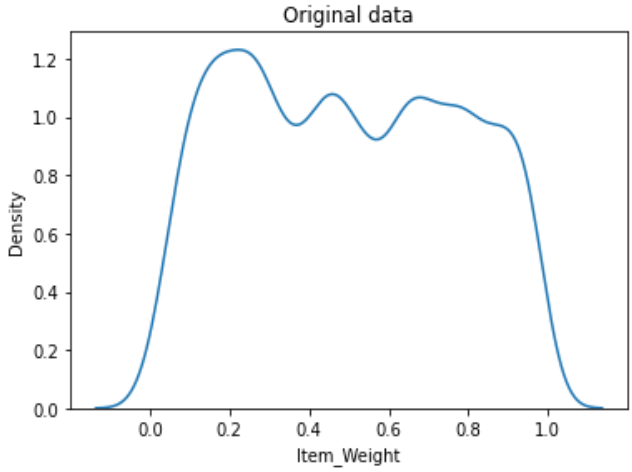
Data Transformation

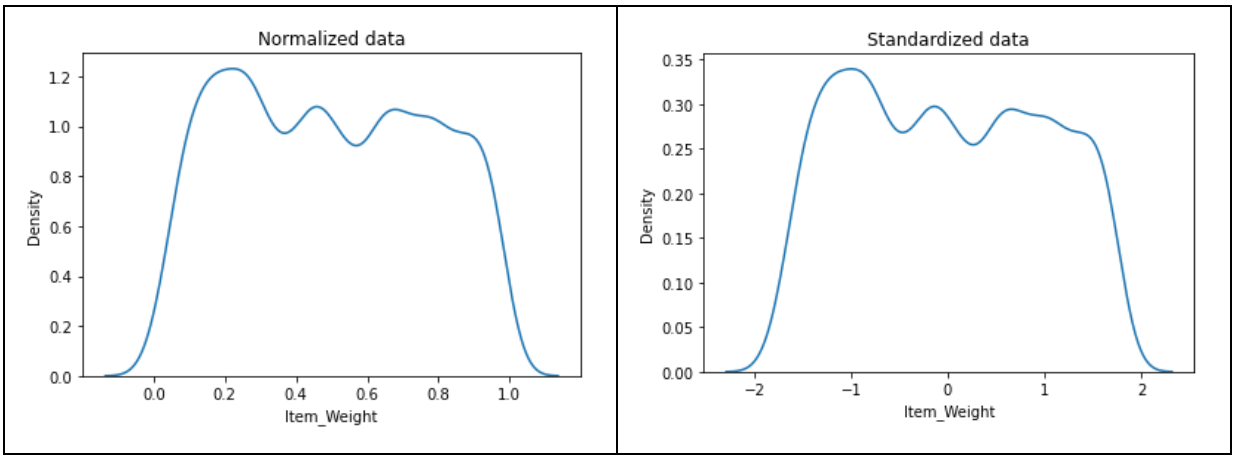
Normalization or Standardization

Normalization is good to use when you know that the distribution of your data does not follow a Normal (Gaussian) distribution. This can be useful in algorithms that do not assume any distribution of the data like K-Nearest Neighbors and Neural Networks.

Standardization, on the other hand, can be helpful in cases where the data follows a Normal (Gaussian) distribution. However, this does not have to be necessarily true. Also, unlike normalization, standardization does not have a bounding range. So, even if you have outliers in your data, they will not be affected by standardization.



However, the choice of using normalization or standardization will **depend on your problem** and the machine learning algorithm you are using.
There is ***no hard and fast rule*** to tell you when to normalize or standardize your data.



You can always start by fitting your model to raw, normalized and standardized data and **compare the performance for the best results**. With any of these techniques as we can see the density/pattern of the data points remains the same. Only the scale of the data points gets changed after applying these Scaling Methods.