UNIVERSITY OF ALBERTA
CMPUT 267 Fall 2024

Midterm Exam 1

Do Not Distribute

Duration: 75 minutes

Last Name: ______________________________

First Name: ______________________________

Carefully read all of the instructions and questions. Good luck!

---

1. **Do not turn this page** until instructed to begin.
2. Ensure that your exam package contains 9 pages.
3. **Only the scantron will be marked**. All of your answers must be clearly marked on the scantron.
4. Use **pencil only** to fill out the scantron (preferably an HB or #2 pencil).
5. **Erase mistakes completely** on the scantron to avoid misreading by the scanner.
6. **Mark answers firmly and darkly**, filling in the bubbles completely.
7. This exam consists of **25 questions**. Each question is worth **1 mark**. The exam is worth a total of **25 marks**.
8. Some questions may have **multiple correct answers**. To receive **full marks**, you must select **all correct answers**. If you select only **some** of the correct answers, you will receive **partial marks**. Selecting an incorrect option will cancel out a correct one. For example, if you select two answers—one correct and one incorrect—you will receive zero points for that question. If the number of incorrect answers exceeds the correct ones, your score for that question will be zero. **No negative marks** will be given.

Question 1. [1 MARK]

Which of the following represents the set of all linear functions of the form $f(x) = wx + b$ where $w, b \in \mathbb{R}$?

A. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = wx + b,\ w, b \in \mathbb{R}\}$

B. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = w^2b^2x,\ w, b \in \mathbb{R}\}$

C. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = wx,\ w \in \mathbb{R}\}$

D. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = b,\ b \in \mathbb{R}\}$

Question 2. [1 MARK]

Which of the following is a valid definition of a function $\mathcal{A} : (\mathbb{R} \times \mathbb{R})^2 \to \{f \mid f : \mathbb{R} \to \mathbb{R}\}$?

A. $\mathcal{A}((a, b), (c, d)) = f$, where $f(x) = ax + b$

B. $\mathcal{A}((a, b), (c, d)) = f$, where $f(x) = cx + d$

C. $\mathcal{A}((a, b), (c, d)) = f$, where $f(x) = a + bx$

D. $\mathcal{A}((a, b), (c, d)) = f$, where $f(x) = ax - b$

Question 3. [1 MARK]

Given $\mathcal{D} = ((x_1, y_1), \ldots, (x_n, y_n))$, the loss function $\ell(\hat{y}, y) = |\hat{y} - y|$, and the predictor $f(x) = wx + b$, which of the following represents the total loss calculated by summing $\ell(f(x), y)$ over all data points $(x, y) \in \mathcal{D}$?

A. $\sum_{i=1}^{n} |wx_i + b - y_i|$

B. $\sum_{(x,y)\in\mathcal{D}} |wx + b - y|$

C. $\sum_{i=1}^{n} (wx_i + b - y_i)^2$

D. $\prod_{(x,y)\in\mathcal{D}} |wx + b - y|$

Question 4. [1 MARK]

Let $g(x, y) = x^2 + y$ where $x, y \in \mathbb{R}$. What is

$$\int_{\mathcal{X}} \sum_{y\in\mathcal{Y}} g(x, y)\, dx$$

where $\mathcal{Y} = \{1, 2, 3\}$ and $\mathcal{X} = [0, 2]$?

A. 15

B. 18

C. 20

D. 22

Question 5. [1 MARK]

Let $h(x, y) = x + y$ where $x, y \in \mathbb{R}$. What is

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} h(x, y)\, dy\, dx$$

where $\mathcal{Y} = [2, 4]$ and $\mathcal{X} = [1, 3]$?

A. 18

B. 20

C. 22

D. 24

Question 6. [1 MARK]

Let X be a discrete random variable uniformly distributed over the outcome space $\mathcal{X} = \{1, 2, 3, 4, 5\}$. The probability mass function (pmf) of X is given by $p(x) = \frac{1}{5}$ for each $x \in \mathcal{X}$. Which of the following statements are true?

A. The probability that X is less than 4 is 0.5.

B. The expected value $\mathbb{E}[X] = 3$.

C. The probability that X is an even number is 0.4.

D. The probability that X is greater than or equal to 4 is 0.4.

Question 7. [1 MARK]

Suppose you roll two fair six-sided dice. Let the first die be represented by the random variable $D_1 \in \{1, 2, 3, 4, 5, 6\}$ and the second die by $D_2 \in \{1, 2, 3, 4, 5, 6\}$. Which of the following sets represents the outcome space of the random variable $D = (D_1, D_2)$?

A. $\{(d_1, d_2) \mid d_1, d_2 \in \{1, 2, 3, 4, 5, 6\}\}$

B. $\{2, 3, 4, \ldots, 12\}$

C. $\{1, 2, 3, 4, 5, 6\}$

D. $\{d_1 + d_2 \mid d_1, d_2 \in \{1, 2, 3, 4, 5, 6\}\}$

Question 8. [1 MARK]

Suppose you randomly select a letter from the English alphabet. The outcome space is $\mathcal{X} = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \ldots, \mathrm{Z}\}$. Which of the following is an event?

A. $\{\mathrm{A}, \mathrm{E}, \mathrm{I}, \mathrm{O}, \mathrm{U}\}$

B. $\{\mathrm{AA}, \mathrm{BB}\}$

C. M

D. $\{\mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{F}, \mathrm{G}\}$

Question 9. [1 MARK]

Suppose you have a random variable X representing the time (in minutes) it takes to commute to work. You know X is distributed according to the continuous uniform distribution over the interval $\mathcal{X} = [30, 60]$. Let p be the pdf of X. Which of the following statements are true?

A. $p(45)$ is the probability that $X = 45$.

B. $p(x)$ represents the probability density at point x.

C. The probability that X is between 40 and 50 is $\frac{1}{3}$.

D. The probability that $X = 30$ is $\frac{1}{30}$.

Question 10. [1 MARK]

Suppose that $Y \in \mathbb{R}$ is distributed according to Normal$(0, 1)$. Which of the following statements are true?

A. The probability density function of Y is $p(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ for $y \in \mathbb{R}$.

B. The probability that $Y = 0$ is $\frac{1}{\sqrt{2\pi}}$.

C. Y is a continuous random variable.

D. The variance of Y is 1.

Question 11. [1 MARK]

Suppose you have two discrete random variables $X \in \{0, 1, 2\}$ and $Y \in \{1, 2\}$. The joint probability mass function (pmf) of X and Y is given by the following values:

$$p(0,1) = \frac{1}{12}, \quad p(0,2) = \frac{1}{12}$$
$$p(1,1) = \frac{1}{4}, \quad p(1,2) = \frac{1}{6}$$
$$p(2,1) = \frac{1}{6}, \quad p(2,2) = \frac{1}{4}$$

Which of the following is the marginal pmf of X?

A. $p_X(0) = \frac{1}{6}, \quad p_X(1) = \frac{5}{12}, \quad p_X(2) = \frac{5}{12}$

B. $p_X(0) = \frac{1}{6}, \quad p_X(1) = \frac{1}{3}, \quad p_X(2) = \frac{1}{2}$

C. $p_X(0) = \frac{1}{3}, \quad p_X(1) = \frac{1}{2}, \quad p_X(2) = \frac{1}{6}$

D. $p_X(0) = \frac{1}{2}, \quad p_X(1) = \frac{1}{4}, \quad p_X(2) = \frac{1}{4}$

Question 12. [1 MARK]

Let X_1, X_2, X_3 be independent and identically distributed random variables representing three coin flips, where each coin lands heads (1) with probability 0.6 and tails (0) with probability 0.4. What is the probability that all three flips result in heads or all three flips result in tails?

A. 0.216

B. 0.064

C. 0.28

D. 0.512

Question 13. [1 MARK]

Let X be a random variable taking values in $\mathbb{R}$, and let $Y = 2X + 3$. Which of the following statements are true?

A. Y is a random variable.

B. The outcome space of Y is $\mathbb{R}$.

C. Y is not a random variable.

D. The outcome space of Y is $[3, \infty)$.

Question 14. [1 MARK]

Suppose X_1, X_2, X_3 are independent random variables, each with $X_i \sim \mathcal{N}(5, 4)$. Let $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$. Which of the following statements are true?

A. The expected value $\mathbb{E}[\bar{X}] = 5$.

B. The variance $\text{Var}(\bar{X}) = \frac{4}{3}$.

C. The variance $\text{Var}(X_1) = 4$.

D. The expected value $\mathbb{E}[\bar{X}] = \frac{5}{3}$.

Question 15. [1 MARK]

Let X be a random variable with outcome space $\mathcal{X} = \{a, b, c\}$ and pmf $p(a) = 0.2, \quad p(b) = 0.3, \quad p(c) = 0.5$. The function $f(x)$ is given by:

$$f(a) = 4, \quad f(b) = 6, \quad f(c) = 2$$

What is $\mathbb{E}[f(X)]$?

A. 3.0

B. 3.6

C. 4.0

D. 2.4

Question 16. [1 MARK]

Let X represent the outcome of a biased coin flip, where $\mathbb{P}(X = 1) = 0.8$ (heads) and $\mathbb{P}(X = 0) = 0.2$ (tails). Given that $X = 1$, the conditional distribution of a random variable $N \in \{1, 2, 3\}$ is:

$$\mathbb{P}_{N|X}(N = 1|X = 1) = \frac{1}{5}, \quad \mathbb{P}_{N|X}(N = 2|X = 1) = \frac{2}{5}, \quad \mathbb{P}_{N|X}(N = 3|X = 1) = \frac{2}{5}.$$

What is the conditional expectation $\mathbb{E}[N \mid X = 1]$?

A. 2.0

B. 2.2

C. 1.8

D. 2.4

Question 17. [1 MARK]

Suppose you are tasked with predicting the price of a car based on the following information: the engine size in liters, the number of doors, and the year it was manufactured. Which of the following options correctly specifies the features and the label for this problem?

A. Features: engine size, number of doors; Label: year manufactured

B. Features: year manufactured, price of the car; Label: engine size

C. Features: engine size, number of doors, year manufactured; Label: price of the car

D. Features: price of the car, engine size; Label: number of doors

Question 18. [1 MARK]

Suppose you have a dataset where each instance represents a smartphone. The features are: screen size (in inches), battery capacity (in mAh), and RAM size (in GB). The label is the brand of the smartphone (e.g., Apple iPhone, Samsung Galaxy, Google Pixel). Which of the following statements are true?

A. The features can be represented as an element of $\mathbb{R}^3$.

B. This is a regression problem.

C. The label set is a finite set.

D. This is a classification problem.

Question 19. [1 MARK]

Suppose you are working with a dataset where the features $X \in \mathbb{R}$ and the labels $Y \in \{0, 1\}$. You are using a predictor $f(X)$ which outputs predicted labels. The loss function is defined as:

$$\ell(f(X), Y) = \begin{cases} 1 & \text{if } f(X) \neq Y \\ 0 & \text{if } f(X) = Y \end{cases}$$

Which of the following expressions correctly represent the expected loss?

A. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} \sum_{y \in \{0,1\}} \ell(f(x), y)\, p(x, y)\, dx$

B. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} \ell(f(x), y)\, p(x)\, dx$

C. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} \sum_{y \in \{0,1\}} \ell(f(x), y)\, p(y \mid x)\, p(x)\, dx$

D. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} \ell(f(x), y)\, dx$

Question 20. [1 MARK]

Suppose you roll a six-sided die 4 times and observe the following outcomes: $X_1 = 2$, $X_2 = 5$, $X_3 = 3$, $X_4 = 6$.
What is the sample mean of these 4 rolls?

A. 3.5

B. 5.0

C. 4.0

D. 4.5

Question 21. [1 MARK]

Suppose you are given a predictor $f(x) = 5 + 0.5x$, which models the relationship between the number of years of experience x and the salary y (in tens of thousands of dollars) of an employee. You are provided with the following dataset of 4 (x, y) pairs:

$$\mathcal{D} = ((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)) = ((1, 5.5),\ (2, 6.0),\ (3, 6.0),\ (4, 7.5)).$$

Let the loss function be the squared loss $\ell(f(x), y) = (f(x) - y)^2$.
Calculate $\hat{L}(f) = \frac{1}{4}\sum_{i=1}^{4} \ell(f(x_i), y_i)$ (which is an estimate of the expected squared loss $L(f) = \mathbb{E}[\ell(f(x), y)]$).

A. 0.25

B. 0.5

C. 0.0625

D. 0.125

Question 22. [1 MARK]

Consider the function $g(w) = e^{3w} + 2w^2$, where $w \in \mathbb{R}$. What is the second derivative $g''(w)$?

A. $g''(w) = 3e^{3w} + 4w$

B. $g''(w) = 9e^{3w} + 4$

C. $g''(w) = e^{3w} + 4$

D. $g''(w) = 9e^{3w} + 2$

Question 23. [1 MARK]

Consider the convex function $g(w) = 3w^2 - 12w + 7$, where $w \in \mathbb{R}$.
What is the minimum value of $g(w)$ and at what w is it achieved?

A. The minimum value is 1 at $w = 2$.

B. The minimum value is -5 at $w = 2$.

C. The minimum value is -5 at $w = -2$.

D. The minimum value is 7 at $w = 0$.

Question 24. [1 MARK]

Consider the convex function $g(w) = \sum_{i=1}^{4}(w - x_i)^2$, where $x_1 = 1$, $x_2 = 3$, $x_3 = 5$, $x_4 = 7$, and $w \in \mathbb{R}$. What value of w minimizes $g(w)$?

A. $w = 5$

B. $w = 3$

C. $w = 4$

D. $w = 0$

Question 25. [1 MARK]

Consider the convex function $g(w_1, w_2) = w_1^2 + 4w_2^2 - 6w_1 - 16w_2 + 50$, where $w_1, w_2 \in \mathbb{R}$.
At what values of w_1 and w_2 is $g(w_1, w_2)$ minimized?

A. $w_1 = 2, \quad w_2 = 3$

B. $w_1 = 3, \quad w_2 = 2$

C. $w_1 = 2, \quad w_2 = 2$

D. $w_1 = 3, \quad w_2 = 3$