CS 307 Introduction to Machine Learning
Fall 2022
Midterm Exam I
October 31, 2022

Name:

Instructions:

- This test is open books and notes. There are three questions.
- Answer each question either in the space provided or on your own papers. If you make a mess, clearly indicate your answer. If we cannot read your answers, you will receive a score of zero.
- You do not need to show your work unless you are explicitly requested to. However, if you are asked to show your work, you will not receive any credit unless you do so.
- You have 95 minutes to complete the test *and* upload it to gradescope as a single PDF file. It is extremely important for you to make sure that your PDF includes **every** page you want to submit. Under no circumstances can we accept any pages that you forget/fail to include in the PDF after the test.
- Submit your test to gradescope. The submission folder will be closed at 11am. No submissions will be accepted after the submission folder is closed.

	Question	Score		Max score
1.	Short Answers	___	/	35
2.	Decision Trees and Naive Bayes	___	/	45
3.	Open Questions	___	/	20
	Total Score	___	/	100

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) =

   P(2|S) =

   P(3|S) =

   Entropy(S) =

   (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.

   (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.

(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.

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

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

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

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

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

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

2. **Decision Trees and 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.

(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.

(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.

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

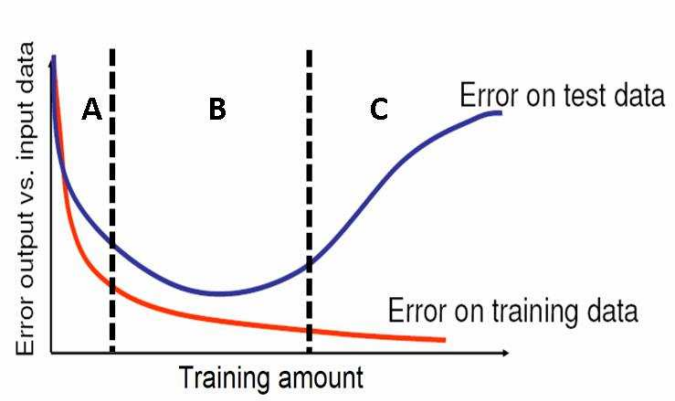
i. Are these methods unsupervised or supervised? Explain your answer.

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

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

3. **Open Questions (20 points total)**

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



Which region (i.e. A, B, or C) denotes underfitting, and which denotes overfitting? Explain your answer.

(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?

(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?