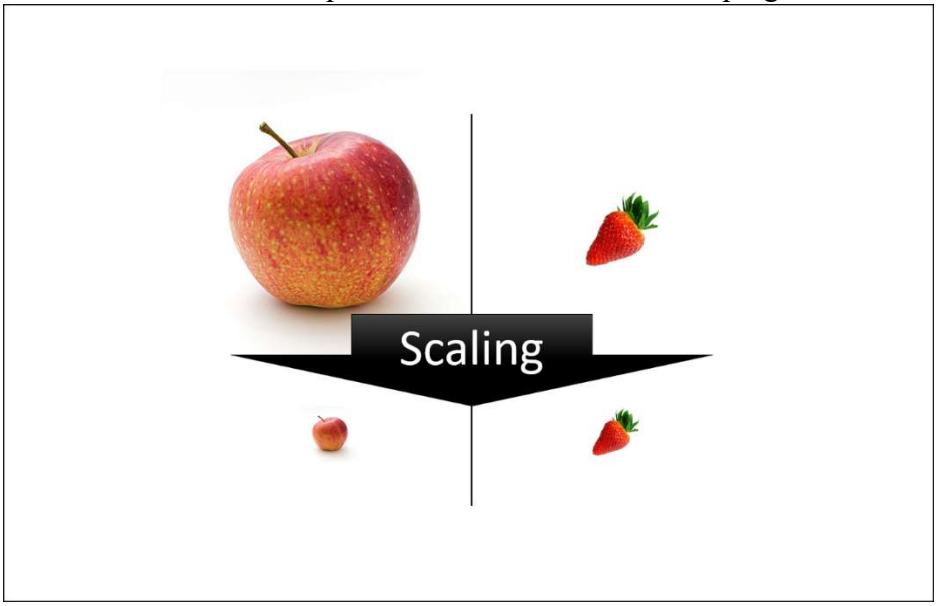
Feature Scaling

In this blog let's try to understand the Normalization and Standardization methods in Feature Scaling manually.

- What is Feature Scaling?
  - ✓ Feature scaling is used to normalize the data in to the particular range of same scale to make it easier to predict the model while developing it.



- Two types of Feature Scaling:
  1. Normalization
  2. Standardization

Let's try to understand both methods by using this simple data set

Data set x = {1,2,3,4,5}

1.Normalization

- ➢ In feature scaling, Normalization can be done by using the minimax Scaler function.
- ➢ It will normalize our data into the range of 0 to 1.

Formular for minimax scaler:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Let’s try to scale down our data into normalized data by using this scaler formula

Our x values are = {1,2,3,4,5}

Implement these values into the formula and get the new values of x`

X` = xi-min(x)/max(x)-min(x)

1.) If (xi = 1)

X` = 1-1/5-1 = 0/4 = 0

2.) If (xi = 2)

X` = 2-1/5-1 = 1 / 4 = 0.25

3.) If (xi = 3)

X` = 3-1/5-1 = 2/4= 0.5

4.) If (xi = 4)

X` = 4-1/5-1 = 3/4 = 0.75

5.) If (xi = 5)

X` - 5-1/5-1 = 4/4 = 1

So, finally we have converted our data set into normalized.

X` = {0, 0.25, 0.5, 0.75, 1}

We can see all the data of X` are in between the range of (0 and 1) so it means we have done the proper implementation of normalization method to scale down the data.

2. Standardization

- Standardization used to get the different scale values from different variables and put it all in the same scale. it also seems like normalization, but here we can’t able to set the limit of the scale values it will automatically scaled based on the input data.
- This process can be done by using the Z score Formula
- In standardization the Mean(μ) value will be 0 and standard deviation (σ) will be 1.

Standardization Z Score Formula:

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean
σ = Standard Deviation

Let's try to scale down our data into Standardized by using the z score formular

Our x values are ε= {1,2,3,4,5}

Implement these values in to the formula and get the new values of Z

I. Find out the Mean(μ)
   - Mean is nothing but average of given data

   Formula to finds the mean is

   μ = Ɛx/N

   N = Total count of values in the data set = 5

   μ = 1+2+3+4+5/5 = 15/5 = 3

   μ = 3.

II. Find the Standard deviation(σ)

   Formula to finds the sd (σ)

   $$\sigma = \sqrt{\frac{\Sigma (x - u)^2}{N}}$$

   σ = √ (1-3) ² +(2-3) ²+(3-3) ²+(4-3) ²+(5-3) ²/5
   = √ (-2) ² +(-1) ² + (0) ² +(1) ² +(2) ²/5
   = √ 4+1+0+1+4/5
   = √10/5

$\sigma = \sqrt{2}$

$\sigma = 1.414$

We got all the required values to find the Z score.

Let's try to find it out Z value by implanting this all values in the Z score formula.

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

i.) If x = 1

Z = 1-3/1.414

= -2/1.414

Z = - 1.414

ii.) If x = 2

Z = 2-3/1.414

= -1/1.414

Z = - 0.7072

iii.) If x = 3

Z = 3-3/1.414

= 0/1.414

Z = 0

iv.) If x = 4

Z = 4-3/1.414

= 1/1.414

Z = 0.7072

v.) If x = 5

Z = 5-3/1.414

= 2/1.414

Z = 1.414

Z = {-1.414, -0.7072, 0, 0.7072, 1.4141}

So, finally we have converted our data set into Standardized

Z = {-1.414, -0.7072, 0, 0.7072, 1.4141}

Finally, we can conclude that:

The data set is x = {1,2,3,4,5}

1. After Normalization the values are X` = {0, 0.25, 0.5, 0.75, 1}
2. After standardization the values are Z= {-1.414,-0.7072,0,0.7072,1.4141}