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

Simple Linear Regression (SLR) is a supervised learning and statistical modeling technique used to describe and quantify the relationship between one independent variable and one dependent variable. It assumes that this relationship can be represented using a straight-line equation. SLR is widely used in data analysis, economics, engineering, and machine learning to understand trends and patterns in data. The purpose of Simple Linear Regression is multifold. It helps in identifying whether a relationship exists between two variables, understanding the nature of that relationship, predicting future values based on past data, and supporting data-driven decision-making. Because of its simplicity and interpretability, SLR is often the first model used in regression analysis.

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

Simple Linear Regression relies on several assumptions to ensure that the model produces valid and reliable results. First, the linearity assumption states that the relationship between the independent and dependent variables must be linear. Second, independence assumes that observations are independent and errors are not correlated. Third, homoscedasticity means that the variance of residuals remains constant across all values of the independent variable. Fourth, the normality assumption requires that residuals follow a normal distribution. Lastly, the absence of significant outliers is important, as extreme values can heavily influence the regression line.

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

The mathematical equation for Simple Linear Regression is given by $y = \beta■ + \beta■x + \varepsilon$. Here, y represents the dependent variable whose value we want to predict. The variable x is the independent predictor variable. The term $\beta■$ is known as the intercept, which represents the value of y when x equals zero. The coefficient $\beta■$ is the slope of the line and indicates the rate at which y changes with respect to x. The error term ε captures the random noise or unexplained variation in the data.

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

A common real-world application of Simple Linear Regression is predicting a student's exam score based on the number of hours studied. In this case, hours studied is the independent variable, and exam score is the dependent variable. Other practical examples include predicting house prices based on area, estimating salary based on years of experience, forecasting sales based on advertising expenditure, and estimating electricity consumption based on temperature.

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

The method of least squares is a mathematical approach used to estimate the parameters of a linear regression model. It works by minimizing the sum of squared differences between the actual observed values and the values predicted by the regression model. By squaring the errors, larger deviations are penalized more heavily, and all error values become positive. This method ensures that the regression line fits the data points as closely as possible in an overall sense.

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

Logistic Regression is a supervised learning algorithm used primarily for classification tasks. Instead of predicting continuous values, it predicts the probability that a given input belongs to a particular class using a sigmoid (logistic) function. Unlike Linear Regression, which outputs values over an infinite range, Logistic Regression restricts outputs between 0 and 1. Linear Regression is used for predicting quantities, whereas Logistic Regression is used for predicting categories such as yes/no, true/false, or pass/fail.

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

Three commonly used evaluation metrics for regression models are Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). MAE calculates the average absolute difference between predicted and actual values. MSE computes the average of squared errors, giving more weight to large errors. RMSE is the square root of MSE and expresses the error in the same unit as the target variable, making it easier to interpret.

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

R-squared, also known as the coefficient of determination, measures the proportion of variance in the dependent variable that is explained by the independent variable. An R-squared value closer to 1 indicates a strong explanatory power of the model, while a value closer to 0 indicates weak performance. It is commonly used to compare different regression models and evaluate goodness of fit.

Question 9: Write Python code to fit a simple linear regression model using scikit-learn and print the slope and intercept.

In Python, the scikit-learn library provides an efficient implementation of linear regression. The model is trained using sample data, after which the slope and intercept are printed to understand the learned relationship between variables.

```
from sklearn.linear_model import LinearRegression

import numpy as np


# Independent variable

X = np.array([1, 2, 3, 4, 5]).reshape(-1, 1)


# Dependent variable

y = np.array([2, 4, 6, 8, 10])


# Create and train the model

model = LinearRegression()

model.fit(X, y)


# Output results

print("Slope (Coefficient):", model.coef_[0])

print ("Intercept:", model. Intercept_)
```

Slope (Coefficient): 2.0

Intercept: 0.0

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

The coefficients in a simple linear regression model provide meaningful insights into the relationship between variables. The intercept represents the predicted value of the dependent variable when the independent variable is zero. The slope indicates how much the dependent variable is expected to

increase or decrease for every one-unit increase in the independent variable. A positive slope indicates a direct relationship, while a negative slope indicates an inverse relationship.