Question 1 : What is Simple Linear Regression (SLR)? Explain its purpose.
Answer:

Let's first understand what does the Regression word mean here:

Regression means to establish a relation between two or more than two variables. In which one variable is independent (usually mapped at x axis) and the other is dependent over x.

Simple Linear Regression:

It attempts to determine the strength and characteristics of the relation between one independent variable(x) and other dependent variable(y) or in other words we can also explain it as

Simple Linear Regression (SLR) is a statistical method used to model the relationship between **one independent variable (input)** and **one dependent variable (output)** using a **straight line**. It assumes that the change in the dependent variable is directly proportional to the change in the independent variable.

The equation of SLR is:

$y=b0+b1xy = b0 + b1x.$

Purpose of Simple Linear Regression

- To **predict** the value of one variable based on another.
- To **analyze** and understand the strength and direction of the relationship between variables.
- To help in **forecasting**, **trend analysis**, and **decision making**.

Example:
Predicting house price (y) based on area in square feet (x).

Question 2: What are the key assumptions of Simple Linear Regression?
Answer:

Simple Linear Regression is based on the following key assumptions:

1. **Linearity**
   There is a linear relationship between the independent variable (X) and dependent variable (Y).

2. **Independence of Errors**
   Residuals (error terms) are independent of each other.

3. **Homoscedasticity**
   The variance of residuals is constant across all values of X (no pattern in error spread).

4. **Normality of Errors**
   Residuals follow a normal distribution.

5. **No Multicollinearity**
   Since SLR has only one independent variable, this assumption is automatically satisfied.

6. **X is measured without error**
   (The independent variable is assumed to be accurate.)

These assumptions ensure that the model provides reliable predictions and valid statistical conclusions.

Question 3: Write the mathematical equation for a simple linear regression model and explain each term.
Answer:

The mathematical equation for a Simple Linear Regression (SLR) model is:

$y=b0+b1x+\varepsilon$ Where:

- **y** → Dependent variable (predicted output)
- **x** → Independent variable (input feature)
- **b_0 (Intercept)** → Value of y when x = 0; it shifts the regression line up or down
- **b_1 (Slope/Regression Coefficient)** → Change in y for a one-unit change in x
- **ε (Error term/Residual)** → Represents the difference between actual and predicted values (captures noise and unobserved factors)

This equation forms a straight line and is used to understand and predict the relationship

between two variables.

Question 4: Provide a real-world example where simple linear regression can be applied.
Answer:

A common real-world example of applying Simple Linear Regression is **predicting house prices based on the size of the house (area in square feet)**.

- **Independent variable (X):** Area of the house (sq. ft.)
- **Dependent variable (Y):** Selling price of the house

As the area of the house increases, the price generally increases in a roughly linear manner. A linear regression model can help estimate the price of a house based on its size.

Other example alternatives (you may mention any one):

- Predicting **student marks** based on **study hours**
- Predicting **sales** from **advertising spend**
- Predicting **temperature** based on **altitude**

Question 5: What is the method of least squares in linear regression?
Answer:

Answer:
The **method of least squares** is a technique used in linear regression to find the best-fitting line through the data points. It works by **minimizing the sum of the squared differences (errors)** between the actual values and the predicted values.

These differences are called **residuals**:

Residual=(yactual−ypredicted)

The least squares method chooses slope ($\mathbf{b}_1$) and intercept ($\mathbf{b}_0$) such that:

∑(yactual−ypredicted)^2
is as **small as possible**.

Purpose:

- To obtain the **most accurate** regression line
- To **reduce overall prediction error**

In simple words:
- The method of least squares finds a line where the total squared error between actual and predicted values is minimum.

Question 6: What is Logistic Regression? How does it differ from Linear Regression?
Answer:

Answer:
Logistic Regression is a statistical classification algorithm used to predict a **categorical outcome**, usually binary (0 or 1), based on input variables. Instead of fitting a straight line, it uses the **sigmoid (logistic) function** to map predicted values to probabilities between **0 and 1**.

Difference between Logistic Regression and Linear Regression

Feature	Linear Regression	Logistic Regression
Output type	Continuous values (e.g., price, marks)	Categorical values (e.g., spam/not spam)
Prediction range	Any real number	Between 0 and 1 (probability)
Model curve/line	Straight line	S-shaped Sigmoid curve
Loss function	Mean Squared Error (MSE)	Log Loss / Binary Cross-Entropy
Purpose	Regression tasks	Classification tasks
Decision boundary	Not applicable	Yes (e.g., classify as 1 if $p \geq 0.5$)

In simple words:

- **Linear Regression predicts a number.**

- **Logistic Regression predicts a class (using probability).**

Question 7: Name and briefly describe three common evaluation metrics for regression models.
Answer:

Three commonly used evaluation metrics for regression models are:

1. **Mean Absolute Error (MAE)**
   - Measures the **average absolute difference** between actual and predicted values.
   - Less sensitive to outliers.

$$MAE = 1/n(\sum |y_{actual} - y_{predicted}|)$$

2. **Mean Squared Error (MSE)**
   - Measures the **average of squared errors** between actual and predicted values.
   - Penalizes large errors more.

$$MSE = 1/n\ ((y_{actual} - y_{predicted})2)$$

3. **R^2 Score (Coefficient of Determination)**
   - Indicates how well the model explains the variability of the dependent variable.
   - Values range from **0 to 1** (higher means better fit).

$$R2 = 1 - SSres/SStot$$

Question 8: What is the purpose of the R-squared metric in regression analysis?
Answer:

The **R-squared (R^2)** metric measures how well a regression model explains the variability of

the dependent variable. It indicates the **goodness of fit** of the model.

The value of R^2 ranges from **0 to 1**:

- **$R^2 = 1$** → Model perfectly explains the data
- **$R^2 = 0$** → Model does not explain any variability
- Higher R^2 → Better model performance

Purpose:

- To show the **proportion of variation in the output (Y)** that is explained by the input variable(s).
- To help compare the performance of different regression models.

In simple words:
R-squared tells how well the regression line fits the data.

Question 9: Write Python code to fit a simple linear regression model using scikit-learn and print the slope and intercept. (Include your Python code and output in the code box below.)
Answer:

```
from sklearn.linear_model import LinearRegression
import numpy as np

#sample data
x = np.array([[1], [2], [3], [4], [5]])  # Independent variable
y = np.array([2, 4, 5, 4, 5])

# Create and train the model
model = LinearRegression()
model.fit(x, y)

# Print the slope and intercept
print("Slope (b1):", model.coef_[0])
print("intercept(b0):", model.intercept_)
```

Output:

```
Slope (b1): 0.6
intercept(b0): 2.2
```

Question 10: How do you interpret the coefficients in a simple linear regression model?

Answer:

In a Simple Linear Regression model:

$y=b0+b1x$

There are two coefficients:

1. **Intercept (b_0):**
    - It represents the value of the dependent variable **y** when the independent variable **x = 0**.
    - It shows where the regression line crosses the Y-axis.
2. **Slope (b_1):**
    - It shows how much **y changes for every one-unit increase in x**.
    - If **b_1 is positive**, y increases as x increases.
    - If **b_1 is negative**, y decreases as x increases.
    - The magnitude of b_1 shows the **strength** of the relationship.

In simple words

- **b_0 tells where the line starts**
- **b_1 tells how steep the line is and the direction of the relationship**

This interpretation helps us understand how strongly the independent variable affects the prediction.