DATA SCIENCE WITH PYTHON : DATA NORMALIZATION #1857

Presented by: Deepthi M

Batch Number: 05

Serial Number: 172

What is normalization?

Normalization is a technique that updates the values and distributes the values in the same range through out the dataset.

Why normalization?

Lets us imagine a dataset that have the columns age and their income.

Age	Income
25	1800000
30	1000000
78	2500000
67	2000000

When we apply model it becomes a bias estimation.
Hence, we need to convert the distribution of values to the values which has the same range.
One of the famous techniques is normalization.

Types of normalization:
min-max normalization
Formula:

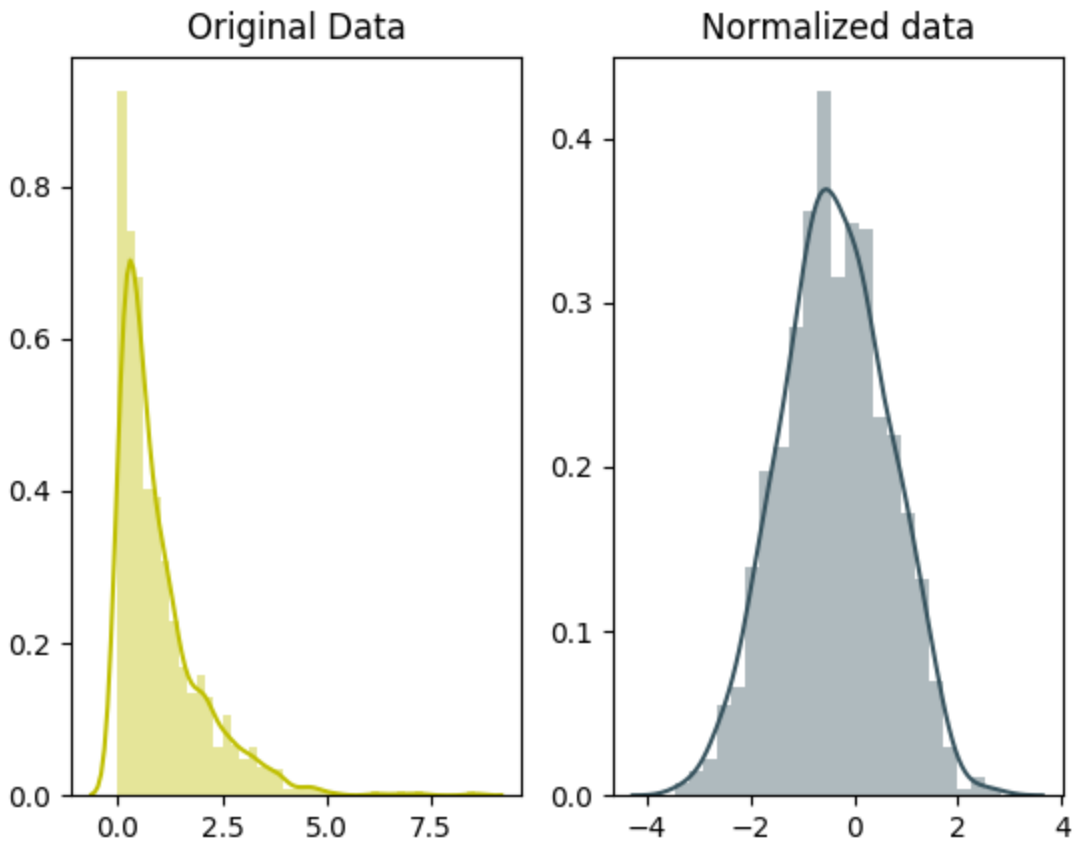
$$x' = \frac{x - min(x)}{max(x) - min(x)}$$

Mean normalization
Formula:

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

Z-score normalization
Formula:

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


Original Data
Normalized data
0.0
0.2
0.4
0.6
0.8
0.0
2.5
5.0
7.5
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Advantages:

- It coverts the values in the dataset to the same distribution.
- It helps to avoid the bias estimation.

Disadvantages:

- It also coverts the outliers to the same distribution which might impact the model performance.