Assignment 1.1

Q1. Load energy dataset.

Q2. Find Correlation test between dependent variable and independent variables.(alpha=0.05)

Use H0 :ρ = 0 (There is no correlation between two variables)

H1 : ρ≠ 0 (There is correlation between two variables)

Output

Variable Correlation coefficient p-value

Justify answer.

Q3. Find the Coefficients , T.Values and P.Values.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Coefficients | T.Values | P.Values |
|  |  |  |  |

Find the variables which are negatively co-related.

Q4. Find the linear regression equation using function as well as formula given below

B =

Find whether the residual error is normally distributed or not.

Q5. Apply the Baysian Linear Regression and find the following values in Table

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Coeff. | Quantile(2.5%) | Qunatile(97.5%) |
| Slope |  |  |  |
| X1 |  |  |  |
| X2 |  |  |  |
| - |  |  |  |
| - |  |  |  |
| Xn |  |  |  |

.

Q6. Find the linear regression equation and plot it with linear regression plot .

Q7.Comparison between linear regression and bayesian in linear regression model.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | RMSE | MAD(median(e-median(e)) | MAPE |
| LR |  |  |  |
| LR with Baysian |  |  |  |

MAD (Mean Absolute Deviance) and MAPE (Mean Absolute Percentage Error)

It is calculated as:

**MAPE** = (1/n) \* Σ(|actual – forecast| / |actual|) \* 100

where:

* **Σ** – A symbol that means “sum”
* **n** – Sample size
* **actual** – The actual data value
* **forecast** – The forecasted data value

Conclude your Results.

Assignment 2.1

**Q1. Load** USA housing dataset.

Q2. Find Features of Data-set and Number and Types.

Q3.plot the USA housing histogram for price.

Q4.plot Heat map of the correlation between each of columns.

Q5. Plot Histograms and correlation scatter plots.

Q6.Is Data visualization and Fit for Linear Regression? Check.(Use Correlation of each feature with y)

Q7. Divide the dataset as 40% and 60% into test data and training data, respectively.

Q8.Apply Linear Regression on Training Dataset and Calculate Coefficients.

Q9.Apply Predict() Method to calculate Predicted value for Test Dataset.

Q10.Calculate Error matrix MAE, MSE, and RMSE

Q11.Using distplot to plot predicted values and test values for correctness.(Scatter plot: Y Test versus Prediction)

Q12. Plot the Residual Histogram.

Supplement:

MAE =

MSE =

RMSE =