Final Part II - DSC 40A, Winter 2024

Instructions

- This is a 50-minute exam consisting of 4 questions worth a total of 40 points.
- The only allowed resource is your hand-written reference notes.
- No calculators.
- Please write neatly and stay within the provided boxes.
- You may fill out the **front page only** until you are instructed to start.

Statement of Academic Integrity

By submitting your exam, you are attesting to the following statement of academic integrity.

I will act with honesty and integrity during this exam.

Name:	Solutions
PID:	A12345678
Seat you are in:	

Version - A

Chess Pieces

A set of chess pieces has 32 pieces. 16 of these are black, and 16 of these are white. **In each color**, the 16 pieces are

- 8 pawns,
- 2 bishops,
- 2 knights,
- 2 rooks,
- 1 queen, and
- 1 king.

When there are multiple pieces of a given color and type (for example, 8 white pawns), we will assume they are **indistinguishable** from one another.

1. (12 points) Consider an experiment where each of n people selects one piece from their own set of 32 chess pieces, uniformly at random. The **result** of the experiment is a description of the **colors and types** of the pieces each person selected. For example, if $n = 3$, one possible result is:

- Person 1 selected a white knight.
- Person 2 selected a black queen.
- Person 3 selected a black pawn.

a) (4 points) How many results are possible for this experiment with n people?

◯ 2^n

◯ 6^n

● 12^n

◯ 16^n

◯ 32^n

◯ None of the above.

b) (4 points) As n become large, what fraction of n people has selected a black pawn? Choose the answer that's closest to the actual fraction.

◯ 1/32

◯ 1/12

● 1/4

◯ 1/2

◯ None of the above.

c) (4 points) True or False: Having two pawns is **independent** of having two white pieces.

● True

◯ False

2. (8 points) You run the k-means clustering algorithm on a dataset and it converges to a certain clustering with associated inertia I. You then duplicate each data point in the dataset and run k-means again on this twice-as-big dataset, with the same initial centroids as before. Which of the following is true? **Select all that apply.**

■ The centroids found will be the same as before.

☐ The inertia will be the same as before, I.

■ The inertia will be twice as much as before, $2I$.

☐ None of the above.

3. (8 points) With the help of the Alien from Bayesian Galaxy, Issac built a Little Hadron Collider in his garage to continue testing some fundamental principles of nature. The alien set up a fixed target at one end of the collider, and ask Issac to shoot quarks from the other end. Everytime the quark hit the target, the Alien will tell Issac whether a new hadron is formed or not. Due to the energy limitation in Issac's garage, he could only generate up, down, top, and bottom quarks (Strange and Charm quark would have consumed too much energy). For each quark Issac created, he has a record of:

- the type [up, down, top, bottom]
- the state [particle, antiparticle]
- the color charge [red, green, blue], and
- the outcome [hadron formed, hadron not formed]

Issac created 10 quarks, and these 10 quarks are recorded in the table below.

type	**state**	**color**	**Formed Hadron**
down	particle	green	formed
top	particle	red	formed
top	antiparticle	blue	formed
up	particle	red	formed
up	antiparticle	blue	formed
bottom	antiparticle	red	not formed
down	antiparticle	blue	not formed
down	particle	blue	not formed
top	particle	green	not formed
up	antiparticle	green	not formed

Since the alien is from Bayesian galaxy, they want Issac to develop a Naive Bayes classifier to predict whether he'll be successful in forming a hadron under the following quark conditions:

- the type is up
- the state is particle
- the color is blue

a) (4 points) Naive Bayes predicts that, given a up-particle-blue quark, the probability a hadron formed is k times the probability a hadron is not formed, for an integer value of k. What is k?

◯ 1 ◯ 2 ● 3 ◯ 4

b) (4 points) What would be the value of k if you change up quark in the previous collision to top quark, but keep everything else the same (i.e. top-particle-blue quark)?

◯ 1 ◯ 2 ● 3 ◯ 4

4. (12 points) In hospital, there are 10 patients who have recently been exposed to COVID. The hospital purchased a new COVID testing machine. The machine could comprehensively analyze multiple body measurements to produce a COVID score for each patient, ranging from 1 to 10. According to the instruction manual of the machine, a higher COVID score means the patient is more likely to be infected by COVID, while a lower COVID score means the patient is less likely to be infected. The COVID score of each patient is listed below:

Patient Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
COVID Score	7	9	9	3	5	4	6	1	8	2

Suppose Patient 1, Patient 2, Patient 3, and Patient 4 are actually infected by COVID, while all other patients are not infected.

a) (4 points) Based on the patient's COVID score, Dr. Issac uses a threshold of 2.5 to diagnose COVID. That means, he diagnose all patients with COVID Score above 2.5 to be COVID-infected, and all patient with Diagnostic Score below 2.5 to be non-infected. What is the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) of Dr. Issac's diagnose? Each block should be filled with an integer.

TP = 4 TN = 2 FP = 4 FN = 0

b) (4 points) Dr. Albert is somewhat skeptical about technology, leading him to adopt a cautious approach in diagnosing patients. Initially, he sets a diagnostic threshold at 5.5. Patients scoring above this threshold are initially diagnosed as positive. However, Dr. Albert conducts a secondary diagnosis for all patients who received a positive diagnosis initially. During this secondary diagnosis, he reclassifies Patient 9's positive diagnoses as negative. What is the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) of Dr. Albert's diagnose? Each block should be filled with an integer.

TP = 3 TN = 5 FP = 1 FN = 1

c) (4 points) Which Doctor's diagnoses is better?

● Albert ○ Issac ○ They are equally good