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

Answer:

Simple Linear Regression (SLR) is a statistical method used to model and analyze the relationship between **one independent variable (X)** and **one dependent variable (Y)** by fitting a straight line to the observed data.

The mathematical form of SLR is:

Y=β0+β1X+εY = \beta_0 + \beta_1 X + \varepsilonY=β0+β1X+ε

where:

- β0\beta_0β0 = intercept (value of Y when X = 0)
- β1\beta_1β1 = slope (change in Y for a one-unit change in X)
- ε\varepsilonε = error term (random variation)

Purpose of Simple Linear Regression

The main purposes of SLR are:

1. **To understand the relationship** between an independent variable and a dependent variable.
2. **To predict the value** of the dependent variable based on a given value of the independent variable.
3. **To quantify the effect** of the independent variable on the dependent variable using the slope.
4. **To identify trends** and patterns in data using a simple linear model.

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

Answer:

The key assumptions of Simple Linear Regression (SLR) are the conditions that must be satisfied for the model to produce reliable and valid results:

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

2.Independence
The observations are independent of each other, meaning the error terms are not correlated.

3.Homoscedasticity
The variance of the error terms is constant across all values of the independent variable.

4.Normality of Errors
The residuals (errors) are normally distributed with a mean of zero.

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

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

Answer:

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

$Y=\beta0+\beta1X+\varepsilon$

Explanation of each term:

- **Y**: The dependent (response) variable that we want to predict or explain.
- **X**: The independent (predictor) variable used to explain changes in Y.
- **β0**: The intercept of the regression line; it represents the expected value of Y when X=0.
- **β1**: The slope of the regression line; it shows the change in Y for a one-unit increase in X.
- **ε**: The error term, which captures the effect of factors other than X that influence Y.

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

Answer:

A real-world example of Simple Linear Regression is predicting house prices based on house size.

- Independent variable (X): Size of the house (in square feet)
- Dependent variable (Y): Price of the house

Using simple linear regression, we can model the relationship between house size and price with a straight line. The model helps us understand how much the house price is expected to increase for each additional square foot of area and allows us to predict the price of a house given its size.

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

Answer:

The method of least squares in linear regression is a technique used to estimate the parameters of a regression line by minimizing the differences between the observed values and the values predicted by the model.

In Simple Linear Regression, it works by finding the values of the intercept (β0) and slope (β1) that minimize the sum of the squared residuals, where a residual is the difference between the observed value and the predicted value.

Mathematically, the objective is to minimize:

$\sum(Yi - \hat{Y}i)2$

where:

- Yi = observed values
- $\hat{Y}i$ = predicted values from the regression line

Purpose of the Least Squares Method:

- It ensures the best-fitting straight line through the data points.
- Squaring the residuals gives more weight to larger errors and avoids cancellation of positive and negative errors.
- It provides unique and optimal estimates for regression parameters.

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

Answer:

Logistic Regression is a supervised learning algorithm used for classification problems , where the target variable is categorical (e.g., 0/1, yes/no, spam/not spam).

How it works:

1. It predicts probabilities of an event occurring (e.g., probability of an email being spam).
2. Uses a sigmoid function (S-shaped curve) to map predictions to probabilities between 0 and 1
3. Threshold (usually 0.5) determines the classification (e.g., >0.5 = class 1, <0.5 = class 0).

Linear Regression:

1. Output: Logistic Regression predicts probabilities for classification; Linear Regression predicts continuous values 📈.
2. Use case: Logistic for categorical targets (spam/not spam); Linear for continuous targets (house prices).
3. Function: Logistic uses sigmoid; Linear uses a straight line

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

Answer:

Three common evaluation metrics for regression models are:

1. Mean Absolute Error (MAE)
   It measures the average of the absolute differences between the actual and predicted values. It shows how far predictions are from actual values on average.

2. Mean Squared Error (MSE)
   It calculates the average of the squared differences between actual and predicted values. Larger errors are penalized more due to squaring.

3. Root Mean Squared Error (RMSE)
   It is the square root of MSE and represents the standard deviation of prediction errors. It is expressed in the same units as the target variable, making it easier to interpret.

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

Answer:

The purpose of the R-squared metric in regression analysis is to measure how well a regression model explains the variation in the dependent variable.

- R-square represents the proportion of total variance in the dependent variable that is explained by the independent variable(s).

- Its value ranges from 0 to 1, where a higher value indicates a better model fit.

- An R-square value of 1 means the model explains all the variability in the data, while 0 means it explains none.

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:

```
# Import required libraries
import numpy as np
from sklearn.linear_model import LinearRegression

# Sample data
X = np.array([1, 2, 3, 4, 5]).reshape(-1, 1)  # Independent variable
y = np.array([2, 4, 6, 8, 10])              # Dependent variable

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

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

Output:

```
Slope (Coefficient): 2.0
Intercept: 0.0
```

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

Answer:

In a Simple Linear Regression model, the coefficients are interpreted as follows:

- **Intercept (β0):**
  The intercept represents the expected value of the dependent variable (Y) when the independent variable (X) is zero.

- **Slope (β1):**
  The slope indicates the average change in the dependent variable (Y) for a one-unit increase in the independent variable (X).

    - If $\beta 1 > 0$, Y increases as X increases.

    - If $\beta 1 < 0$, Y decreases as X increases.