به نام خدا

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

دانشکده برق

یادگیری ماشین – نیم‌سال دوم ۱۴۰۱–۱۴۰۲

تمرین تئوری احتمال و قانون بیز، سری اول، درس یادگیری ماشین

Collaboration Policy

You are to complete this assignment individually. However, you may discuss the general algorithms and ideas with classmates, TAs, peer mentors and instructor in order to help you answer the questions. But we require you to:

- Not explicitly tell each other the answers.
- Not to copy answers or code fragments from anyone or anywhere.
- Not to allow your answers to be copied.
- Not to get any code from the Web.

If you have any questions regarding this assignment, please contact Mr. Alireza Rahmati.

Telegram ID: AliReza_AR

Submit by 15th of Farvardin, 1402. 11.59pm.

Question 1: [5 points]

In a country, 51% of adults are females. One adult is randomly selected for a survey involving credit card usage.

A) Find the prior probability that the selected person is a male.

B) It is later learned that the selected survey subject was a cigar smoker. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars. Use this additional information to find the probability that the selected subject is a male.

Question 2: [5 points]

An aircraft emergency locator transmitter (ELT) is a device designed to transmit a signal in the case of a crash. The Altigauge Manufacturing Company makes 80% of the ELTs, the Bryant Company makes 15% of them, and the Chartair Company makes the other 5%. The ELTs made by Altigauge have a 4% rate of defects, the Bryant ELTs have a 6%rate of defects, and the Chartair ELTs have a 9% rate of defects (which helps to explain why Chartair has the lowest market share).

A) If an ELT is randomly selected from the general population of all ELTs, find the probability that it was made by the Altigauge Manufacturing Company.

B) If a randomly selected ELT is then tested and is found to be defective, find the probability that it was made by the Altigauge Manufacturing Company.

Question 3: [5 points]

A table has been collected about poisoning the school's students.

	The student claimed to be sick	The student claimed not to be sick
Positive test	80	5
Negative test	3	11

A) If one of the 99 test subjects is randomly selected, what is the probability of getting a subject who is sick?

B) A test subject is randomly selected and is given a disease test. What is the probability of getting a subject who is not sick, given that the test result is negative?

Question 4: [20 points]

For each of the following functions $F_i(c)$, state whether or not $F_i(c)$ is the CDF. If not, state which of the properties of a CDF it violates. If so, find the corresponding PMF or PDF.

A)

$$F_1(c) = \begin{cases} 0 & c \leq 0 \\ \frac{1}{3}c & 0 < c \leq 3 \\ 0.005c^2 & 3 < c \leq 10 \\ 0 & 10 < c \end{cases}$$

B)

$$F_2(c) = \begin{cases} 0 & c \leq 2 \\ 0.2c & 2 < c \leq 3 \\ 0.05c & 3 < c \leq 5 \\ 0.01 & 5 < c \end{cases}$$

C)

$$F_3(c) = \begin{cases} 0 & c \leq 1 \\ 0.2c & 0 < c \leq 3 \\ 0.05c & 3 < c \leq 10 \\ 0 & 10 < c \end{cases}$$

Question 5: [15 points]

Random variable x is a distributed with the following PDF:

$$f_x(u) = \begin{cases} \cos(x) & 0 \leq x \leq A \\ 0 & O.W. \end{cases}$$

A) What is the value of the constant A?

B) What is the corresponding CDF related to $F_x(c)$?

C) What is $E(x)$?

D) What is $Var(x)$?

Question 6: [5 points]

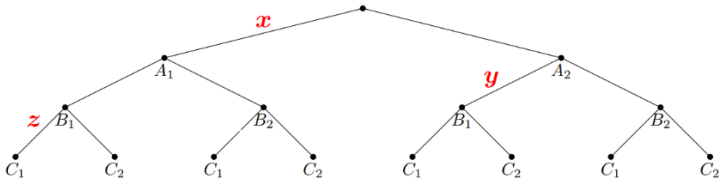
There are 7 ball (5 red and 2 green) in a box. A random ball is selected and then it is replaced by a ball of another color, then a second ball is drawn.

A) What is the probability that the first ball is green?

B) What is the probability that the second ball is green while the first selected ball was red?

Question 7: [5 points]

Find the probability of each variable using the tree below.



A) What probability does the x variable represent? $(P(x) = ?)$

B) What probability does the y variable represent? $(P(y) = ?)$

C) What probability does the z variable represent? $(P(z) = ?)$

Question 8: [15 points]

According to Amirkabir University of technology survey, the probability that a student will graduate within 5 years is 0.8. The score in ML (machine learning), which taken values in [0,20], of the students who are graduated (G) within 5 years follows a normal distributed with mean of 17 and standard deviation of 1 and for students who are not graduated (nG) within 5 years follows a normal distribution with mean 13 and standard deviation 2.3. Moreover, let x denoted the score of a student in ML exam.

A) What is the probability that a student will graduate within 5 years and given 18 score in the ML exam? $(P(G|x = 18) = ?)$

B) What is the minimum score of ML exam that a student must get to be graduate? $(P(G|x) \geq P(nG|x))$

Question 9: [25 points]

According to the previous question, suppose you have another metric y and it is their last high school's class score, which taken values in [0,20]. Assume that the probability density function $P(x, y|G)$ and $P(x, y|nG)$ are bi-variate Gaussian.

$$p(x, y|G) \sim \mathcal{N}(\mu_G, \Sigma_G)$$

$$p(x, y|nG) \sim \mathcal{N}(\mu_{nG}, \Sigma_{nG})$$

with:

$$\mu_G = \begin{bmatrix} 17 \\ 16 \end{bmatrix}, \quad \Sigma_G = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\mu_{nG} = \begin{bmatrix} 13 \\ 15 \end{bmatrix}, \quad \Sigma_{nG} = \begin{bmatrix} 4 & 1 \\ 1 & 5 \end{bmatrix}$$

A) What is the probability that a student graduates within 5 years and their ML score are 18.5?

B) Determine the optimal decision boundary between the two classes G and nG (your variable must be x and y).

What to submit:

• Answer the questions completely.

Good luck 😊