# Homework Assignment 3

**Due: Friday, October 3, 2025, 11:59 p.m. Mountain time**
**Total marks: 20**

**Policies:**

For all multiple-choice questions, note that multiple correct answers may exist. However, 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. Similarly, if the number of incorrect answers selected exceeds the correct ones, your score for that question will be zero. Please note that it is not possible to receive negative marks. **You must select all the correct options to get full marks for the question.**

This PDF version of the questions has been provided for your convenience should you wish to print them and work offline.

**Only answers submitted through the Canvas quiz system will be graded. Please do not submit a written copy of your responses.**

## Question 1. [1 MARK]

Suppose you are tasked with predicting the age of a house based on the following information: the size of the house in square feet, the number of bedrooms, and the year it was built. Which of the following options correctly specifies the features and the label for this problem?

a. Features: size of the house, number of bedrooms, year built; Label: age of the house

b. Features: year built, number of bedrooms; Label: size of the house

c. Features: age of the house, size of the house; Label: number of bedrooms

d. Features: number of bedrooms, year built; Label: size of the house

## Question 2. [1 MARK]

Imagine you want to predict the likelihood that a customer will purchase a product based on their past purchase history, the amount of time spent on the website, and their demographic information (e.g., age, gender, location). Which of the following options correctly specifies the features and the label for this problem?

a. Features: time spent on the website, past purchase history; Label: demographic info

b. Features: likelihood of purchase, past purchase history; Label: time spent on the website

c. Features: demographic info, likelihood of purchase; Label: past purchase history

d. Features: past purchase history, time spent on the website, demographic info; Label: likelihood of purchase

## Question 3. [1 MARK]

You are provided with the daily temperatures, humidity levels, and wind speeds in a city. You need to predict whether the city's next recorded temperature will exceed the highest recorded temperature for that day. Which of the following options correctly specifies the features and the label for this problem?

a. Features: daily temperature, humidity levels, wind speeds; Label: whether the next recorded temperature exceeds the record high

b. Features: wind speeds, daily temperature; Label: humidity levels

c. Features: whether the next recorded temperature exceeds the record high; Label: daily temperature, humidity levels, wind speeds

d. Features: humidity levels, whether the next recorded temperature exceeds the record high; Label: wind speeds

## Question 4. [1 MARK]

Suppose you have a dataset where each instance represents a car. The features are: weight of the car (in kg), engine power (in horsepower), and the number of seats. The label is the price of the car in dollars. True or False: This is a regression problem.

## Question 5. [1 MARK]

You are working on a dataset where each instance contains information about a flower. The features are petal length, petal width, and sepal length (all in cm), and the label indicates the species of the flower (one of three categories: species A, species B, or species C). True or False: This is a classification problem.

## Question 6. [1 MARK]

Consider a dataset where each instance records a patient's age, weight, and height. The label represents the patient's blood type: A, B, AB, or O. True or False: This is a classification problem.

## Question 7. [1 MARK]

Suppose you are working with a dataset where the features $X \in \mathcal{X} = \mathbb{R}$ and the labels $Y \in \mathcal{Y} = \mathbb{R}$. You are using a predictor $f(X)$ and the absolute loss function $\ell(f(X), Y) = |f(X) - Y|$, where $f(X)$ represents the prediction and $Y$ represents the label.
Which of the following expressions correctly represents the expected loss?

a. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} |f(x) - y| p(x, y)\, dy$

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

c. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x) - y| p(y \mid x) p(x)\, dy\, dx$

d. $\mathbb{E}[\ell(f(X), Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} \log(1 + |f(x) - y|) p(x, y)\, dy\, dx$

## Question 8. [1 MARK]

Suppose you flip a fair coin 5 times and observe the following outcomes: $X_1 = 1$, $X_2 = 0$, $X_3 = 1$, $X_4 = 0$, and $X_5 = 1$, where 1 represents heads and 0 represents tails. What is the sample mean of these 5 coin flips?
(Do not write your answer as a fraction. Instead, express it as a decimal number, rounded to two decimal places if necessary. For example, write 0.33 instead of 1/3.)

## Question 9. [1 MARK]

Suppose you are given a predictor $f(x) = 10 - 0.2x$, which models the relationship between the age of a house $x$ (in years) and its price $y$ (in hundreds of thousands of dollars). You are provided with the following dataset of 5 $(x, y)$ pairs:

$$\mathcal{D} = ((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)) = ((1, 9.5), (2, 9), (3, 8.7), (4, 8), (5, 7.8))$$

Let the loss function be the squared loss $\ell(f(x), y) = (f(x)-y)^2$. Calculate $\hat{L}(f) = \frac{1}{5}\sum_{i=1}^{5} \ell(f(x_i), y_i)$ (which is an estimate of the expected squared loss $L(f) = \mathbb{E}[\ell(f(x), y)]$ using the sample mean over the dataset).
(Do not write your answer as a fraction. Instead, express it as a decimal number, rounded to two decimal places if necessary. For example, write 0.33 instead of 1/3.)

## Question 10. [1 MARK]

Suppose you are given the same setup as in the previous question, but the loss function is changed to the absolute loss $\ell(f(x), y) = |f(x) - y|$. Calculate $\hat{L}(f) = \frac{1}{5}\sum_{i=1}^{5} \ell(f(x_i), y_i)$.
(Do not write your answer as a fraction. Instead, express it as a decimal number, rounded to two decimal places if necessary. For example, write 0.33 instead of 1/3.)

## Question 11. [1 MARK]

Suppose you are given a predictor

$$f(x) = \begin{cases} 1 & \text{if } x > 40 \\ 0 & \text{otherwise,} \end{cases}$$

which models the relationship between the length of an email $x$ (in words) and its classification $y$ (spam = 1, not spam = 0). You are provided with the following dataset of 5 $(x, y)$ pairs:

$$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)\} = \{(20, 0), (50, 1), (35, 1), (60, 0), (45, 1)\}$$

Let the loss function be the 0-1 loss

$$\ell(f(x), y) = \begin{cases} 1 & \text{if } f(x) \neq y \\ 0 & \text{if } f(x) = y. \end{cases}$$

Calculate $\hat{L}(f) = \frac{1}{5}\sum_{i=1}^{5} \ell(f(x_i), y_i)$.
(Do not write your answer as a fraction. Instead, express it as a decimal number, rounded to two decimal places if necessary. For example, write 0.33 instead of 1/3.)

## Question 12. [1 MARK]

Let $X$ be a normally distributed random variable with mean $\mu$ and variance $\sigma^2$, denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$. Consider a sample of $n$ independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$, each following the same distribution as $X$. The sample mean $\hat{\mu}_n$ is defined as:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This sample mean $\hat{\mu}_n$ serves as an estimator for the true mean $\mu$. What are the expected values of $\hat{\mu}_{10}$ and $\hat{\mu}_{100}$?

a. $\mathbb{E}[\hat{\mu}_{10}] = \mu - \sigma$ and $\mathbb{E}[\hat{\mu}_{100}] = \mu - \sigma$

b. $\mathbb{E}[\hat{\mu}_{10}] = \mu + \sigma$ and $\mathbb{E}[\hat{\mu}_{100}] = \mu + \sigma$

c. $\mathbb{E}[\hat{\mu}_{10}] = \mu$ and $\mathbb{E}[\hat{\mu}_{100}] = \mu$

d. $\mathbb{E}[\hat{\mu}_{10}] = \mu + \frac{\sigma}{\sqrt{10}}$ and $\mathbb{E}[\hat{\mu}_{100}] = \mu + \frac{\sigma}{\sqrt{100}}$

## Question 13. [1 MARK]

Consider the same setting as in the previous question. What are the variances of $\hat{\mu}_{10}$ and $\hat{\mu}_{100}$, and which one is smaller?

a. $\text{Var}[\hat{\mu}_{10}] = \frac{\sigma^2}{10}$ and $\text{Var}[\hat{\mu}_{100}] = \frac{\sigma^2}{100}$. $\text{Var}[\hat{\mu}_{100}] < \text{Var}[\hat{\mu}_{10}]$

b. $\text{Var}[\hat{\mu}_{10}] = \frac{\sigma^2}{100}$ and $\text{Var}[\hat{\mu}_{100}] = \frac{\sigma^2}{10}$. $\text{Var}[\hat{\mu}_{100}] > \text{Var}[\hat{\mu}_{10}]$

c. $\text{Var}[\hat{\mu}_{10}] = \sigma^2$ and $\text{Var}[\hat{\mu}_{100}] = \sigma^2$. $\text{Var}[\hat{\mu}_{100}] = \text{Var}[\hat{\mu}_{10}]$

d. $\text{Var}[\hat{\mu}_{10}] = \frac{\sigma}{\sqrt{10}}$ and $\text{Var}[\hat{\mu}_{100}] = \frac{\sigma}{\sqrt{100}}$. $\text{Var}[\hat{\mu}_{100}] < \text{Var}[\hat{\mu}_{10}]$

## Question 14. [1 MARK]

True or False: If $w^*$ is the point where $g(w)$ attains its minimum value, then $w^*$ is also the point where $-g(w)$ attains its maximum value.

## Question 15. [1 MARK]

True or False: For any function $g$, the minimum value of $g(w)$ is always equal to the maximum value of $-g(w)$.

## Question 16. [1 MARK]

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

a. $g''(w) = e^w + 2w$

b. $g''(w) = e^w + 2$

c. $g''(w) = e^w$

d. $g''(w) = 2w$

## Question 17. [1 MARK]

Consider the function $g(w) = w^4 - 4w^2 + 1$, where $w \in \mathbb{R}$.
Given that the second derivative of $g(w)$ is:

$$g''(w) = 12w^2 - 8$$

**True** or **False**: The function $g(w)$ is convex over its entire domain.

## Question 18. [1 MARK]

Consider the convex function $g(w) = 2w^2 - 4w + 3$, 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 $w = 1$.

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

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

d. The minimum value is 3 at $w = 0$.

## Question 19. [1 MARK]

Consider the convex function $g(w) = \sum_{x \in \mathcal{X}} (w - x)^2$, where $\mathcal{X} = \{1, 2, 3, 4, 5\}$ and $w \in \mathbb{R}$. What is the minimum value of $g(w)$ and at what $w$ is it achieved?

a. The minimum value is 10 at $w = 3$.

b. The minimum value is 10 at $w = 2$.

c. The minimum value is 15 at $w = 3$.

d. The minimum value is 20 at $w = 4$.

## Question 20. [1 MARK]

Consider the predictor $f(x) = xw$, where $w \in \mathbb{R}$ is a one-dimensional parameter, and $x$ represents the feature with no bias term. Suppose you are given a dataset of $n$ data points $\mathcal{D} = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$, where each $y_i$ is the target variable corresponding to feature $x_i$. Let the loss function be the squared loss $\ell(f(x), y) = (f(x) - y)^2$. The estimate of the expected loss for a parameter $w \in \mathbb{R}$ is defined as the following convex function:

$$\hat{L}(w) = \frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2$$

What is the formula for $\hat{w} = \arg\min_{w \in \mathbb{R}} \hat{L}(w)$ ?

a. $\hat{w} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$

b. $\hat{w} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$

c. $\hat{w} = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i y_i}$

d. $\hat{w} = \frac{\sum_{i=1}^n (y_i - x_i)}{n}$