DSC 40A - Homework 3

Due: Tuesday, August 1 at 11:59pm

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Homeworks are due to Gradescope by 11:59pm on the due date. You can use a slip day to extend the deadline by 24 hours.

Homework will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should **always explain and justify** your conclusions, using sound reasoning. Your goal should be to convince the reader of your assertions. If a question does not require explanation, it will be explicitly stated.

Homeworks should be written up and turned in by each student individually. You may talk to other students in the class about the problems and discuss solution strategies, but you should not share any written communication and you should not check answers with classmates. You can tell someone how to do a homework problem, but you cannot show them how to do it.

For each problem you submit, you should **cite your sources** by including a list of names of other students with whom you discussed the problem. Instructors do not need to be cited.

This homework will be graded out of 50 points. The point value of each problem or sub-problem is indicated by the number of avocados shown.

Notes:

- Please **assign pages to questions** when you upload your submission to Gradescope.

Problem 1. k-Means Clustering

For parts (a), (b), and (c) of this question, we'll use the six data points given below, $\vec{x_1}$ through $\vec{x_6}$.

$$\vec{x_1} = \begin{bmatrix} 9 \\ 50 \end{bmatrix}, \vec{x_2} = \begin{bmatrix} 28 \\ 25 \end{bmatrix}, \vec{x_3} = \begin{bmatrix} 3 \\ 55 \end{bmatrix}, \vec{x_4} = \begin{bmatrix} 30 \\ 20 \end{bmatrix}, \vec{x_5} = \begin{bmatrix} 6 \\ 58 \end{bmatrix}, \vec{x_6} = \begin{bmatrix} 33 \\ 22 \end{bmatrix}$$

Just by looking at the data, you should be able to roughly identify two clusters. Let's see how k-means clustering finds these clusters algorithmically.

a) Using $\vec{x_1}$ and $\vec{x_2}$ as initial centroids, trace through one iteration of the k-means algorithm by hand. What are the two centroids and what are the two clusters found after this first iteration?

Problem 2. Probability Rules for Three Events

a) The multiplication rule for two events says

$$P(A \cap B) = P(A) \cdot P(B|A)$$

Use the multiplication rule for two events to prove the multiplication rule for three events:

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|(A \cap B))$$

Hint: You can think of $A \cap B \cap C$ as $(A \cap B) \cap C$.

b) Suppose E, F, and G are events. Explain in words why

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G).$$

Intuitively, the relationship between $\cap$ and $\cup$ is similar to the relationship between multiplication and addition; if e, f, g are numbers, then $(e + f) \cdot g = e \cdot g + f \cdot g$ as well.

c) The general addition rule for any two events says:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Use the general addition rule for two events to prove the general addition rule for three events:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Hint: You will need to use the result of part (b).

d) A survey was administered to 500 Formula One (F1) Racing fans asking about their predictions for the 2023 F1 season. Each respondent named 3 drivers that they predicted would finish in the top 3. The survey revealed the following information:

- 20 respondents' predictions did not include any of Max Verstappen, Charles Leclerc, and Sergio Perez.
- 350 responses included Max Verstappen.
- Of the 350 respondents who said Max Verstappen, 240 also said Sergio Perez.
- Of the 350 respondents who said Max Verstappen, 150 also said Charles Leclerc
- 300 respondents said Sergio Perez.
- Of the 300 respondents who said Sergio Perez, 140 also said Charles Leclerc.
- 90 respondents predicted all three of Max Verstappen, Charles Leclerc, and Sergio Perez.

Suppose we randomly select one survey participant. What is the probability that they predicted that Charles Leclerc would be among the top 3 this year?

Problem 3. Stringle

In this problem, we will look at a made-up game called Stringle. Each day, a random six-letter string is chosen, and players have to try to guess what it is.

In Stringle, any six-letter string of uppercase letters is allowed, as long as it does not have any repeated letters. The string does not have to make sense as an English word. For example, the string of the day might be ZVODUP. Any valid string is equally likely to be chosen each day.

a) Consider A, E, I, O, U, and Y to be vowels. What is the probability that today's Stringle string and yesterday's Stringle string both start with a vowel?

b) What is the probability that today's Stringle string or yesterday's Stringle string starts with a vowel?

c) What is the probability that today's Stringle string includes no vowels?

d) What is the probability that today's Stringle string includes all vowels?

e) What is the probability that today's Stringle string includes the letter J?

f) What is the probability that today's Stringle string is exactly the same as yesterday's Stringle string?

Problem 4. Independence and Conditional Independence

Consider the sample space $S = \{a, b, c, d, e, f, g\}$ with associated probabilities given in the table below.

outcome	a	b	c	d	e	f	g
probability	$\frac{5}{21}$	$\frac{2}{21}$	$\frac{1}{21}$	$\frac{4}{21}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{3}{21}$

Let $X = \{d, e\}$ and $Y = \{e, f\}$. Remember to show your work for all calculations.

a) Are X and Y independent?

b) Determine if X and Y are conditionally independent given each of the following events Z.

1. $Z = \{a, b, d, e, f, g\}$
2. $Z = \{a, d, e, f, g\}$
3. $Z = \{d, e, f, g\}$

Problem 5. Independence and Complements

Let E and F be two events in a sample space S, with $0 < P(F) < 1$.

a) If $P(E|F) = P(E|\overline{F})$, must it be true that E and F are independent? Provide a proof of independence, or give a counterexample by specifying a sample space S and two dependent events E and F that satisfy the given conditions.

b) If $P(E|F) = P(\overline{E}|F)$, must it be true that E and F are independent? Provide a proof of independence, or give a counterexample by specifying a sample space S and two dependent events E and F that satisfy the given conditions.

Problem 6. Probability Theory

Let S be a sample space, and let A, E_1, E_2, E_3 be events in that sample space. Suppose that $E_1 \cap E_2$, $E_1 \cap E_3$, and $E_2 \cap E_3$ are all empty. Given the following probabilities, find $P(E_2|A)$:

$$\begin{aligned} P(E_1) &= 1/6 \quad & P(A|E_1) &= 1/9 \\ P(E_2) &= 1/3 \quad & P(A|E_2) &= 1/7 \\ P(E_3) &= 1/2 \quad & P(A|E_3) &= 1/5 \end{aligned}$$

Problem 7. Double Deck of Cards

A standard deck of cards contains 52 cards. There are 13 cards in each of 4 suits (hearts $\heartsuit$, spades $\spadesuit$, diamonds $\diamondsuit$, and clubs $\clubsuit$.) Within a suit, the 13 cards each have a different rank. In ascending order, these ranks are 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace.

You are playing a four-player card game using two regular decks of cards. Each player will be dealt 26 cards as follows:

- The first deck of cards will be randomly shuffled and dealt out, 13 cards to each of the four players.
- Then the second deck of cards will be randomly shuffled and dealt out, 13 cards to each of the four players.

a) Let Event A = 'you are dealt two Kings of Hearts' and Event B = 'some other player has a pair of Kings of Hearts'. Are these two events independent of each other?

b) Let Event A = 'you are dealt two Kings of Hearts' and Event C = 'some other player has a pair of Aces of Hearts'. Are these two events independent of each other?

c) Now suppose that you know you do not have any Aces of Hearts. Let Event A = 'you are dealt two Kings of Hearts' and Event C = 'some other player has a pair of Aces of Hearts'. Now that you know you don't have any Aces of Hearts, are these two events independent of each other?

d) Let Event A = 'you are dealt two Kings of Hearts' , Event C = 'some other player has a pair of Aces of Hearts', and Event D = 'you do not have any Aces of Hearts'. State the result of parts (b) and (c) in words, in terms of independence and conditional independence.