Department of Electronic and Telecommunication Engineering
University Of Moratuwa

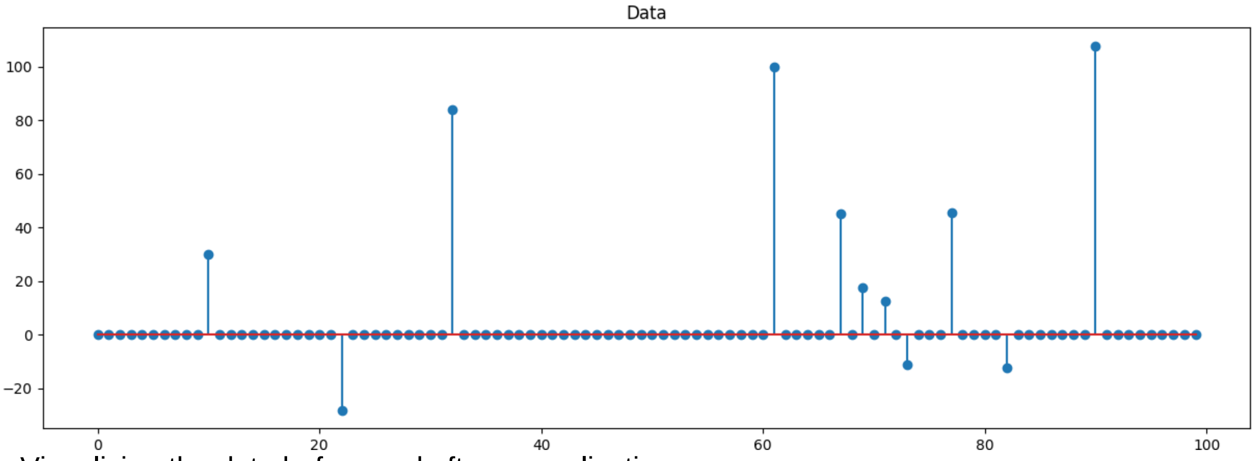
Assignment 01

Index No.	Name
200702H	P.D.G.U.M.B. WEERASINGHE

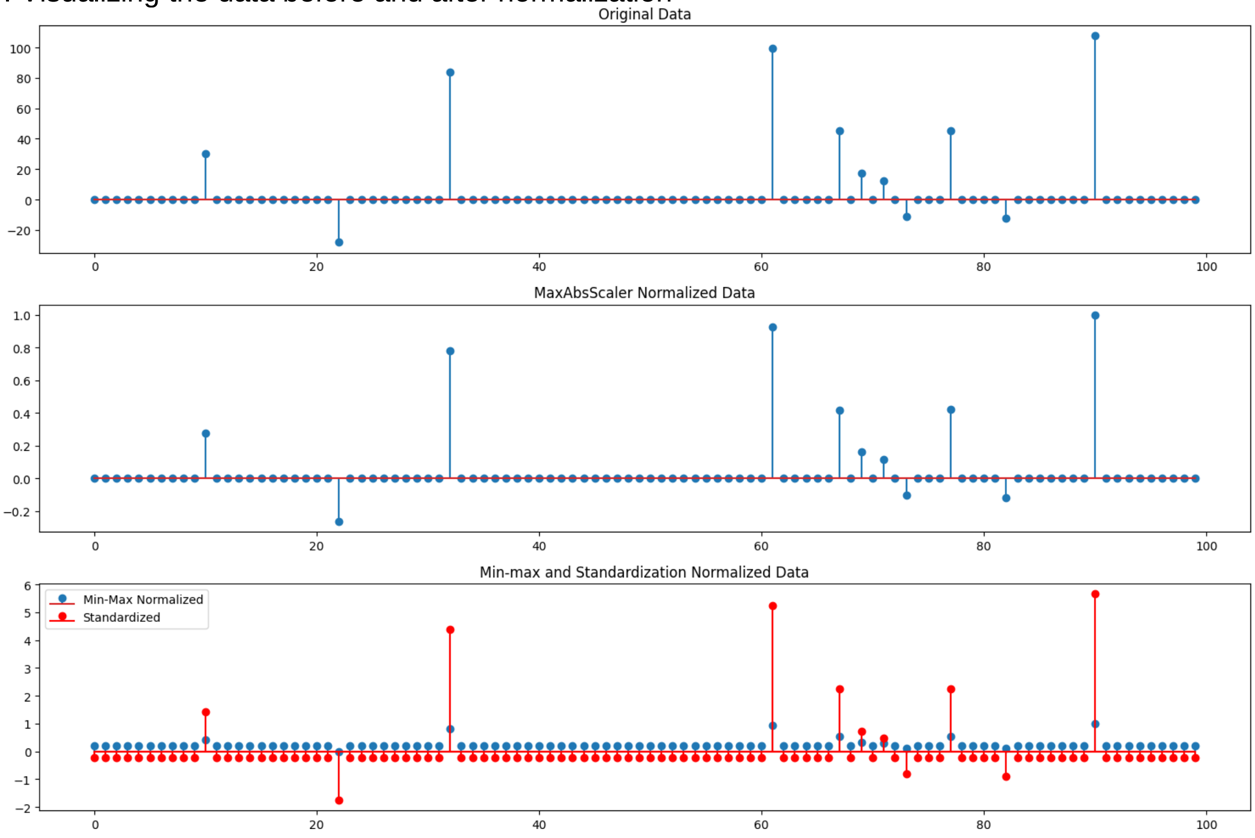
This assignment is submitted as a partial fulfillment of the module EN3150 – Pattern Recognition

1. Data preprocessing

1) Plotted signal



. Visualizing the data before and after normalization



5)
```
Number of Non-Zero Elements:
    Before the normalization
                        Original Data: 11
```

```
After the normalization
                   MaxAbsScaler Normalized Data: 11
                   Min-Max Normalized Data: 99
                   Standardized Data: 100
```

6)

factor	**MaxAbsScaler**	**Min-Max Scaling**	**Standardization (Z-score)**
Scaling Range	[-1, 1]	[0, 1]	centers the data around 0 with a standard deviation of 1
Distribution Shape	Doesn't change the shape of the data's distribution. It only scales the values while maintaining their proportionality.	Doesn't change the shape	centers it around the mean. It may lead to a more symmetric distribution.
Use Case	-When it is wanted to preserve the relationships between features and no need to change the overall distribution shape. -When dealing with sparse data or want to ensure that all features have values within a bounded range.	-When it is needed to bound data within a specific range, such as [0, 1]. -Where algorithms are sensitive to the scale of input features.	-When it needs to ensure that the data has a mean of 0 and a standard deviation of 1. -Used in statistical analysis and when applying algorithms like Principal Component Analysis (PCA) or clustering.
Other special facts	It preserves the relative scale between features	It preserves the order of values, meaning that if one data point is originally larger than another, it will still be larger after scaling.	It scales the data based on the standard deviation, which makes it suitable for algorithms that assume a Gaussian (normal) distribution.

7)

Factor	**MaxAbsScaler**	**Min-Max Scaling**	**Standardization (Z-score)**
Distribution	Does not change the distribution's shape. It scales the data linearly,	Min-max scaling does not change the shape of the distribution but	Standardization centers the data around 0 and scales it based on the standard deviation. This

	maintaining the relative distance between data points.	scales it to fit within the range [0, 1].	can make the distribution more symmetric, resembling a standard normal distribution (mean of 0 and standard deviation of 1).
Structure	It preserves the structure of the data in terms of feature relationships.	It preserves the order of values, maintaining relationships between features.	It centers the data around the mean, which can affect the relationships between features.
Scale	The scale is bounded between -1 and 1. If data has outliers, they might be heavily compressed.	The scale is bounded between 0 and 1, which can be helpful when algorithms are sensitive to feature scales. However, it may not handle outliers well.	The scale is unbounded, but it typically has a mean of 0 and a standard deviation of 1.

- MaxAbsScaler retains the distribution and structure of the data while scaling it to a common range.
- Min-Max scaling shifts the data to a common range, but it doesn't preserve the original distribution and structure.
- Standardization centers the data around mean 0 and scales it, which may change the distribution and structure.

For this kind of data, MaxAbsScaler is recommended because it scales the data while preserving its distribution and structure. This is particularly useful when the original distribution and structure of the data are important for downstream analysis

2. Linear Regression on Real World Data

Output-

```
sample   index   TV      radio   newspaper   sales
0        1       230.1    37.8    69.2        22.1
1        2       44.5     39.3    45.1        10.4
2        3       17.2     45.9    69.3         9.3
3        4       151.5    41.3    58.5        18.5
4        5       180.8    10.8    58.4        12.9
```

-Question 4-

```
RSS (Training): 432.8207076930262
RSS (Testing): 126.96389415904413
RSE (Training): 1.6603673672483137
RSE (Testing): 1.8524191207426803
MSE (Training): 2.705129423081414
MSE (Testing): 3.1740973539761033
R2 (Training): 0.8957008271017818
R2 (Testing): 0.899438024100912
T-Statistics: TV          28.543587
radio        19.517950
newspaper     0.391761
dtype: float64
P-Values: TV          8.166150e-64
radio        1.016134e-43
newspaper    6.957694e-01
dtype: float64

coefficients [0.04472952 0.18919505 0.00276111]
```

-Question 5-

Significant Features: ['TV', 'radio']

-Question 6-

Highest Contributor to Sales: radio

-Question 7-

Hypothetical Sales with $25,000 for TV and $25,000 for Radio (No Newspaper):
5851.093359915444

Discussion

5)

```
 P-Values:
TV             8.166150e-64
radio          1.016134e-43
newspaper      6.957694e-01

since all p- values are lesser than 0.05 we can determine that there is a
relationship between advertising budgets and sales
```

Relationship between advertising budgets and sales can be discussed based on the coefficients obtained from the linear regression model and the provided dataset:

1. **TV Advertising**:
   - The coefficient for TV advertising is $0.0447. This positive coefficient indicates that for every additional $1,000 spent on TV advertising, there is an estimated

increase in sales of approximately $44.70. This suggests that TV advertising has a moderately positive impact on sales in this context.

2. **Radio Advertising**:
   - Justification: The coefficient for radio advertising is 0.1892. This positive coefficient is higher than that of TV advertising, indicating that for every additional $1,000 spent on radio advertising, there is an estimated increase in sales of approximately $189.20. This suggests that radio advertising has a stronger positive impact on sales compared to TV advertising.
3. **Newspaper Advertising**:
   - Justification: The coefficient for newspaper advertising is 0.00276. This coefficient is very close to zero and negative, indicating that for every additional $1,000 spent on newspaper advertising, there is an estimated minimal decrease in sales of approximately $2.76. This suggests that newspaper advertising has almost no significant impact on sales or may even have a slightly negative effect.

In summary, the relationship between advertising budgets and sales in this case can be characterized as follows:

- Both TV and radio advertising budgets positively contribute to sales, with radio advertising having a stronger impact per dollar spent.
- Newspaper advertising budgets show little to no positive influence on sales and might even have a slightly negative effect.

Therefore, allocating more budget to radio advertising, followed by TV advertising, appears to be a more effective strategy for increasing sales, while newspaper advertising may not be a worthwhile investment in this context.

6) – highest contributor – Radio

7)

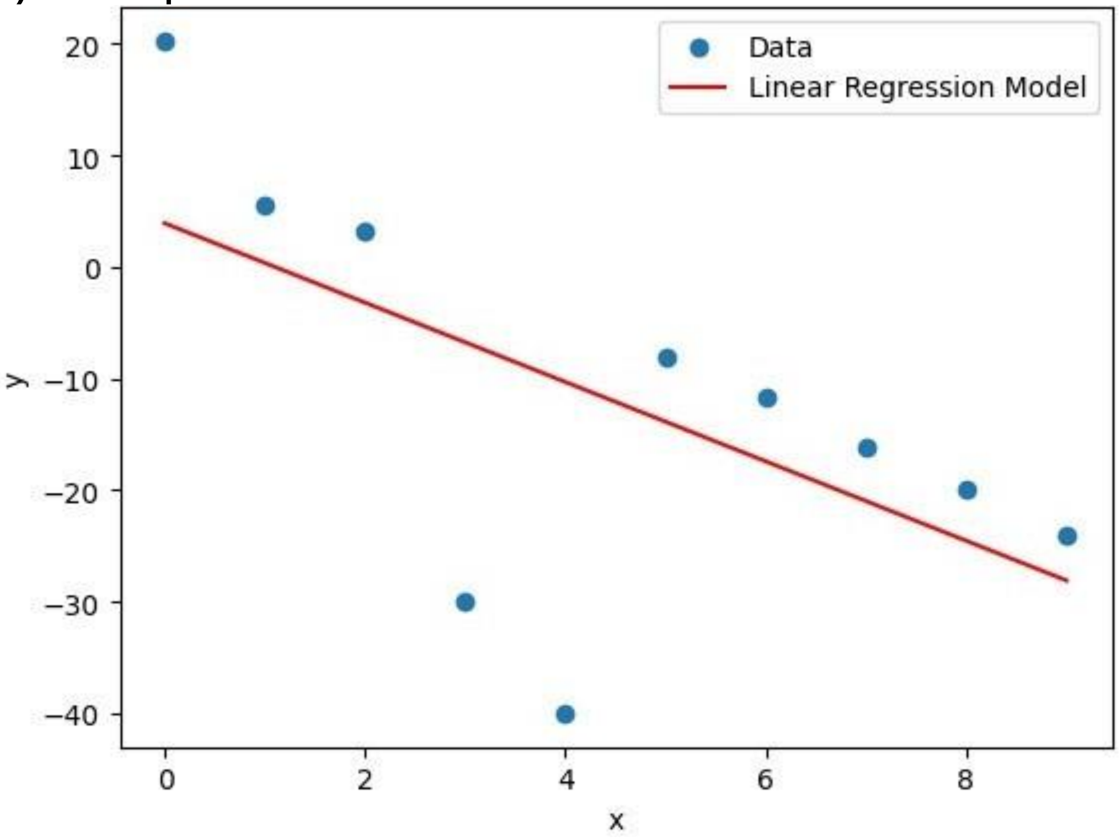
Based on the analysis using the trained linear regression model:

1. Allocating $25,000 to both TV and radio advertising individually (total budget of $50,000) yields higher sales compared to investing $50,000 in either television or radio advertising individually.

This suggests that a balanced allocation of budget between TV and radio advertising is more effective in driving sales than putting all the budget into one channel. However, it's important to note that this conclusion is based on the specific data and model used for analysis, and real-world results may vary depending on various factors such as market conditions and audience preferences.

3.Linear Regression Impact on Outliers

2)- Scatter plot



4)- Loss Function values(output)

```
Loss for Model 1: 0.09956155329216698
Loss for Model 2 (Linear Regression Model): 0.09947190999487436
```

Linear regression model(model 2) has a less loss compared to model 1

5) – Determine the Most Suitable Model

In this case, it is needed the model with the lowest loss, which indicates a better fit to the data. So, Model 2 (Linear Regression Model) is the most suitable model for this dataset because it has a lower loss compared to Model 1.

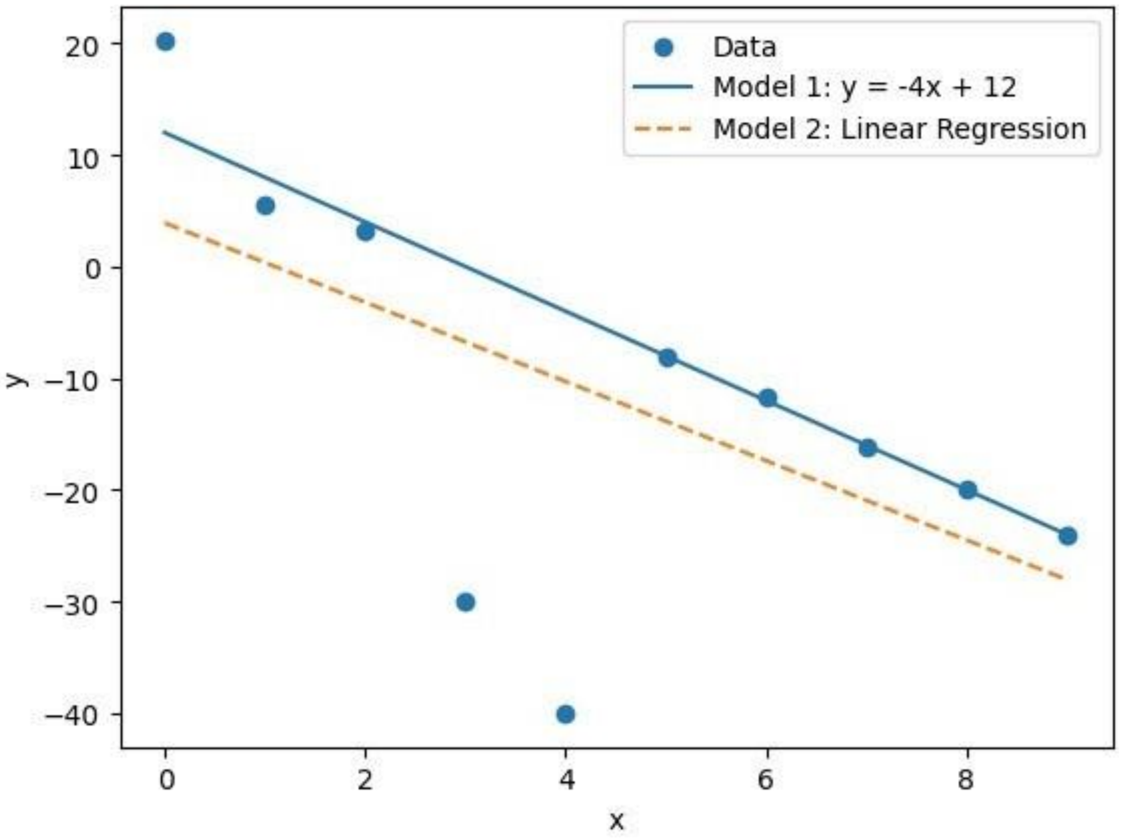
6) – How Does the Robust Estimator Reduce Impact of Outliers

The robust estimator reduces the impact of outliers by giving less weight to data points that have large residuals (i.e., the difference between the observed and predicted values). In the loss function,

$$L(\theta,\beta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(y_i-\hat{y}_i)^2}{(y_i-\hat{y}_i)^2+\beta^2}\right).$$

, the term $(y_i-\hat{y}_i)^2$ measures the squared residual, and the term $(y_i-\hat{y}_i)^2+\beta^2$ is a combination of squared residual and β^2, which is a constant. By using this formulation, the loss function assigns a smaller weight to data points with larger residuals, effectively reducing the influence of outliers on the model parameters.

7) – **plotting models**



8) – **Impact on β in the Loss Function**

The β term in the loss function,

$$L(\theta,\beta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(y_i-\hat{y}_i)^2}{(y_i-\hat{y}_i)^2+\beta^2}\right).$$

acts as a regularization parameter. It controls the impact of outliers on the loss. A larger β value will make the loss function less sensitive to outliers, as it increases the denominator and reduces the weight of the outliers. Conversely, a smaller β value makes the loss function more sensitive to outliers, as it increases the weight of the residuals. The choice of β depends on the specific problem and the desired level of robustness to outliers.