**Aug 22, 2023**

**Max marks:25**
**Assignment-1**

**Important Instructions:**

- Adding relevant lucid caption below/above to each figure/table is integral part of the exercise.
- Select appropriate font size in the figures.
- Font Style : Times new Roman
- Font Size : 12 for text.
- Your report should carry Content Table/Index with page numbers.
- Appropriately commented MATLAB codes for all the exercises should be appended as Annexure in the report.

---

**Download the data from the following link:**

**https://docs.google.com/spreadsheets/d/1AYXB6icx1psxWSh9DNPXikDzqk1Sbjdf/edit?usp=sharing&ouid=107767635652201956438&rtpof=true&sd=true**

**Dataset Description**

| Name | Size | Bytes | Class | Attributes |
|---|---|---|---|---|
| T | 89x12 | 13521 | table | |

### 1.1 Data preprocessing

a. Check the dataset for the missing values.

b. Check the dataset for the outliers for each relevant column. Use mean and median methods for checking the outliers. Write down the methodology for each methods used with necessary description. Explain the results with the help of plots with proper label and legends.
Ref: https://in.mathworks.com/help/matlab/ref/isoutlier.html#bvlllts-method

c. Remove or Fill the outliers. Provide supporting argument behind the selecting the removing or filling the outliers. Refer the methods for removal and filling given in the link and use any 2 methods and justify your answer.

**d. Detect and Remove:**
https://in.mathworks.com/help/matlab/ref/rmoutliers.html#mw_f334bff5-af8a-45f3-a025-2d4bb1909464

**e. Detect and Replace:**
https://in.mathworks.com/help/matlab/ref/filloutliers.html#bvlnf4n-1-fillmethod

f. Divide the data in input and output where inputs are from column 2 to column 7 and outputs are from column 8 to column 12. Print the summary of the newly created input and output vector/tables.

g. Take any output and normalize the elements using z-score. Also write function for computing the z-score.
Ref: https://in.mathworks.com/help/matlab/ref/double.normalize.html#mw_e6886c44-2923-4074-844a-f3e1a447359d

### 1.2 Linear Regression

a. Create training and testing dataset using holdout.
b. Build multiple linear regression model considering all inputs and each output. Compute the slope and intercept.
c. Find out the $R^2$ for all the models.
d. Show the predicted and actual output with absolute error as the bar plot in the same figure. Ensure proper legends and labelling.

### 1.3 Support vector regression (SVR)

a. Create training and testing dataset using holdout.
b. Build support vector regression model for predicting each output. Consider all the inputs.
c. Find out the $R^2$ for all the models.
d. Show the predicted and actual output with absolute error as the bar plot in the same figure for each model. Ensure proper legends and labelling.
e. Repeat SVR model development (from b to d) using cross validation.
f. Discuss the results obtained using holdout and cross-validation.

### 1.4 Decision Tree Regression (DTR)

a. Build DTR model for predicting each output. Consider all the inputs.
b. Find out the $R^2$ for all the models.

c. Show the predicted and actual output with absolute error as the bar plot in the same figure for each model. Ensure proper legends and labelling.

### 1.5 Performance comparison

a. Create bar plot for RMSE and Coefficient of determination.

b. Write detailed conclusion on the performance of each model using the bar plots.