



AIDI 1000 – AI Algorithms & Mathematics Assignment #1 (75 points)

This assignment is to be completed at home and is
due: Wednesday, Oct 2, 11:59pm

Plagiarism is the presentation of someone else's work as your own. Submitting work that is completed or directly influenced by a third party or information copied from the internet is considered plagiarism. Plagiarism will result in an academic misconduct which will result in a zero in the work completed or a zero in the course.

General Instructions

- You have the flexibility to solve this assignment either manually or using Python programming language. You cannot use both methods.
- If you decide to solve this assignment manually, submit a pdf file containing the solution to all the questions. You must write down all the necessary steps how you reach at the final answer. If you do not follow this rule, it might affect your scores. Your handwriting should be clear and clean.
- If you decide to solve this assignment using Python, submit a jupyter notebook (.ipynb file) containing the solution to all the questions. You must write down all the necessary steps how you reach at the final answer. You can make use of docstrings/comments and markdown cells for this purpose. If you do not follow this rule, it might affect your scores. Your code should be clear and readable. Use different cells for different questions. Do not solve multiple questions in one cell.
- Submit your solution file in blackboard by the due date/time.
- Submit only one solution file, no multiple files are allowed.
- Late submissions are not accepted and will result in a grade of zero.
- Submissions through emails are not accepted and will result in a grade of zero.
- Submission of a wrong file will result in a grade of zero.
- Plan to submit your assignment earlier, do not wait for last minute to submit it as unexpected technical issues may hinder your submission and you can lose your marks.
- Submission pattern: FirstName-LastName-Assignment1.pdf
or FirstName-LastName-Assignment1.ipynb

Q1: Recent crime statistics collected over the past year reveal the following figures. 81% of all people arrested were male, 15% of all people arrested were under the age of 18, 8% of all people arrested were females and under the age of 18.

Let M represent the event that the person arrested is male, and $U18$ represents the event that the person arrested is under the age of 18.

Part (a): Complete the table below. Use two decimals in each of your answers. **(9 points)**

	M	M^c	Row Probabilities
$U18$			
$U18^c$			
Column Probabilities			

Part (b): A person who was arrested in the past year is randomly chosen. What is the probability that this person is male or under the age of 18? Use two decimals in your answer. **(2 points)**

Part (c): Find the probability that the person chosen in Part (b) is neither male nor under the age of 18. Use two decimals in your answer. **(1 points)**

Part (d): What is the probability that all people arrested in the past year are male and at least 18 years of age? Use two decimals in your answer. **(1 points)**

Part (e): Are the events M and $U18$ mutually exclusive? Select the most appropriate reason below. **(2 points)**

A: M and $U18$ are not mutually exclusive events, because $P(M \cap U18) \neq P(M) \cdot P(U18)$

B: M and $U18$ are mutually exclusive events, because $P(M \cup U18) = 0$

C: M and $U18$ are mutually exclusive events, because $P(M \cap U18) = P(M) \cdot P(U18)$

D: M and $U18$ are mutually exclusive events, because $P(M \cap U18) = 0$

E: M and $U18$ are not mutually exclusive events, because $P(M \cap U18) \neq 0$

Q2: Suppose that 40% of the voters in a city are in favor of a ban of smoking in public buildings. Suppose 5 voters are to be randomly sampled. Find the probability that:

- (a) 2 favor the ban. **(2 points)**
- (b) Less than 4 favor the ban. **(4 points)**
- (c) At least 1 favor the ban. **(4 points)**

Q3: A recent survey of employed Canadians found that 40%, or 4-in-10, would find it difficult to meet their financial obligations if their paycheque was delayed by one-week. You are to randomly select two employed Canadians. Compute the probability that:

- (a) Both would find it difficult to meet their financial obligations if their paycheque was delayed by one-week. **(2 points)**
- (b) Neither would find it difficult to meet their financial obligations if their paycheque was delayed by one-week. **(2 points)**
- (c) At least one of the two would find it difficult to meet their financial obligations if their paycheque was delayed by one-week. **(2 points)**
- (d) Suppose you are to randomly pick n -employed Canadians in such a way that the probability of at least one of them will not be able to meet their financial obligations if their paycheque is delayed by one -week is 0.95. Compute the minimum number of employed Canadians you would have to randomly select. In other words, compute the sample size n . **(4 points)**

Q4: Most graduate schools of business require applicants for admission to take the SAT examination. Scores on the SAT are roughly normally distributed with a mean of 530 and a standard deviation of 110. What is the probability of an individual scoring above 500 on the SAT? **(5 points)**

Q5: The Edwards's Theater chain has studied its movie customers to determine how much money they spend on concessions. The study revealed that the spending distribution is approximately normally distributed with a mean of 4.11 dollar and a standard deviation of 1.37 dollar. What percentage of customers will spend less than 3.00 dollar on concessions? **(5 points)**

Q6: There are three types of coins which have different probabilities of landing heads when tossed:

- Type A coins are fair, with probability 0.5 of heads
- Type B coins are bent and have probability 0.6 of heads
- Type C coins are bent and have probability 0.9 of heads

Suppose I have a drawer containing 5 coins: 2 of type A, 2 of type B, and 1 of type C. I reach into the drawer and pick a coin at random. Without showing you the coin, I flip it once and get heads (i.e. event D). What is the probability it is:

type A (i.e. $P(H = A|D)$)? Type B (i.e. $P(H = B|D)$)? Type C (i.e. $P(H = C|D)$)?

Fill out table below with your answers.

(15 points)

hypothesis	prior	likelihood	posterior
H	$P(H)$	$P(D H)$	$P(H D)$
A			
B			
C			

Q7: Consider the following dataset of four rows and three features (Malicious, Viagara, Meet) with class labels (ham and spam). Suppose we see a message having these features $M_5 = (\text{Malicious} = \text{'yes'}, \text{Viagara} = \text{'no'}, \text{Meet} = \text{'yes'})$, What is the probability that it is a spam or ham? **(15 points)**

S.No	Malicious	Viagara	Meet	class
M_1	yes	yes	yes	spam
M_2	no	no	yes	ham
M_3	yes	no	yes	spam
M_4	no	yes	no	ham