Assignment Code: DA-AG-010

Regression & Its Evaluation | **Assignment**

Instructions: Carefully read each question. Use Google Docs, Microsoft Word, or a similar tool to create a document where you type out each question along with its answer. Save the document as a PDF, and then upload it to the LMS. Please do not zip or archive the files before uploading them. Each question carries 20 marks.

Total Marks: 100

Question 1: What is Simple Linear Regression?

Answer:

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

Answer:

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

Answer:

Question 4: What is Multiple Linear Regression?

Answer:

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

Answer:

Question 6: Implement a Python program to fit a Simple Linear Regression model to the following sample data:

- X = [1, 2, 3, 4, 5]
- Y = [2.1, 4.3, 6.1, 7.9, 10.2]

Plot the regression line over the data points.

(*Include your Python code and output in the code box below.*)

Answer:

Question 7: Fit a **Multiple Linear Regression** model on this sample data:

- Area = [1200, 1500, 1800, 2000]
- Rooms = [2, 3, 3, 4]
- Price = [250000, 300000, 320000, 370000]

Check for multicollinearity using VIF and report the results.
(*Include your Python code and output in the code box below.*)

Answer:

Question 8: Implement **polynomial regression** on the following data:

- X = [1, 2, 3, 4, 5]

- `Y = [2.2, 4.8, 7.5, 11.2, 14.7]`

Fit a **2nd-degree polynomial** and plot the resulting curve.

(*Include your Python code and output in the code box below.*)

Answer:

Question 9: Create a **residuals plot** for a regression model trained on this data:

- `X = [10, 20, 30, 40, 50]`
- `Y = [15, 35, 40, 50, 65]`

Assess heteroscedasticity by examining the spread of residuals.

(*Include your Python code and output in the code box below.*)

Answer:

Question 10: Imagine you are a data scientist working for a real estate company. You need to predict house prices using features like area, number of rooms, and location. However, you detect **heteroscedasticity** and **multicollinearity** in your regression model. Explain the steps you would take to address these issues and ensure a robust model.

Answer: