**Table 1.8.** Difference between standartization and normalization (Practice with əɞ 2022)

|  |  |
| --- | --- |
| **Normalization** | **Standartization** |
| Minimum and maximum value of features are used for scaling | Mean and standard deviation is used for scaling |
| It is used when features are of different scales | It is used when we want to ensure zero mean and unit standard deviation |
| Scales values between or | It is not bounded to a certain range |
| It is really affected by outliers. | It is much less affected by outliers |
| Scikit-Learn provides a transformer called *MinMaxScaler* for Normalization | Scikit-Learn provides a transformer called *StandardScaler* for standardization |
| This transformation squishes the n-dimensional data into an n-dimensional unit hypercube | It translates the data to the mean vector of original data to the origin and squishes or expands |
| It is useful when there is no knowledge about the distribution | It is useful when the feature distribution is Normal or Gaussian |
| It is a often called as Scaling Normalization | It is a often called as *Z-Score* Normalization |

Normalization (scaling normalization) is a rescaling of the data from the original range so that all values are within the range of 0 and 1. Normalization requires to know or be able to accurately estimate the minimum and maximum observable values from the available data. Normalization typically rescales the values into a range of [0,1] and means changing the shape of the distribution of the data.

Standardizing a dataset involves rescaling the distribution of values so that the mean of observed values is 0 and the standard deviation is 1. It is sometimes referred to as *“*whitening” This can be thought of as subtracting the mean value or centering the data. Standardization assumes that the observations fit a Gaussian distribution (bell curve) with a well behaved mean and standard deviation. The data can still be standardized if this expectation is not met, but the results obtained may be not reliable. Standartisation is a often called as *Z-Score Normalization*. The Z*-score or standard score* indicates how many standard deviations the data element is from the mean. Since mean and standard deviation are sensitive to outliers, this standardization is sensitive to outliers. Standartisation typically means rescaling the data to have a mean of 0 and a standard deviation of 1 (unit variance) i. e. changing range of the data.