Regression & Its Evaluation | Assignment Answers

Question 1: What is Simple Linear Regression?

Simple Linear Regression is a statistical method used to model the relationship between a dependent variable and one independent variable. It fits a straight line (regression line) through the data to predict the value of the dependent variable based on the independent variable. The general form is $Y = \beta0 + \beta1X + \varepsilon$, where $\beta0$ is the intercept, $\beta1$ is the slope, and ε is the error term.

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

The key assumptions are:

1. Linearity: The relationship between X and Y is linear.
2. Independence: Observations are independent.
3. Homoscedasticity: Constant variance of residuals.
4. Normality: Residuals are normally distributed.
5. No multicollinearity: (not relevant in simple regression, but in multiple regression).

Question 3: What is heteroscedasticity, and why is it important to address in regression models?

Heteroscedasticity refers to the situation where the variance of residuals is not constant across all levels of the independent variable(s). It violates a key assumption of linear regression and can lead to inefficient estimates and biased standard errors, impacting hypothesis testing. Detecting and correcting heteroscedasticity ensures more reliable model performance.

Question 4: What is Multiple Linear Regression?

Multiple Linear Regression is an extension of simple linear regression that models the relationship between a dependent variable and two or more independent variables. The general form is $Y = \beta0 + \beta1X1 + \beta2X2 + ... + \beta nXn + \varepsilon$. It helps in predicting outcomes when multiple factors influence the response variable.

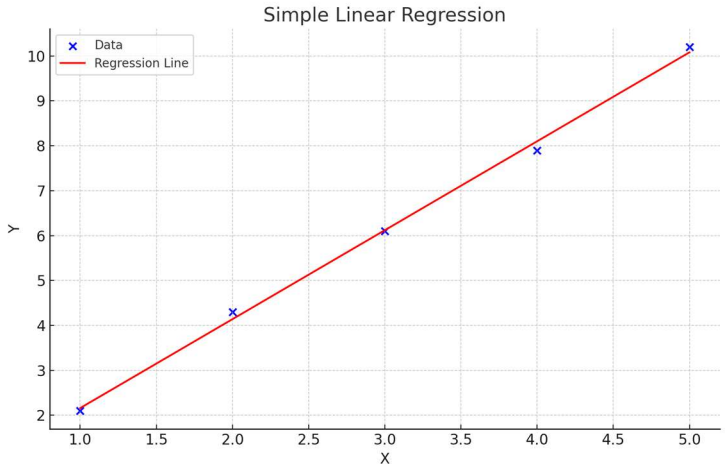
Question 5: What is polynomial regression, and how does it differ from linear regression?

Polynomial Regression is a type of regression that models the relationship between the independent variable and the dependent variable as an nth-degree polynomial. Unlike linear

regression, which fits a straight line, polynomial regression can model curves by including higher-degree terms (e.g., X^2, X^3).

Question 6: Python program for Simple Linear Regression

Python Code Output:

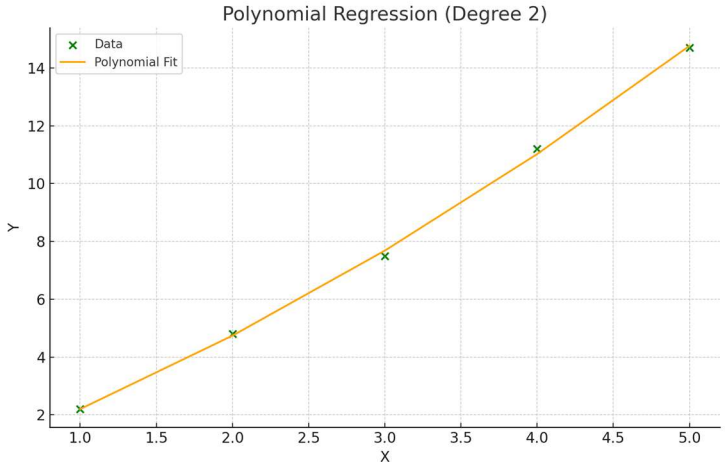


Question 7: Multiple Linear Regression and VIF

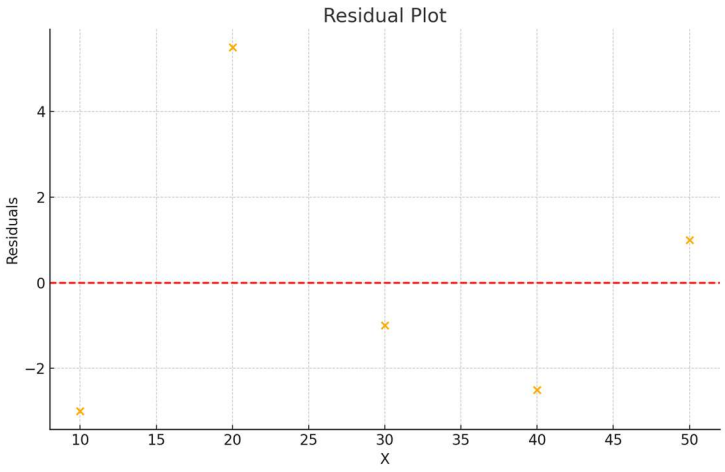
VIF Results:

Feature	VIF
const	34.210526
Area	7.736842
Rooms	7.736842

Question 8: Polynomial Regression



Question 9: Residuals Plot



Question 10: Addressing Heteroscedasticity and Multicollinearity

To address heteroscedasticity:
- Apply transformations like log or square root to stabilize variance.
- Use heteroscedasticity-robust standard errors.

To handle multicollinearity:
- Calculate VIF and remove or combine highly correlated predictors.
- Apply dimensionality reduction techniques like PCA.
- Regularization techniques like Ridge or Lasso Regression can also help mitigate multicollinearity.

These steps improve model reliability and ensure more accurate predictions.