Assignment : 9

Objective: Analyze the **Boston Housing Dataset** to build and compare the performance of **Multiple Regression, Ridge Regression, and Lasso Regression** models using **Root Mean Squared Error (RMSE)** as the evaluation metric.

1. Data Loading and Exploration

a. Load the **Boston Housing Dataset** from sklearn.datasets or from a CSV file using pandas.
b. Convert it into a Pandas DataFrame and display the first **5 rows** of the dataset.
c. How many **rows and columns** are there in the dataset?
d. Display the **data types** of all columns. Which columns are **categorical**, and which are **numerical**?
e. Are there any **missing values** in the dataset? If yes, handle them appropriately.
f. Display the **basic statistical details** (mean, median, standard deviation, etc.) of all numerical columns

2. Feature Selection

a. Select the following columns as input features (**independent variables**):

- **CRIM** (per capita crime rate)
- **RM** (average number of rooms per dwelling)
- **AGE** (proportion of owner-occupied units built before 1940)
- **DIS** (weighted distances to five Boston employment centers)
- **LSTAT** (percentage lower status of the population)
  b. Use the **MEDV** (Median value of owner-occupied homes in $1000s) as the **target variable (dependent variable)**.

3. Data Splitting

a. Split the dataset into **training and testing sets**, ensuring **80% of the data** is used for training and **20% for testing**.
b. Display the **shapes** of the training and testing datasets.

4. Model Implementation and Training

a. Multiple Regression

i. Train a **Multiple Regression** model using the training data.
ii. Display the **coefficients** and **intercept** of the trained model.

b. Ridge Regression

i. Train a **Ridge Regression** model with an **alpha value of 1.0** using the training data.
ii. Display the **coefficients** and **intercept** of the trained Ridge model.

c. Lasso Regression

i. Train a **Lasso Regression** model with an **alpha value of 1.0** using the training data.
ii. Display the **coefficients** and **intercept** of the trained Lasso model.

5. Model Evaluation

a. Use all three models (**Multiple Regression, Ridge Regression, and Lasso Regression**) to make **predictions** on the test dataset.
b. Calculate and display the following **performance metrics** for each model:

- **Root Mean Squared Error (RMSE)**
- **R^2 Score**
  c. Present the **performance metrics** for the three models in a **comparison table**.

6. Comparison and Analysis

a. Based on the **RMSE and R^2 scores**, which model performs the best?
b. Experiment with **different alpha values** (e.g., **0.1, 5, 10**) for **Ridge and Lasso Regression**. How do these changes affect the **performance metrics** (**RMSE and R^2 Score**)?
c. **Visualize the actual vs. predicted prices** for all three models using **scatter plots**.