Homework Assignment 3

Due: Friday, October 4, 2024, 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.**

While the syllabus initially indicated the need to submit a paragraph explaining the use of AI or other resources in your assignments, this requirement no longer applies as we are now utilizing eClass quizzes instead of handwritten submissions. Therefore, you are **not** required to submit any explanation regarding the tools or resources (such as online tools or AI) used in completing this quiz.

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 eClass 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

Solution:

The correct answer is:

- **a.** Features: size of the house, number of bedrooms, year built; Label: age 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: past purchase history, time spent on the website, demographic info; Label: likelihood of purchase

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

Solution:

The correct answer is:

- **a.** 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

Solution:

The correct answer is:

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

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.

Solution:

Answer: True.
Explanation: This is a regression problem because the label (car price) can take any real value. In other words, the label has a notion of order and can be represented as a real number $\mathbb{R}$, making it suitable for regression.

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.

Solution:

Answer: True.
Explanation: This is a classification problem because the label (flower species) has no notion of order. The labels are drawn from a finite set of categories (species A, B, or C), and there is no inherent ranking or numerical value associated with them, making this 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.

Solution:

Answer: True.
Explanation: This is a classification problem because the label (blood type) has no notion of order. The labels come from a finite set of categories (A, B, AB, or O), and there is no inherent ranking or numerical relationship between them, which makes this a classification task.

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

Solution:

The correct answers are:

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

Explanation: Both expressions **b** and **c** are correct. In **b**, the expected loss is expressed directly using the joint probability density $p(x, y)$, where the integral covers both X and Y over their entire domain $\mathbb{R}$. In **c**, the joint probability $p(x, y)$ is factorized using the product rule as $p(y \mid x)p(x)$, where $p(y \mid x)$ is the conditional probability of Y given X, and $p(x)$ is the marginal distribution of X.

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.)

Solution:

The correct value is 0.6.
Explanation: The sample mean $\hat{\alpha}$ is calculated as the average of the observed outcomes:

$$\hat{\alpha} = \frac{1+0+1+0+1}{5} = \frac{3}{5} = 0.6$$

Therefore, the sample mean of the 5 coin flips is 0.6.
Note: The sample mean represents the proportion of heads observed in the 5 flips. Since 3 out of 5 flips resulted in heads, the sample mean is 0.6.

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.)

Solution:

The correct value for the estimated expected squared loss is 0.764.
Explanation: The expected squared loss is estimated by calculating the average squared error over the 5 data points. For each data point (x_i, y_i), the squared loss is:

$$\ell(f(x_i), y_i) = (f(x_i) - y_i)^2$$

First, calculate the predictions for each x_i using the predictor $f(x) = 10 - 0.2x$:

$$f(x_1) = 10 - 0.2(1) = 9.8, \quad f(x_2) = 10 - 0.2(2) = 9.6$$
$$f(x_3) = 10 - 0.2(3) = 9.4, \quad f(x_4) = 10 - 0.2(4) = 9.2, \quad f(x_5) = 10 - 0.2(5) = 9$$

Now, calculate the squared losses for each pair:

$$\ell(f(x_1), y_1) = (9.8 - 9.5)^2 = 0.09, \quad \ell(f(x_2), y_2) = (9.6 - 9)^2 = 0.36$$
$$\ell(f(x_3), y_3) = (9.4 - 8.7)^2 = 0.49, \quad \ell(f(x_4), y_4) = (9.2 - 8)^2 = 1.44, \quad \ell(f(x_5), y_5) = (9 - 7.8)^2 = 1.44$$

The total squared loss is:

$$0.09 + 0.36 + 0.49 + 1.44 + 1.44 = 3.82$$

The sample mean of the squared loss is:

$$\hat{L}(f) = \frac{3.82}{5} = 0.764$$

Therefore, the estimate of the expected squared loss using the sample mean is 0.764.

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.)

Solution:

The correct value for the estimated expected absolute loss is 0.8.
Explanation: The expected absolute loss is estimated by calculating the average absolute error over the 5 data points. For each data point (x_i, y_i), the absolute loss is:

$$\ell(f(x_i), y_i) = |f(x_i) - y_i|$$

First, calculate the predictions for each x_i using the predictor $f(x) = 10 - 0.2x$:

$$f(x_1) = 10 - 0.2(1) = 9.8, \quad f(x_2) = 10 - 0.2(2) = 9.6$$

$$f(x_3) = 10 - 0.2(3) = 9.4, \quad f(x_4) = 10 - 0.2(4) = 9.2, \quad f(x_5) = 10 - 0.2(5) = 9$$

Now, calculate the absolute losses for each pair:

$$\ell(f(x_1), y_1) = |9.8 - 9.5| = 0.3, \quad \ell(f(x_2), y_2) = |9.6 - 9| = 0.6$$

$$\ell(f(x_3), y_3) = |9.4 - 8.7| = 0.7, \quad \ell(f(x_4), y_4) = |9.2 - 8| = 1.2, \quad \ell(f(x_5), y_5) = |9 - 7.8| = 1.2$$

The total absolute loss is:

$$0.3 + 0.6 + 0.7 + 1.2 + 1.2 = 4.0$$

The sample mean of the absolute loss is:

$$\hat{L}(f) = \frac{4.0}{5} = 0.8$$

Therefore, the estimate of the expected absolute loss using the sample mean is 0.8.

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.)

Solution:

The correct value for the estimated expected loss is 0.4.
Explanation: The expected loss is estimated by calculating the average loss over the 5 data points. For each data point (x_i, y_i), the loss is:

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

First, apply the predictor $f(x)$ to each x_i to obtain the predicted labels:

$$f(x_1) = \begin{cases} 1 & \text{if } 20 > 40 \\ 0 & \text{otherwise} \end{cases} = 0$$

$$f(x_2) = \begin{cases} 1 & \text{if } 50 > 40 \\ 0 & \text{otherwise} \end{cases} = 1$$

$$f(x_3) = \begin{cases} 1 & \text{if } 35 > 40 \\ 0 & \text{otherwise} \end{cases} = 0$$

$$f(x_4) = \begin{cases} 1 & \text{if } 60 > 40 \\ 0 & \text{otherwise} \end{cases} = 1$$

$$f(x_5) = \begin{cases} 1 & \text{if } 45 > 40 \\ 0 & \text{otherwise} \end{cases} = 1$$

Next, compare the predicted labels $f(x_i)$ with the actual labels y_i to determine the loss for each data point:

$$\ell(f(x_1), y_1) = \begin{cases} 1 & \text{if } 0 \neq 0 \\ 0 & \text{if } 0 = 0 \end{cases} = 0$$

$$\ell(f(x_2), y_2) = \begin{cases} 1 & \text{if } 1 \neq 1 \\ 0 & \text{if } 1 = 1 \end{cases} = 0$$

$$\ell(f(x_3), y_3) = \begin{cases} 1 & \text{if } 0 \neq 1 \\ 0 & \text{if } 0 = 1 \end{cases} = 1$$

$$\ell(f(x_4), y_4) = \begin{cases} 1 & \text{if } 1 \neq 0 \\ 0 & \text{if } 1 = 0 \end{cases} = 1$$

$$\ell(f(x_5), y_5) = \begin{cases} 1 & \text{if } 1 \neq 1 \\ 0 & \text{if } 1 = 1 \end{cases} = 0$$

The total loss is:

$$0 + 0 + 1 + 1 + 0 = 2$$

The sample mean of the loss is:

$$\hat{L}(f) = \frac{2}{5} = 0.4$$

Therefore, the estimate of the expected loss using the sample mean is 0.4.

Question 12. [1 MARK]

Let X be a normally distributed random variable with mean μ and variance σ^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 μ. What are the expected values of $\hat{\mu}_{10}$ and $\hat{\mu}_{100}$?

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

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

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

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

Solution:

The correct answer is:

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

Explanation:
The sample mean $\hat{\mu}_n$ of n independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance σ^2, is defined as:

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

To find the expected value of $\hat{\mu}_n$, we use the linearity of expectation:

$$\mathbb{E}[\hat{\mu}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \times n\mu = \mu$$

This result shows that the expected value of the sample mean $\hat{\mu}_n$ is equal to the true mean μ, regardless of the sample size n. Therefore:

$$\mathbb{E}[\hat{\mu}_{10}] = \mu \quad \text{and} \quad \mathbb{E}[\hat{\mu}_{100}] = \mu$$

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

Solution:

The correct answer is:

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

Explanation:
For a sample mean $\hat{\mu}_n$ of n independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance σ^2, the variance of the sample mean is given by:

$$\text{Var}[\hat{\mu}_n] = \frac{\text{Var}[X]}{n} = \frac{\sigma^2}{n}$$

Therefore:

$$\text{Var}[\hat{\mu}_{10}] = \frac{\sigma^2}{10}, \quad \text{Var}[\hat{\mu}_{100}] = \frac{\sigma^2}{100}$$

Since $\frac{\sigma^2}{100} < \frac{\sigma^2}{10}$, it follows that:

$$\text{Var}[\hat{\mu}_{100}] < \text{Var}[\hat{\mu}_{10}]$$

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.

Solution:

Answer: True.
Explanation:
If w^* is the point where $g(w)$ attains its minimum value, then by definition, for all w,

$$g(w^*) \leq g(w)$$

Multiplying both sides of the inequality by -1 (which reverses the inequality), we get:

$$-g(w^*) \geq -g(w)$$

This implies that $-g(w^*)$ is greater than or equal to $-g(w)$ for all w. Therefore, w^* is the point where $-g(w)$ attains its maximum value.
In other words, minimizing $g(w)$ is equivalent to maximizing $-g(w)$. Thus, the point w^* that minimizes $g(w)$ also maximizes $-g(w)$.

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 $-g(w)$.

Solution:

Answer: False.
Explanation:
Let m be the minimum value of $g(w)$. Then, the maximum value of $-g(w)$ is $-m$.
For m to be equal to $-m$, it must be that $m = 0$. However, in general, the minimum value m of $g(w)$ is not zero. Therefore, the minimum value of $g(w)$ is **not** equal to the maximum value of $-g(w)$ unless $m = 0$.

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

Solution:

The correct answer is:

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

Explanation:
Given the function:

$$g(w) = e^w + w^2$$

Step 1: Compute the First Derivative
First, find the first derivative of $g(w)$ with respect to w:

$$g'(w) = \frac{d}{dw}\left(e^w + w^2\right) = e^w + 2w$$

Step 2: Compute the Second Derivative
Next, find the second derivative by differentiating $g'(w)$:

$$g''(w) = \frac{d}{dw}\left(e^w + 2w\right) = e^w + 2$$

Therefore, the second derivative of $g(w)$ is:

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

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.

Solution:

Answer: False.
Explanation:
A function is convex over its entire domain if its second derivative is non-negative for all values of w.
Given:

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

Let's analyze the second derivative:
1. **Determine when $g''(w) \geq 0$:**

$$12w^2 - 8 \geq 0 \implies 12w^2 \geq 8 \implies w^2 \geq \frac{8}{12} = \frac{2}{3}$$

$$\implies |w| \geq \sqrt{\frac{2}{3}} \approx 0.816$$

2. **Interpretation:** - For $|w| \geq \sqrt{\frac{2}{3}}$, $g''(w) \geq 0$, indicating that $g(w)$ is convex in these regions. - For $|w| < \sqrt{\frac{2}{3}}$, $g''(w) < 0$, indicating that $g(w)$ is concave in these regions.

Since there exist values of w (specifically, $|w| < \sqrt{\frac{2}{3}}$) where $g''(w) < 0$, the function $g(w)$ is **not** 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 $w = 0$.

Solution:

The correct answer is:

- **a.** The minimum value is 1 at $w = 1$.

Explanation:
To find the minimum value of the function $g(w) = 2w^2 - 4w + 3$, we need to determine the value of w that minimizes $g(w)$. This involves the following steps:
Step 1: Compute the First Derivative
First, find the first derivative of $g(w)$ with respect to w:

$$g'(w) = \frac{d}{dw}(2w^2 - 4w + 3) = 4w - 4$$

Step 2: Find the Critical Point
Set the first derivative equal to zero to find the critical point:

$$4w - 4 = 0 \implies 4w = 4 \implies w = 1$$

Step 3: Calculate the Minimum Value
Substitute $w = 1$ back into the original function to find the minimum value:

$$g(1) = 2(1)^2 - 4(1) + 3 = 2 - 4 + 3 = 1$$

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 $w = 4$.

Solution:

The correct answer is:

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

Explanation:
Given the function:

$$g(w) = \sum_{x \in \mathcal{X}} (w - x)^2 = (w - 1)^2 + (w - 2)^2 + (w - 3)^2 + (w - 4)^2 + (w - 5)^2$$

Step 1: Expand the Function
First, expand each squared term:

$$(w - x)^2 = w^2 - 2wx + x^2$$

Summing over all $x \in \mathcal{X}$:

$$g(w) = \sum_{x=1}^{5}(w^2 - 2wx + x^2) = 5w^2 - 2w\sum_{x=1}^{5} x + \sum_{x=1}^{5} x^2$$

Calculate the sums:

$$\sum_{x=1}^{5} x = 1 + 2 + 3 + 4 + 5 = 15$$

$$\sum_{x=1}^{5} x^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 1 + 4 + 9 + 16 + 25 = 55$$

Substitute these into the function:

$$g(w) = 5w^2 - 2w(15) + 55 = 5w^2 - 30w + 55$$

Step 2: Compute the First Derivative
To find the minimum, compute the first derivative of $g(w)$ with respect to w:

$$g'(w) = \frac{d}{dw}(5w^2 - 30w + 55) = 10w - 30$$

Step 3: Find the Critical Point
Set the first derivative equal to zero to find the critical point:

$$10w - 30 = 0 \implies 10w = 30 \implies w = 3$$

Step 4: Calculate the Minimum Value
Substitute $w = 3$ back into the original function to find the minimum value:

$$g(3) = \sum_{x=1}^{5}(3 - x)^2 = (3-1)^2 + (3-2)^2 + (3-3)^2 + (3-4)^2 + (3-5)^2$$

$$g(3) = 2^2 + 1^2 + 0^2 + (-1)^2 + (-2)^2 = 4 + 1 + 0 + 1 + 4 = 10$$

Therefore, the minimum value of $g(w)$ is 10 at $w = 3$.

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

Solution:

The correct answer is:

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

Explanation:
To find the value of w that minimizes $\hat{L}(w)$, we perform the following steps:
Step 1: Expand $\hat{L}(w)$

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

Step 2: Take the Derivative with Respect to w

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

Step 3: Set the Derivative to Zero to Find the Minimizer

$$2w\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - 2\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) = 0$$

$$w\left(\sum_{i=1}^{n} x_i^2\right) = \sum_{i=1}^{n} x_i y_i$$

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

Therefore, the formula for $\hat{w}$ is:

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