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. You got exam version **0**. Please mark **0 in the special code section in coloum J** of your scantron.
3. Verify that your exam package includes 26 pages, along with a formula sheet and a blank page at the end.
4. **Only the scantron will be marked**. All of your answers must be clearly marked on the scantron.
5. Use **pencil only** to fill out the scantron (preferably an HB or #2 pencil).
6. **Erase mistakes completely** on the scantron to avoid misreading by the scanner.
7. **Mark answers firmly and darkly**, filling in the bubbles completely.
8. This exam consists of **25 questions**. Each question is worth **1 mark**. The exam is worth a total of **25 marks**.
9. 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 exponential functions of the form $f(x) = ae^{bx} + c$ where $a, b, c \in \mathbb{R}$ and $b \neq 0$?

A. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^{bx} + c,\ a, b, c \in \mathbb{R},\ b \neq 0\}$

B. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^{bx^2} + c,\ a, b, c \in \mathbb{R}\}$

C. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ax^b + c,\ a, b, c \in \mathbb{R}\}$

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

E. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^{bx} + cx,\ a, b, c \in \mathbb{R}\}$

Solution 1.

Correct Answer: A.
Explanation:

A. The correct choice since it represents the set of all exponential functions.

B. A quadratic exponent, making it not an exponential function.

C. A power function instead of an exponential function.

D. A restricted form where $b = 1$ and $c = 0$.

E. An additional linear term, which does not belong in an exponential function.

Question 2. [1 MARK]

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

A. $\mathcal{A}((a, b, c), (d, e, g)) = f$, where $f(x_1, x_2) = dx_1 + ex_2^2 + a$

B. $\mathcal{A}((a, b), (d)) = f$, where $f(x) = ax^2 + bx + d$

C. $\mathcal{A}((a, b), (d)) = f$, where $f(x_1, x_2) = ax_1^2 + bx_2 + dx_1$

D. $\mathcal{A}((a, b, c), (d, e, g)) = f$, where $f(x_1, x_2) = ax_1 + bx_2 + c$

E. $\mathcal{A}((a, b), (d)) = f$, where $f(x_1, x_2) = dx_1 + bx_2^2 + a$

Solution 2.

Correct Answers: A., D.
Explanation: The function $\mathcal{A}$ is defined as mapping elements from $(\mathbb{R}^2 \times \mathbb{R})^2$ to a function space $\{f \mid f : \mathbb{R}^2 \to \mathbb{R}\}$. This means that for the input $((a, b, c), (d, e, g))$, the output must be a function of two real variables $f(x_1, x_2)$, mapping from $\mathbb{R}^2$ to $\mathbb{R}$.

A. $f(x_1, x_2) = dx_1 + ex_2^2 + a$: This is a valid function of two variables x_1 and x_2, correctly constructed from the input parameters.

B. $f(x) = ax^2 + bx + d$: This function is in terms of only one variable x, rather than x_1, x_2, making it incorrect. Also, the domain of $\mathcal{A}$ is defined as two tuples each with three real numbers.

C. The domain of $\mathcal{A}$ is defined as two tuples each with three real numbers.

D. $f(x_1, x_2) = ax_1 + bx_2 + c$: This is a simple linear function of two variables, correctly constructed from $\mathcal{A}$'s inputs.

E. The domain of $\mathcal{A}$ is defined as two tuples each with three real numbers.

Question 3. [1 MARK]

Given $\mathcal{D} = ((x_1, y_1), \ldots, (x_n, y_n))$, the loss function $\ell(\hat{y}, y) = (\hat{y} - y)^2$, and the predictor $f(x) = ax + b$, 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. $\sum_{i=0}^{n} (ax_i + b - y_i)^2$

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

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

D. $\sum_{(x,y)\in\mathcal{D}} (ax + b - y)^2$

E. $\sum_{i=0}^{n+1} |ax_i + b - y_i|$

Solution 3.

Correct Answers: B., D.
Explanation: The total loss is computed by summing the squared loss function $\ell(\hat{y}, y) = (\hat{y} - y)^2$ over all data points in $\mathcal{D}$. The correct formula follows:

$$L(f) = \sum_{i=1}^{n} (f(x_i) - y_i)^2 = \sum_{i=1}^{n} (ax_i + b - y_i)^2.$$

A. $\sum_{i=0}^{n} (ax_i + b - y_i)^2$: The summation starts at $i = 0$ instead of $i = 1$. The definition of $\mathcal{D}$ is $((x_1, y_1), \ldots, (x_n, y_n))$, so the index should start at 1.

B. $\sum_{i=1}^{n} (ax_i + b - y_i)^2$: This expression correctly computes the total squared loss by summing from $i = 1$ to $i = n$.

C. $\sum_{i=1}^{n+1} (ax_i + b - y_i)^2$: The summation extends beyond the available data points, as it goes up to $n + 1$, which does not exist in $\mathcal{D} = ((x_1, y_1), \ldots, (x_n, y_n))$.

D. $\sum_{(x,y)\in\mathcal{D}} (ax + b - y)^2$: Instead of using an explicit index, this expression sums over all elements (x, y) in the dataset $\mathcal{D}$. While different in notation, it correctly represents the total squared loss.

E. $\sum_{i=1}^{n+1} |ax_i + b - y_i|$: This expression introduces two errors: it extends to $n + 1$, including a non-existent term, and it applies absolute value instead of squaring.

Question 4. [1 MARK]

Let $h(x, y) = x^2 + y^{-1}$ where $x, y \in \mathbb{R}$. What is

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

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

A. 15

B. 22

C. 22.5

D. 30

E. 48

Solution 4.

Correct Answer: 22
Explanation: We are given the function

$$h(x, y) = x^2 + \frac{1}{y},$$

with $x \in \mathcal{X} = [0, 3]$ and $y \in \mathcal{Y} = \{1, 3\}$. We wish to evaluate

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

For a fixed y, the integral is

$$\int_0^3 h(x, y)\, dx = \int_0^3 \left(x^2 + \frac{1}{y} \right) dx.$$

Since $\frac{1}{y}$ is a constant with respect to x, we can split the integral:

$$\int_0^3 h(x, y)\, dx = \int_0^3 x^2\, dx + \frac{1}{y} \int_0^3 dx.$$

We now compute the two integrals separately:

$$\int_0^3 x^2\, dx = \left[\frac{x^3}{3} \right]_0^3 = \frac{27}{3} = 9,$$

$$\int_0^3 dx = [x]_0^3 = 3.$$

Thus, for a fixed y,

$$\int_0^3 h(x, y)\, dx = 9 + \frac{3}{y}.$$

Evaluating for each y in $\mathcal{Y}$:
- For $y = 1$:

$$\int_0^3 h(x,1)\,dx = 9 + \frac{3}{1} = 12.$$

- For $y = 3$:

$$\int_0^3 h(x,3)\,dx = 9 + \frac{3}{3} = 10.$$

Summing the results, we obtain:

$$\sum_{y\in\{1,3\}} \int_0^3 h(x,y)\,dx = 12 + 10 = 22.$$

Question 5. [1 MARK]

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

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

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

A. $-e + \frac{1}{2}$

B. $-e - \frac{1}{2}$

C. $e + \frac{1}{2}$

D. $e - \frac{1}{2}$

E. $\frac{1}{2}e - \frac{3}{2}$

Solution 5.

Correct Answer: D. $e - \frac{1}{2}$
Explanation: We are given the function

$$h(x,y) = e^x + y,$$

with $x, y \in \mathbb{R}$ and the regions

$$\mathcal{X} = [0,1] \quad \text{and} \quad \mathcal{Y} = [0,1].$$

We wish to evaluate

$$\int_{x=0}^1 \int_{y=0}^1 \left(e^x + y\right)\,dy\,dx.$$

Step 1: Compute the inner integral with respect to y
For a fixed x,

$$\int_{y=0}^1 \left(e^x + y\right)\,dy = e^x \int_0^1 dy + \int_0^1 y\,dy.$$

Since e^x is constant with respect to y,

$$\int_0^1 dy = 1 \quad \text{and} \quad \int_0^1 y\,dy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2}.$$

Thus, the inner integral becomes:

$$\int_0^1 (e^x + y)\ dy = e^x + \frac{1}{2}.$$

Step 2: Compute the outer integral with respect to x
Now, integrate with respect to x:

$$\int_{x=0}^1 \left(e^x + \frac{1}{2}\right)\ dx = \int_0^1 e^x\,dx + \frac{1}{2}\int_0^1 dx.$$

The integral of e^x is:

$$\int_0^1 e^x\,dx = [e^x]_0^1 = e - 1,$$

and the integral of 1 is:

$$\int_0^1 dx = 1.$$

Thus,

$$\int_0^1 \left(e^x + \frac{1}{2}\right)\ dx = (e-1) + \frac{1}{2}.$$

Question 6. [1 MARK]

Let X be a discrete random variable uniformly distributed over the outcome space

$$\mathcal{X} = \{-1, 1, 2, 3, 5\}.$$

The probability mass function (pmf) of X is

$$p(x) = \frac{1}{5} \quad \text{for each } x \in \mathcal{X}.$$

Which of these statements are true?

A. The probability that X is less than 2 is 0.4.

B. The expected value is $\mathbb{E}[X] = 2$.

C. The variance is $\mathrm{Var}(X) = 4$.

D. The probability that X is greater or equal to 3 is 0.6.

E. The probability that X is positive is 0.8.

Solution 6.

Correct Answers: A., B., C., E.
Explanation:

A. For $X < 2$, the outcomes are $\{-1, 1\}$. Therefore,

$$P(X < 2) = \frac{2}{5} = 0.4,$$

so statement A is true.

B. The expected value is

$$\mathbb{E}[X] = \frac{-1 + 1 + 2 + 3 + 5}{5} = \frac{10}{5} = 2,$$

so statement B is true.

C. To compute the variance, first find

$$\mathbb{E}[X^2] = \frac{(-1)^2 + 1^2 + 2^2 + 3^2 + 5^2}{5} = \frac{1 + 1 + 4 + 9 + 25}{5} = \frac{40}{5} = 8.$$

Then,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 8 - 2^2 = 8 - 4 = 4.$$

Hence, statement C is true.

D. For $X \geq 3$, the outcomes are $\{3, 5\}$. Thus,

$$P(X \geq 3) = \frac{2}{5} = 0.4,$$

which does not equal 0.6. Therefore, statement D is false.

E. The positive outcomes are $\{1, 2, 3, 5\}$. Hence,

$$P(X \text{ is positive}) = \frac{4}{5} = 0.8,$$

so statement E is true.

Question 7. [1 MARK]

Suppose you toss three fair coins. Let the first coin be represented by the random variable $C_1 \in \{H, T\}$, the second coin by $C_2 \in \{H, T\}$, and the third coin by $C_3 \in \{H, T\}$. Which of the following sets represents the outcome space of the random variable $C = (C_1, C_2, C_3)$?

A. $\{(c_1, c_2, c_3) \mid c_1, c_2, c_3 \in \{H, T\}\}$

B. $\{H, T\}^3$

C. $\{(c_1, c_2) \mid c_1, c_2 \in \{H, T\}\}$

D. $\{(c_1, c_2, c_3) \mid c_2, c_3 \in \{H, T\},\ c_1 = T\}$

E. $\{(c_1, c_2, c_3, c_4) \mid c_1, c_2, c_3, c_4 \in \{H, T\}\}$

Solution 7.

Correct Answer: A., B.
Explanation:

A. The set
$$\{(c_1, c_2, c_3) \mid c_1, c_2, c_3 \in \{H, T\}\}$$
represents all ordered triples of coin toss outcomes, which is exactly the outcome space for three coin tosses. Thus, this option is true.

B. The notation
$$\{H, T\}^3$$
is a shorthand for the set of all ordered triples where each entry is either H or T, which is equivalent to option 1. Therefore, this option is also true.

C. The set
$$\{(c_1, c_2) \mid c_1, c_2 \in \{H, T\}\}$$
represents the outcome space for two coin tosses, not three. Hence, this option is false.

D. The set
$$\{(c_1, c_2, c_3) \mid c_2, c_3 \in \{H, T\},\ c_1 = T\}$$
restricts the outcomes to those where the first coin is T only, which is a proper subset of the full outcome space. Thus, this option is false.

E. The set
$$\{(c_1, c_2, c_3, c_4) \mid c_1, c_2, c_3, c_4 \in \{H, T\}\}$$
corresponds to the outcome space for four coin tosses, which is not applicable here. Therefore, this option is false.

Question 8. [1 MARK]

Suppose your favorite German football team is playing two consecutive games. In each game, the outcomes are defined as follows:
$$W = \text{win}, \quad D = \text{draw}, \quad L = \text{loss}.$$
Thus, the outcome space for the two games is
$$\mathcal{X} = \{WW,\ WD,\ WL,\ DW,\ DD,\ DL,\ LW,\ LD,\ LL\}.$$
Which of the following are events?

A. $\{WD,\ DW\}$

B. $\{WW,\ WD,\ WL\}$

C. WL

D. $\{W,\ L\}$

E. $\varnothing$(empty set)

Solution 8.

Correct Answers: A., B., E.
Explanation:

A. The set

$$\{WD, DW\}$$

is a subset of $\mathcal{X}$ because both WD and DW are outcomes in the set; hence, it is an event.

B. The set

$$\{WW, WD, WL\}$$

is a subset of $\mathcal{X}$ (representing all outcomes where the team wins the first game); therefore, it is an event.

C. WL is an element of $\mathcal{X}$ but is not expressed as a set. Since events must be subsets of $\mathcal{X}$, this is not an event.

D. $\{W, L\}$ is not a subset of $\mathcal{X}$ because its elements are individual outcomes from a single game rather than two-game outcomes written as strings (like WW, WL, etc.); thus, it is not an event.

E. $\varnothing$ is the empty set, which is a subset of every set (including $\mathcal{X}$); hence, it is an event.

Question 9. [1 MARK]

Suppose you have a random variable Z representing the time (in seconds) it takes for a computer to boot up. You know Z is distributed according to the continuous uniform distribution over the interval

$$\mathcal{Z} = [5, 15].$$

Let p be the probability density function (pdf) of Z. Which of the following statements are true?

A. $p(10)$ is the probability that $Z = 10$.

B. $p(z)$ represents the probability mass at the point z.

C. The probability that Z is between 8 and 12 is 0.4.

D. The probability that $Z = 5$ is 0.

E. The expected value of Z is 10.

Solution 9.

Correct Answers: C., D., E.
Explanation:

A. $p(10)$ is the value of the probability density function at $z = 10$, not the probability that $Z = 10$. In a continuous distribution, the probability of any single point is 0. (False)

B. $p(z)$ represents the probability density at the point z, not the probability mass. The term "probability mass" is used for discrete random variables, so this statement is false. (False)

C. The probability that Z is between 8 and 12 is calculated as

$$P(8 \leq Z \leq 12) = \frac{12 - 8}{15 - 5} = \frac{4}{10} = 0.4.$$

This statement is true.

D. For any continuous random variable, the probability of taking a specific value is 0. Therefore, $P(Z=5)=0$ is true.

E. The expected value of Z for a uniform distribution on $[5,15]$ is

$$E[Z]=\frac{5+15}{2}=10.$$

This statement is true.

Question 10. [1 MARK]

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

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

B. The probability that $X=0$ is $\frac{1}{2}$.

C. X is a continuous random variable.

D. The variance of X is 1.

E. The mean of X is 0.

Solution 10.

Correct Answers: C., E.
Explanation:

A. The stated pdf is

$$p(x)=\frac{1}{\sqrt{2\pi}}e^{-|x|},$$

but the correct probability density function for a Laplace(0,1) distribution is

$$p(x)=\frac{1}{2}e^{-|x|}.$$

Since $\frac{1}{\sqrt{2\pi}} \neq \frac{1}{2}$, this statement is false.

B. For any continuous random variable, the probability that it takes on any specific value is 0. Therefore, $P(X=0)=0$ (not $\frac{1}{2}$), so this statement is false.

C. The Laplace distribution is defined via a continuous probability density function, so X is a continuous random variable. This statement is true.

D. The variance of a Laplace distribution with scale parameter b is $2b^2$. For $b=1$, the variance is $2 \times 1^2 = 2$, not 1. Hence, this statement is false.

E. By definition, the mean of a Laplace(0,1) distribution is 0. This statement is true.

Question 11. [1 MARK]

Suppose you have two discrete random variables X and Y with outcome spaces

$$X \in \{\text{heads}, \text{tails}\} \quad \text{and} \quad Y \in \{\text{fair}, \text{biased}\}.$$

The joint probability mass function (pmf) of X and Y is given by

$$p(\text{heads}, \text{fair}) = \frac{1}{4}, \qquad p(\text{heads}, \text{biased}) = \frac{1}{8},$$
$$p(\text{tails}, \text{fair}) = \frac{1}{4}, \qquad p(\text{tails}, \text{biased}) = \frac{3}{8}.$$

Which of the following is the conditional pmf of X given Y = fair (hint: product rule can be helpful)?

A. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{2}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{1}{2}$

B. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{4}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{3}{4}$

C. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{8}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{3}{8}$

D. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{3}{4}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{1}{4}$

E. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{3}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{2}{3}$

Solution 11.

Correct Answer: B.
Explanation: To determine the conditional pmf $p_{X|Y}(x|\text{biased})$, we use the product rule

$$p_{X|Y}(x|\text{biased}) = \frac{p(x, \text{biased})}{p_Y(\text{biased})},$$

where the marginal pmf of Y at y = biased is given by

$$p_Y(\text{biased}) = p(\text{heads}, \text{biased}) + p(\text{tails}, \text{biased}) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = \frac{1}{2}.$$

Now, we compute the conditional probabilities:

$$p_{X|Y}(\text{heads}|\text{biased}) = \frac{p(\text{heads}, \text{biased})}{p_Y(\text{biased})} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{8} \times 2 = \frac{1}{4},$$

$$p_{X|Y}(\text{tails}|\text{biased}) = \frac{p(\text{tails}, \text{biased})}{p_Y(\text{biased})} = \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{8} \times 2 = \frac{3}{4}.$$

Thus, the conditional pmf of X given Y = biased is

$$p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{4}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{3}{4}.$$

Question 12. [1 MARK]

On a winter day in Edmonton, the probability that a day is freezing is 0.2 and 0.8 that it is adequate. Assuming that the temperature on consecutive days is independent and identically distributed, what is the probability that exactly one out of three consecutive days is freezing?

A. 0.096

B. 0.128

C. 0.256

D. 0.384

E. 0.512

Solution 12.

Correct Answer: D.
Explanation:
Since each day has a 0.2 chance of being freezing and a 0.8 chance of being adequate, and the days are independent, we calculate the probability that exactly one day is freezing by considering the different ways this can occur. There are 3 ways to choose which of the 3 days is the freezing day. For the chosen day, the probability is 0.2, and for the other two days, the probability is 0.8 each. Thus, the probability is given by

$$3 \times (0.2) \times (0.8)^2 = 3 \times 0.2 \times 0.64 = 3 \times 0.128 = 0.384.$$

Question 13. [1 MARK]

Let X be a random variable taking values in $\mathbb{R}$, and define

$$Y = X^2.$$

Which of the following statements are true?

A. Y is a random variable.

B. The outcome space of Y is $[0, \infty)$.

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

D. Y is not a random variable.

E. Y has the same probability distribution as X.

Solution 13.

Correct Answers: A., B.
Explanation:

A. Since $Y = X^2$ is defined as a function of the random variable X (and the squaring function is measurable), Y is itself a random variable. Therefore, statement A is true.

B. For any $x \in \mathbb{R}$, squaring yields $x^2 \geq 0$. Hence, the set of possible values of Y is $[0, \infty)$, not all of $\mathbb{R}$. Thus, statement B is true.

C. As noted above, Y only takes nonnegative values, so its outcome space cannot be $\mathbb{R}$. Therefore, statement C is false.

D. Since we have already established that Y is a random variable, the claim that Y is not a random variable is false. Hence, statement D is false.

E. In general, applying a function (in this case, the square function) to a random variable alters its distribution. Statement E is false.

Question 14. [1 MARK]

Suppose X_1, X_2, X_3 are independent random variables, each with $X_i \sim \mathcal{N}(2, 5)$. Let $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$. Which of the following statements are true?

A. $\mathbb{E}[\bar{X}] = 2$.

B. $\text{Var}(\bar{X}) = \frac{5}{3}$.

C. $\mathbb{E}[\bar{X}] = \frac{2}{3}$.

D. The variance of $\frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$ is smaller than the variance of $\bar{X}$.

E. The variance of $\frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$ is equal or higher than the variance of $\bar{X}$.

Solution 14.

Correct Answers: A., B., D.
Explanation:

A. Using linearity of expectation,

$$\mathbb{E}[\bar{X}] = \frac{1}{3}\Big(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3]\Big) = \frac{1}{3}(2 + 2 + 2) = 2.$$

So, $\mathbb{E}[\bar{X}] = 2$ is true.

B. Since each X_i has variance 5 and the X_i are independent,

$$\text{Var}(X_1 + X_2 + X_3) = 5 + 5 + 5 = 15.$$

Thus, the variance of the sample mean is

$$\text{Var}(\bar{X}) = \frac{15}{3^2} = \frac{15}{9} = \frac{5}{3}.$$

So, $\text{Var}(\bar{X}) = \frac{5}{3}$ is true.

C. The claim that $\mathbb{E}[\bar{X}] = \frac{2}{3}$ contradicts the calculation above, so it is false.

D. For 5 independent observations, define

$$\bar{X}_5 = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5).$$

Since each X_i has variance 5, the variance of their sum is $5 \times 5 = 25$, and thus

$$\text{Var}(\bar{X}_5) = \frac{25}{5^2} = \frac{25}{25} = 1.$$

Since $1 < \frac{5}{3}$, the variance of $\bar{X}_5$ is smaller than that of $\bar{X}$. Therefore, this statement is true.

E. The statement that the variance of $\bar{X}_5$ is equal to or higher than the variance of $\bar{X}$ is false because we have shown that $\text{Var}(\bar{X}_5) = 1$ is less than $\text{Var}(\bar{X}) = \frac{5}{3}$.

Question 15. [1 MARK]

Suppose Y is a random variable with outcome space

$$\mathcal{Y} = \{1, 2, 3, 4\},$$

and probability mass function given by

$$p(1) = 0.1, \quad p(2) = 0.3, \quad p(3) = 0.4, \quad p(4) = 0.2.$$

Let the function f be defined as follows:

$$f(1) = 2, \quad f(2) = 5, \quad f(3) = 1, \quad f(4) = 4.$$

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

A. 2.9

B. 3.1

C. 2.5

D. 3.3

E. 2.7

Solution 15.

Correct Answer: A. 2.9
Explanation:
The expected value of $f(Y)$ is computed as

$$\mathbb{E}[f(Y)] = \sum_{y \in \mathcal{Y}} f(y)\, p(y).$$

Substitute the given values:

$$\mathbb{E}[f(Y)] = 2 \cdot 0.1 + 5 \cdot 0.3 + 1 \cdot 0.4 + 4 \cdot 0.2 = 0.2 + 1.5 + 0.4 + 0.8 = 2.9.$$

Question 16. [1 MARK]

Let X represent whether a bus arrives on time, where $X = 1$ (on time) with probability 0.75 and $X = 0$ (late) with probability 0.25. Given that $X = 1$, the conditional distribution of the number of passengers N boarding the bus is given by

$$\mathbb{P}_{N|X}(N = 10 \mid X = 1) = 0.2, \quad \mathbb{P}_{N|X}(N = 20 \mid X = 1) = 0.5, \quad \mathbb{P}_{N|X}(N = 30 \mid X = 1) = 0.3.$$

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

A. 20

B. 21

C. 22

D. 23

E. 24

Solution 16.

Correct Answer: B. 21
Explanation:
We use the formula for conditional expectation:

$$\mathbb{E}[N \mid X = 1] = \sum_{n \in \{10,20,30\}} n \, \mathbb{P}_{N|X}(n \mid X = 1).$$

Substituting the given probabilities, we have:

$$\mathbb{E}[N \mid X = 1] = 10 \cdot 0.2 + 20 \cdot 0.5 + 30 \cdot 0.3 = 2 + 10 + 9 = 21.$$

Question 17. [1 MARK]

You are given data consisting of the delivery time (in hours) for a package, the distance to the destination (in kilometers), the package weight (in kilograms), and the prevailing weather condition (e.g., "clear", "rainy", "snowy"). Suppose you want to predict the delivery time of a package. Which of the following options correctly identifies the features and the label for this problem?

A. Features: delivery time, distance, package weight; Label: weather condition.

B. Features: package weight, weather condition; Label: delivery time.

C. Features: distance, package weight, weather condition; Label: delivery time.

D. Features: delivery time, weather condition; Label: package weight.

E. Features: distance, weather condition; Label: delivery time.

Solution 17.

Correct Answer: C.
Explanation:

A. The label is set as the weather condition, which is incorrect because we want to predict the delivery time.

B. Although the label is correctly given as delivery time, this option omits distance from the features.

C. This option correctly uses the available predictors (distance, package weight, weather condition) to predict the delivery time.

D. This option incorrectly uses package weight as the label while treating delivery time as a feature.

E. Here the label is correct, but the features are incomplete because package weight is missing.

Question 18. [1 MARK]

Suppose you have a dataset where each instance represents a movie. The dataset includes the following variables: production budget (in millions of dollars), runtime (in minutes), the number of prizes the movie received, and the movie's generated revenue represented as a number with digits behind the decimal point. You wish to predict the movie's generated revenue. Which of the following statements are true?

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

B. The label set is finite.

C. This is a regression problem.

D. This is a classification problem.

E. The label set can be represented as a subset of $\mathbb{R}$.

Solution 18.

Correct Answers: C., E.
Explanation:

A. The features are production budget (in millions), runtime (in minutes), and number of prizes; these can have digits behind the dot and hence - unless assumed whole numbers - cannot be represented as a vector in $\mathbb{N}^3$. (False)

B. The label, generated revenue is a continuous numerical value, so its set is not finite. (False)

C. Since we are predicting a numerical value (revenue), this is a regression problem. (True)

D. Predicting a continuous numerical value is not a classification task. (False)

E. The label, being a positive number, can be represented as a subset of $\mathbb{R}$. (True)

Question 19. [1 MARK]

Let X be a random vector in $\mathbb{R}$ representing a feature extracted from an email, and let $Y \in \{\text{spam}, \text{not spam}\}$ be the corresponding label. Let $g(X)$ be a classifier that outputs a predicted label, and define the loss function as

$$l(g(X), Y) = \begin{cases} 1, & \text{if } g(X) \neq Y, \\ 0, & \text{if } g(X) = Y. \end{cases}$$

Which of the following statements about the expected loss are true?

A. The expected loss can be expressed as

$$\mathbb{E}[l(g(X), Y)] = \int_{\mathbb{R}} \sum_{y \in \{\text{spam,not spam}\}} l(g(x), y)\, p(x, y)\, dx.$$

B. The expected loss can be written as

$$\mathbb{E}[l(Y, g(X))] = \int_{\mathbb{R}} \sum_{y \in \{\text{spam,not spam}\}} l(y, g(x))\, p(x \mid y)\, dx.$$

C. The expected loss is given by

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

D. The expected loss is computed as

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

E. The expected loss is calculated with respect to the distribution $\mathbb{P}_{X,Y}$.

Solution 19.

Correct Answers: A., E.
Explanation:

A. This expression correctly computes the expected loss by summing over the two possible values of Y and integrating over X using the joint density $p(x, y)$. (True)

B. This expression attempts to write the expected loss using the conditional density $p(x \mid y)$, but it omits the necessary weighting by $p(y)$ (i.e., $p(x, y) = p(x \mid y)p(y)$). Even though the loss function is symmetric, missing $p(y)$ makes the expression incorrect. (False)

C. This expression only integrates over x with respect to $p(x)$ and does not sum over the possible values of y, so it does not properly account for the variation in Y. (False)

D. This expression omits any probability density altogether, so it does not represent an expectation. (False)

E. This statement correctly notes that the expected loss is calculated with respect to the joint distribution $\mathbb{P}_{X,Y}$. (True)

Question 20. [1 MARK]

Suppose you roll a six-sided die 4 times and obtain the following outcomes:

$$X_1 = 2, \quad X_2 = 5, \quad X_3 = 3, \quad X_4 = 6.$$

The sample mean is $\bar{X} = 4$. What is the sample variance of these 4 rolls?

A. $\frac{10}{3}$

B. 3.0

C. 4.0

D. $\frac{8}{3}$

E. 2.5

Solution 20.

Correct Answer: E.
Explanation:
Using the definition of the sample variance,

$$\text{Var}[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

with $n = 4$ and $\bar{X} = 4$, we first compute the squared deviations:

$$(2-4)^2 = 4, \quad (5-4)^2 = 1, \quad (3-4)^2 = 1, \quad (6-4)^2 = 4.$$

The sum of these squared deviations is:

$$4 + 1 + 1 + 4 = 10.$$

Thus, the sample variance is:

$$\text{Var}[\bar{X}] = \frac{10}{4} = 2.5.$$

Question 21. [1 MARK]

Suppose you are given a predictor $f(x) = 4 + x$, 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: $\mathcal{D} = \{(1, 4.2), (2, 5.7), (3, 6.8), (4, 8.5)\}$. Let the loss function be the absolute loss $\ell(f(x), y) = \left|f(x) - y\right|$. Calculate $\hat{L}(f) = \frac{1}{4}\sum_{i=1}^{4} \ell(f(x_i), y_i)$. Which of the following statements are true?

A. $\hat{L}(f) = 0.425$

B. $\hat{L}(f) = 0.450$

C. $\hat{L}(f) = 0.400$

D. $\hat{L}(f) = 0.500$

E. $\hat{L}(f)$ is an estimate of the expected absolute loss $L(f)$.

Solution 21.

Correct Answer: B., E.
Explanation:
We first compute the predictions and the absolute losses for each data point:

$$\begin{aligned}
f(1) &= 4 + 1 = 5, \quad \ell(f(1), 4.2) = |5 - 4.2| = 0.8, \\
f(2) &= 4 + 2 = 6, \quad \ell(f(2), 5.7) = |6 - 5.7| = 0.3, \\
f(3) &= 4 + 3 = 7, \quad \ell(f(3), 6.8) = |7 - 6.8| = 0.2, \\
f(4) &= 4 + 4 = 8, \quad \ell(f(4), 8.5) = |8 - 8.5| = 0.5.
\end{aligned}$$

The total loss is:

$$0.8 + 0.3 + 0.2 + 0.5 = 1.8.$$

Thus, the estimated expected loss is:

$$\hat{L}(f) = \frac{1.8}{4} = 0.45.$$

Now, evaluating each statement:

A. $\hat{L}(f) = 0.425$ (False)

B. $\hat{L}(f) = 0.450$ (True)

C. $\hat{L}(f) = 0.400$ (False)

D. $\hat{L}(f) = 0.500$ (False)

E. $\hat{L}(f)$ is an estimate of the expected absolute loss $L(f)$. (True)

Question 22. [1 MARK]

Consider the function $f(x) = \ln(3x+1) + 2x^2$, where $x \in \mathbb{R}$. What is the first derivative $f'(x)$?

A. $f'(x) = \frac{1}{3x+1} + 4x$

B. $f'(x) = \frac{3}{3x+1} + 2x$

C. $f'(x) = \frac{3}{(3x+1)^2} + 4x$

D. $f'(x) = \frac{3}{3x+1} + 4x$

E. $f'(x) = \frac{3}{3x+1} + 4$

Solution 22.

Correct Answer: D.
Explanation:
We differentiate $f(x) = \ln(3x+1) + 2x^2$ term by term.
For the first term, using the chain rule:

$$\frac{d}{dx}\ln(3x+1) = \frac{1}{3x+1} \cdot 3 = \frac{3}{3x+1}.$$

For the second term:

$$\frac{d}{dx}(2x^2) = 4x.$$

Thus, the derivative is:

$$f'(x) = \frac{3}{3x+1} + 4x.$$

Question 23. [1 MARK]

Consider the function $f(x) = \frac{1}{12}x^4 - 4x^2 + 2$, where its second derivative is: $f''(x) = x^2 - 8$. Which of the following statements are true?

A. f is convex on $\mathbb{R}$.

B. $f''(x) = x^2 - 8$ is negative for some $x \in \mathbb{R}$.

C. f is not convex on $\mathbb{R}$.

D. f is convex on a subset of $\mathbb{R}$.

Solution 23.

Correct Answer: B., C, D.
Explanation:
We have

$$f''(x) = x^2 - 8.$$

Note that when $x = 0$,

$$f''(0) = 0^2 - 8 = -8 < 0,$$

so there exist values of x (in fact, for all $|x| < 2\sqrt{2}$) where $f''(x) < 0$. Therefore:

A. f is not convex on all of $\mathbb{R}$ because convexity requires $f''(x) \geq 0$ for every $x \in \mathbb{R}$. (false)

B. The statement that $f''(x)$ is negative for some x is true.

C. True.

D. Since $f''(x) \geq 0$ for $|x| \geq 2\sqrt{2}$, f is convex on those intervals; that is, f is convex on a subset of $\mathbb{R}$ (true).

Question 24. [1 MARK]

Consider the convex function $h(z) = \sum_{i=1}^{3} \big(5z - a_i\big)^2$, where $a_1 = 2$, $a_2 = 4$, $a_3 = 6$, and $z \in \mathbb{R}$. Which value of z minimizes $h(z)$?

A. $z = 0.8$

B. $z = 1.0$

C. $z = 0.6$

D. $z = 1.2$

Solution 24.

Correct Answer: A.
Explanation:
We are given

$$h(z) = (5z-2)^2 + (5z-4)^2 + (5z-6)^2.$$

To find the minimizing value of z, we compute the first derivative $h'(z)$ and set it equal to zero. Differentiate each term using the chain rule:

$$\frac{d}{dz}(5z-2)^2 = 2(5z-2)\cdot 5 = 10(5z-2),$$

$$\frac{d}{dz}(5z-4)^2 = 2(5z-4)\cdot 5 = 10(5z-4),$$

$$\frac{d}{dz}(5z-6)^2 = 2(5z-6)\cdot 5 = 10(5z-6).$$

Thus, the derivative is:

$$h'(z) = 10\big[(5z-2) + (5z-4) + (5z-6)\big] = 10\big[15z - (2+4+6)\big] = 10(15z-12).$$

Setting $h'(z) = 0$:

$$10(15z-12) = 0 \quad \Longrightarrow \quad 15z - 12 = 0 \quad \Longrightarrow \quad z = \frac{12}{15} = \frac{4}{5} = 0.8.$$

Therefore, the value of z that minimizes $h(z)$ is 0.8.

Question 25. [1 MARK]

Suppose you are working within the empirical risk minimization (ERM) framework for supervised learning. Let f be a predictor with parameters θ and let the loss function be $\ell(f(x), y)$. The empirical risk is defined as

$$\hat{L}(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(f(x_i), y_i)$$

over a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$. Which of the following statements correctly describe the estimation and optimization steps in ERM?

A. The estimation step involves computing the empirical risk, which serves as an approximation of the true risk

$$L(f) = \mathbb{E}_{(x,y)\sim\mathbb{P}_{X,Y}}[\ell(f(x), y)]$$

because the true data distribution is unknown.

B. The optimization step involves finding the model parameters θ that minimize the empirical risk $\hat{L}(f)$.

C. The optimization step always guarantees that the global minimum of the empirical risk is found.

D. The estimation step is necessary because we use the training data to approximate the expected loss over the true distribution.

E. The expected risk $L(f)$ is computed with respect to the true distribution $\mathbb{P}_{X,Y}$, which is generally not directly available.

Solution 25.

Correct Answer: A., B., D., E.
Explanation:

A. True since the true distribution $\mathbb{P}_{X,Y}$ is unknown.

B. True because the optimization step seeks the parameters θ that minimize $\hat{L}(f)$.

C. False; in general, particularly for non-convex loss functions, the optimization step does not guarantee finding the global minimum.

D. True.

E. True since $R(f)$ is defined with respect to the true (but usually unknown) distribution $\mathbb{P}_{X,Y}$.

For your notes (1/4)

For your notes (2/4)

For your notes (3/4)

For your notes (4/4)