Department of Electronic and Telecommunication Engineering
University of Moratuwa

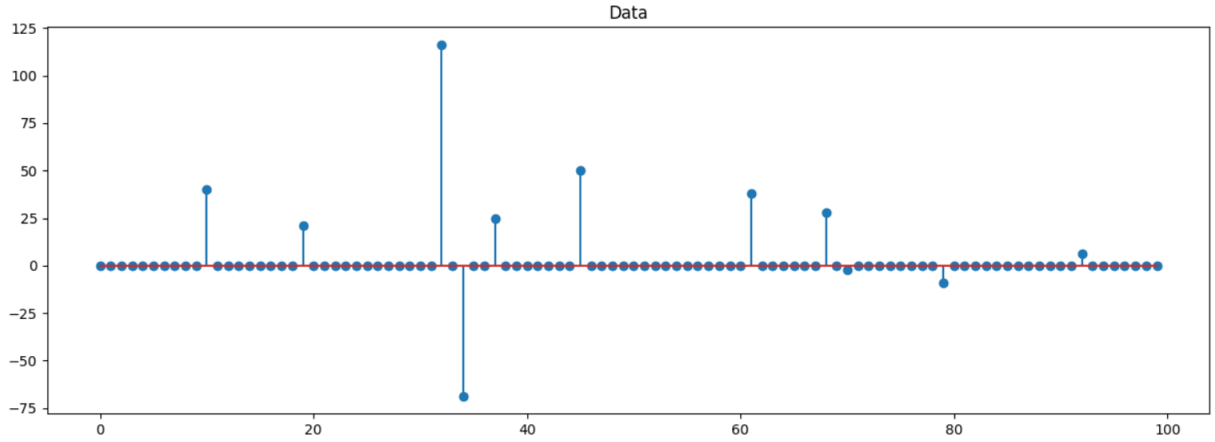
BM3150 – Pattern recognition

Learning from data and related challenges and linear models for regression.

200003P
T.L Abeygunathilaka

1. Data Preprocessing

Index No: 200003P

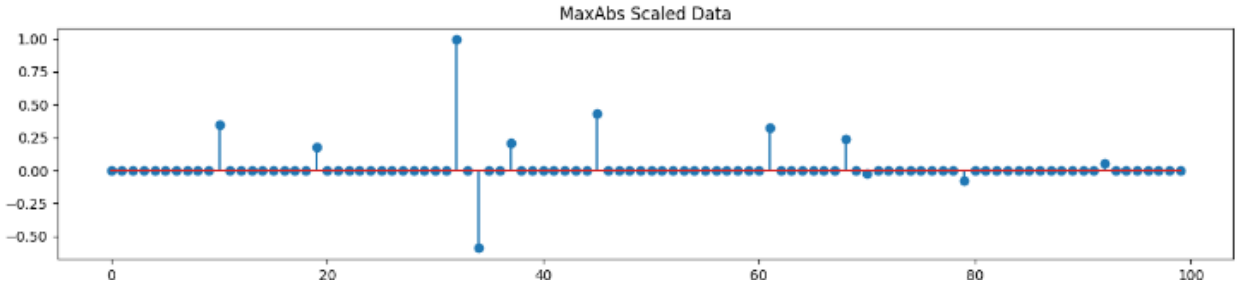


Normalization methods

1. Max Abs Scaler

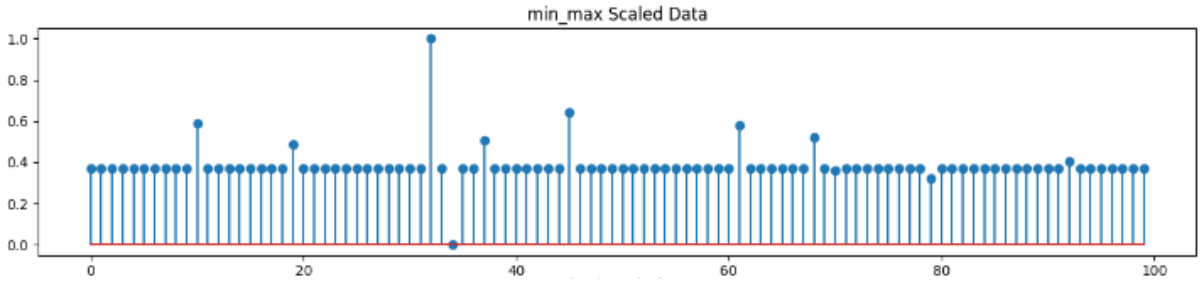
$$X_i' = \frac{X_i}{abs(X_{max})}$$

Sklearn.preprocessing library is used to determine the Max Abs Scaler



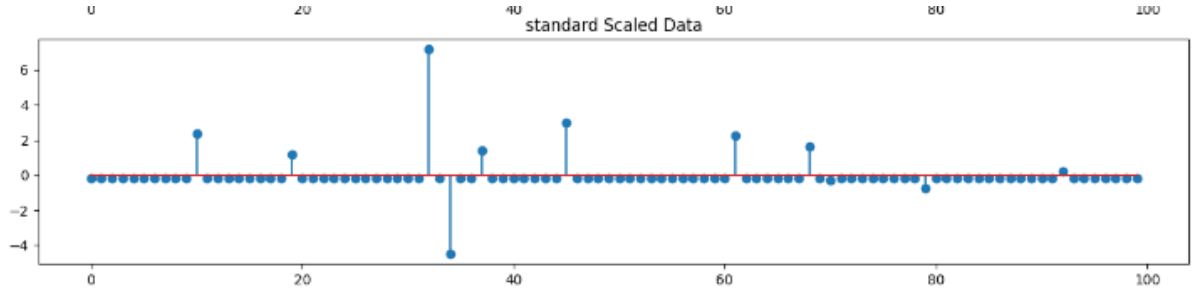
2. Min Max Scaler

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$



3. Standarlization scaler

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma}.$$



- **How many non-zero elements in the data before the normalization and after the normalization.**
  - Before Normalization - 11 points
  - After MaxAbsScaler – 11 points
  - After Min-Max Scaler – 99 points
  - After standard Normalization – 100 points

- **Compare how each normalization method scales the data and its impact on structure of the data.**
  - Max Abs scaler

$$X_i' = \frac{X_i}{abs(X_{max})}$$

MaxAbs Scaler linearly scales the data in such a way that the resulting scaled dataset falls within the range of [-1, 1]. This scaling method retains the original sign (positive or negative) of each data point and does not change the overall distribution of the data. As a consequence, MaxAbs Scaler does not

mitigate or reduce the influence of outliers in the dataset; it simply scales all data points based on the maximum absolute value to maintain their relative proportions.

- Min Max scaler

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

It transforms each data point so that it falls within the range [0, 1], with 0 corresponding to the minimum value in the dataset and 1 corresponding to the maximum value. This scaling method ensures that the scaled data retains the same proportionality as the original data but is bound within a specific range.

- Standard Scaler

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma}.$$

After applying standard scaling the mean of the scaled dataset indeed becomes 0, and the standard deviation becomes 1. This transformation centers the dataset around zero and scales it in such a way that it has a standard deviation of 1.

- **Discuss the effects of each normalization method on the data's distribution, structure, and scale. Which normalization approach you recommend for this kind of data and what is the reason behind this?**

In this context, where we don't know about the outliers because the signal is sparse and the primary objective is to scale the data for easier processing.

MaxAbs Scaler scales the data into the range of [-1, 1], which can be a suitable choice when we want to maintain the relative proportions of the data without introducing any distortions due to scaling. Additionally, the fact that MaxAbs Scaler preserves zero values is particularly advantageous in cases where the dataset contains many

zeros, as it retains the sparsity and can reduce computational costs, especially when dealing with sparse matrices or datasets.

Therefore, the suitable scaler is <u>Max Abs Scaler</u>

2. Linear regression on real world data

	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

Linear regression model for 3 input data set

$$y = w_0 + w_1x_1 + w_2x_2 + w_3x_3$$

x_1 – TV, x_2 – Radio, x_3 – Newspaper, y – sales

w1 = 0.04458402

w2 = 0.19649703

w3 = -0.00278146

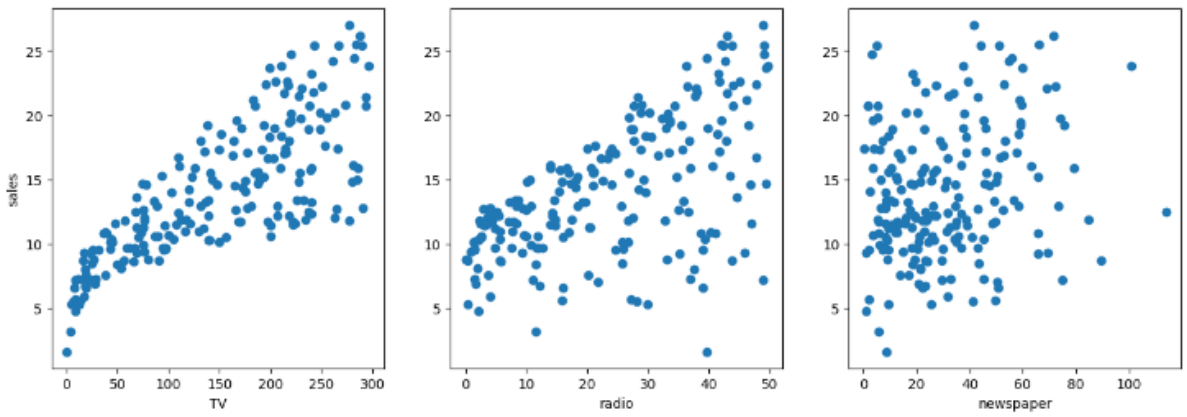
w0 = 2.9948

Parameter	Value for train set	Values for test set
RSS	385.0903609310249	176.08473165798722
RSE	1.571154974751395	2.2116153702791994
MSE	2.406814755818906	4.40211829144968
R2	0.9067114990146383	0.8601145185017869
Std Error	W0- TV 0.020020 radio 0.046569 newspaper 0.051812 w1= TV 0.000001 radio 0.000007 newspaper 0.000008	w0= TV 0.143985 radio 0.346286 newspaper 0.407027 w1= TV 0.000009 radio 0.000060 newspaper 0.000076
T statistics	w1 = TV 42.980703 radio 75.264916 newspaper -0.986517	tw1= TV 15.152338 radio 25.351631 newspaper -0.320080

	tw0 =TV 21.166743 radio 13.878127 newspaper 13.157233	w0 = TV 7.892645 radio 5.089365 newspaper 4.694286
P Value	w1 – [1.11651611e-088 8.02602658e-125 1.62703482e- 001] w0 – [5.11297874e-48 2.46737328e-29 2.24669597e-27]	w1 – [1.68706802e-17 7.48382493e-25 3.75378883e-01] w0 – [1.14913586e-09 5.71526361e-06 1.90458389e-05]

$w1$ and $w2$ exhibit very low p-values, signifying their statistical significance. This implies that there is a meaningful connection between the money spent on TV and radio advertising and the resulting sales. Essentially, changes in these advertising budgets have a substantial impact on sales.

Conversely, the p-value for $w3$, related to newspaper advertising spending, fails to reach the conventional levels of statistical significance, such as 0.05 or 0.01. This indicates that there isn't strong enough evidence to reject the Null Hypothesis ($H0$) for $w3$. In simpler terms, it suggests that variations in the newspaper advertising budget do not appear to significantly affect sales.

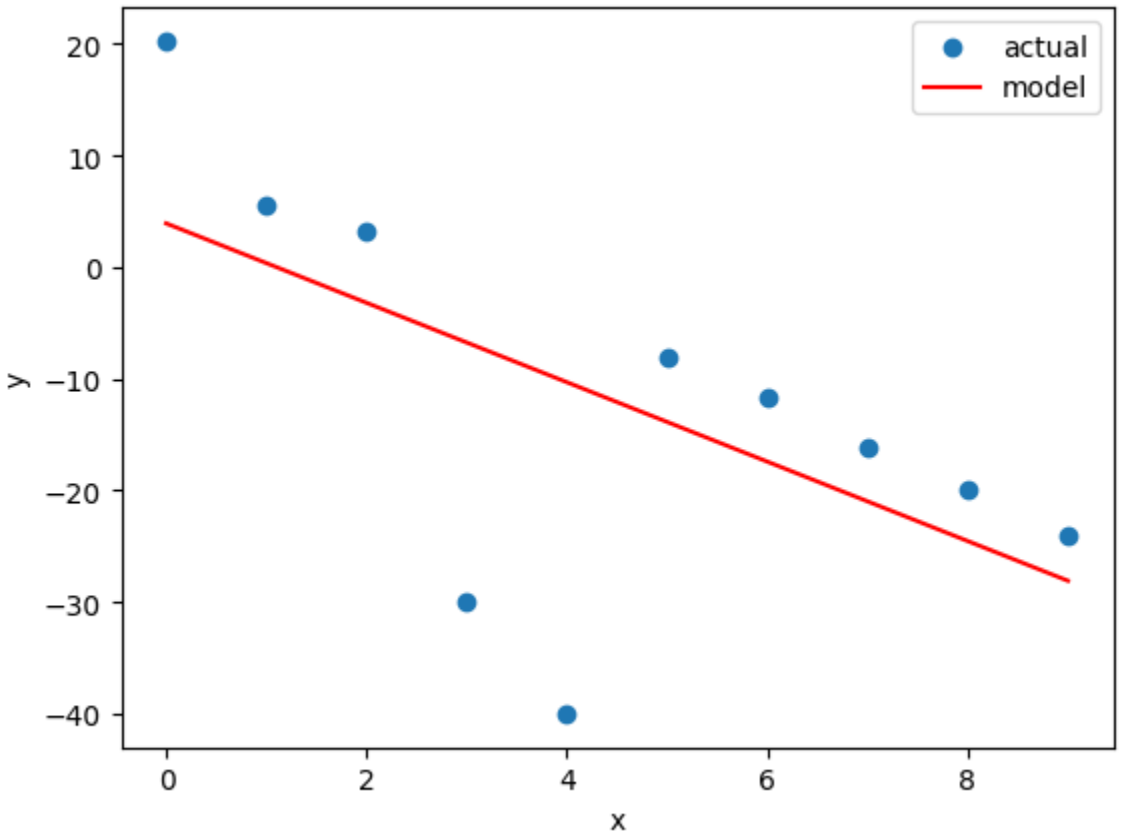



```
predict 1 [25,25,0]:  [9.02191939]
predict 2 [50,0,0]:  [5.22409404]
predict 3 [0,50,0]:  [12.81974474]
```

Based on above predictions 3rd method is the best one

Spend $50000 on radio makes the best outcome.

3. Linear regression impact on outliers



Regression model:

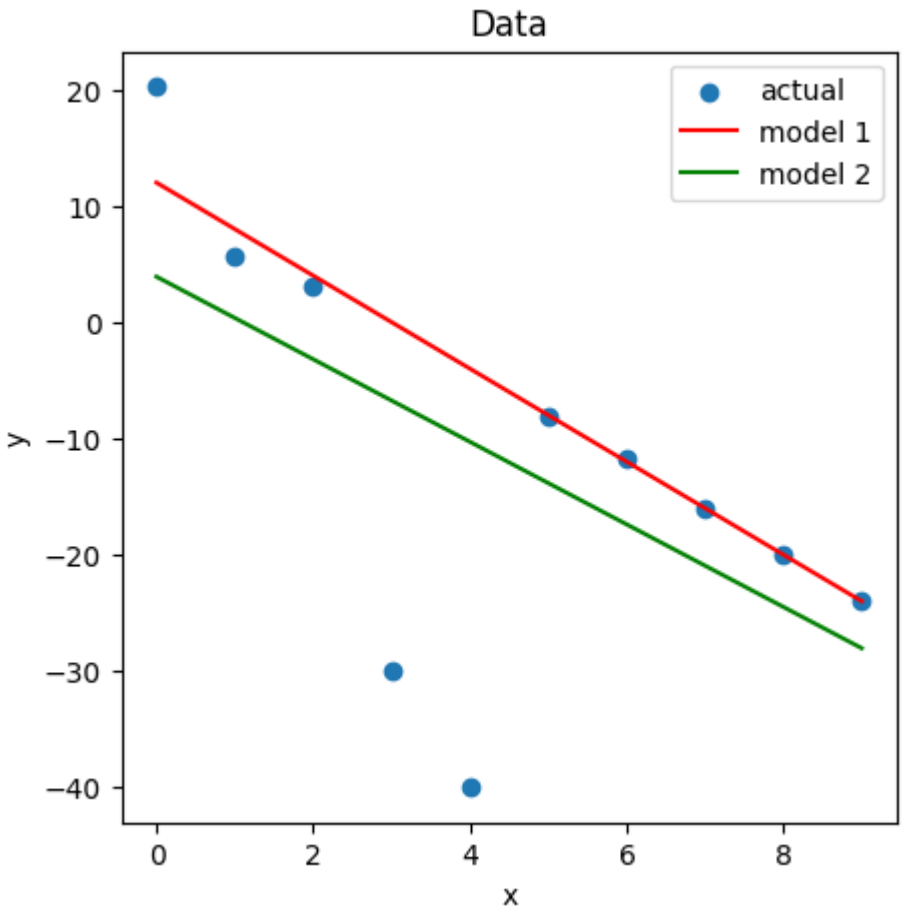
```
coefficients [-3.55727273]
intercept 3.916727272727277
```

Loss function for calculate the loss

```
def loss(w, beta, x,y,N):
    # N = len(x)
    loss_val = 0
    for i in range(N):
        loss_val += ((y.iloc[i]-(w[0] + w[1]*x.iloc[i]))**2)/(((y.iloc[i]-(w[0] +
w[1]*x.iloc[i]))**2)+beta**2)
    loss_val/=N
    return loss_val
```

loss for model 1: `0.435416`

loss for model 2: `0.972847`



The parameter β plays a crucial role in down weighting the impact of outliers in a model. When β is set to a high value, the loss function approaches something like the Mean Squared Error (MSE), because the term $(yi - ypredicti)^2 + \beta^2$ becomes approximately equivalent to β^2. Conversely, using very small β values doesn't effectively down weight outliers, as $(yi - ypredicti)^2 + \beta^2$ is nearly equal to $(yi - ypredicti)^2$.

Therefore, selecting an appropriate β value is critical and should be based on the specific characteristics of the data and the model's predictions. For instance, if the values of $(yi - ypredicti)^2$ fall within a narrow range like (-10, 10), then using a β value of 1 may be suitable. However, if the values of $(yi - ypredicti)^2$ are within a wider range like (-100, 100), then a β value of 1 may not effectively address the presence of outliers.