Y=False)	2/4
P(B=1|Y=False)	4/4
P(C=1|Y=False)	3/4

Classify

P(Y=False|A=1,B=0,C=1)

= P(A=1|Y=False)*(1 - P(B=1|Y=False))* P(C=1|Y=False)*P(Y=False)

= 0

P(Y=True|A=1,B=0,C=1)

= P(A=1|Y=True)*(1 - P(B=1|Y=True))* P(C=1|Y= True)*P(Y= True)

= 0.083

➔ **True**

My Comment Box
This is the perfect test for testing Naïve Bayes construction. <u>Timing:</u> 8 minutes

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

(c) **(10 pts)** Would it be possible to add just one instance to the dataset that would cause the Naive Bayes classifier to change its classification of the instance (A=1,B=0,C=1)? You should justify your answer by showing the relevant calculations.

We try to add (A=1,B=0,C=1) where Y=False and reconstruct the table.

Concept	Result
P(Y=True)	6/11
P(A=1|Y=True)	5/6
P(B=1|Y=True)	3/6
P(C=1|Y=True)	2/6
P(Y=False)	5/11
P(A=1|Y=False)	3/5
P(B=1|Y=False)	4/5
P(C=1|Y=False)	4/5

Classify

P(Y=False|A=1,B=0,C=1)

= P(A=1|Y=False)*(1 - P(B=1|Y=False))* P(C=1|Y=False)*P(Y=False)

= 0.043

P(Y=True|A=1,B=0,C=1)

= P(A=1|Y=True)*(1 - P(B=1|Y=True))* P(C=1|Y= True)*P(Y= True)

= 0.076

- ➔ No change, still **True**

My Comment Box
This is a very great test for testing Naïve Bayes understanding. This problem requires a small spark in people's creativity, and the tedious formulation can be made very simple because of what have been previously been done. Timing: 9 minutes

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

(d) **(10 pts)** In this question, we want to compare Decision Trees and Nave Bayes.

i. Are these methods unsupervised or supervised? Explain your answer.

Both are supervised, because they used labels.

ii. Which one is discriminative and which is generative? What do these terms mean.

Decision tree is discriminative, which is a category of approaches that use discriminating features to perform desired tasks (e.g. classification).

Naïve Bayes is generative, which is a category of approaches that estimates the distribution of related variables to perform desired tasks (e.g. classification).

iii. Which one do you prefer for a classification task? Explain your answer.

Answers may vary, but students must provide supporting reasons.

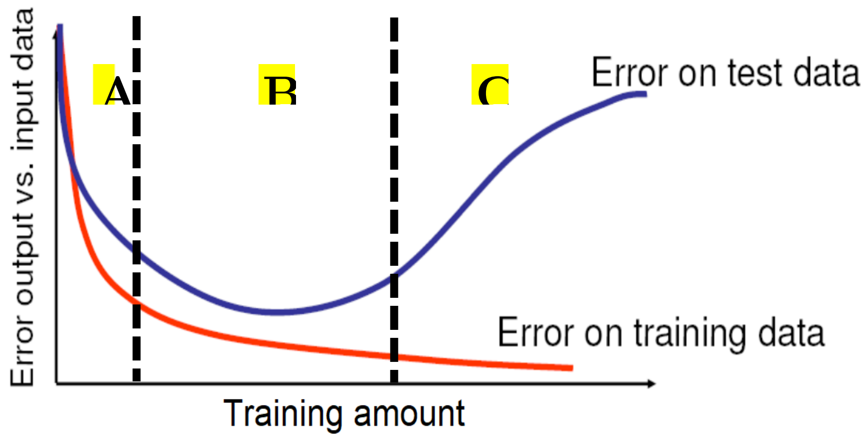
My Comment Box
This is a short test for overview of the course. Some people can copy and paste from the Internet. Subtask iii. is a giveaway, but it requires students to think about strengths and weaknesses of each approach. Timing: 5 minutes

This is the main section for testing students' skills in model building and overall understanding in what we covered about Decision Tree and Naïve Bayes. **Total for section 2:** 40 minutes

Maximum Total Esimated Time: 40 + 40 + 10 = 90 minutes

3. Open Questions (20 points total)

(a) **(6 pts)** Consider the figure below.



From the above figure, which region (i.e. **A**, **B**, or **C**) denotes underfitting, and which denotes overfitting? How can we tell?

Region A denotes underfitting because **both test error and training error are still high**.

Region C denotes overfitting becausue **training error is much lower than testing error**.

(b) **(7 pts)** When overfitting occurs with Decision Trees, we can use pruning. But why do we know that pruning helps us deal with overfitting?

Because pruning helps **reduce the biases** of a **complex tree trained on a finite dataset**, which **does not represent completeness**, and it can be **littered with outliers** and **inessential attributes**.

(c) **(7 pts)** Given a classification function f(a, b, c) = y, and two functions g and h, we know that the hypothesis a = g(b) and c = h(a, b) is true. After sampling the data from f, we use Nave Bayes to approximate its label y. Does Nave Bayes underfits f? Why?

Yes. Because **Naïve Bayes assumes independence of variables**, while **a depends on b**, and **c depends on both a and b**.

This section only requires fast thinking ~ a bit of in-depth understanding of what we have covered, but answers are short.

Total for section 3: 10 minutes