**Feature Scaling:**

It is a technique to standardize the independent features present in the data set in a fixed range.

**Note**: It is the last operation we perform ML model training.

**Types of Feature Scaling:**

1. **Standardization (Z-Score Normalization)**
2. **Normalization**

**Use Cases for types of Feature Scaling:**

|  |  |
| --- | --- |
| **Most Cases are Solved by** | **Standardization** |
| **If Minimum and Maximus is known in dataset** | **Normalization (Min Max Scalar)** |
| **If Outliers present in the dataset** | **Normalization (Robust Scaling)** |
| **If data set is highly sparse** | **Normalization (Max Absolute Scaling)** |

1. **Standardization (Z-SCORE):**

It data transformation techniques performing statistical operation in data where transformed valued outcome (***Mean = 0***, and ***Standard deviation = 1)***

**Mathematical Intuition (Formula):**

[Standardized Values: Example](https://www.statisticshowto.com/wp-content/uploads/2009/09/aformula.bmp)

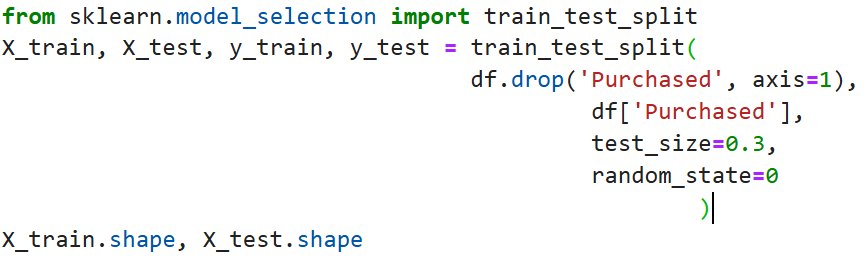
You calculate a standardized value (a z-score), using the above formula. The symbols are:

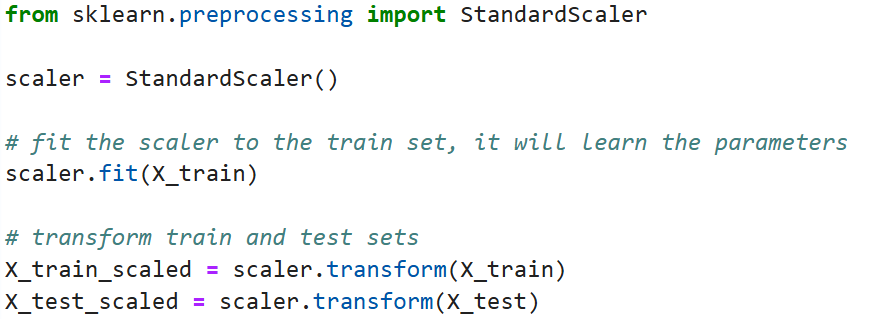
* **X**: the [observation](https://www.statisticshowto.com/observation-in-statistics/)(a specific value that you are calculating the z-score for).
* **Mu (μ):** the [mean](https://www.statisticshowto.com/mean/).
* **Sigma (σ):** the [standard deviation](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/).

1. **Standardization (Z-SCORE):**

**Standardization (Z-SCORE) Scripts Intuition:**

**Code Snippets:**





1. **Standardization (Z-SCORE):**

**When to use Standardization:**

|  |  |
| --- | --- |
| **K-Means** | Use the Euclidean distance measure. |
| **KNN** | Measure the distance between pairs of samples and these distances are influenced by the measurement units. |
| **PCA** | Try to get the feature with maximum variance. |
| **ANN** | Apply Gradient Descent. |
| **Gradient Descent** | Theta calculation becomes faster after feature scaling and learning rate in the update equation of Stochastic Gradient Descent is the same for every parameter. |

1. **Normalization:**

It is a technique to change the values of numerical columns in the dataset to use common scale, without distorting differences in the ranges of values or losing information.

Numeric column = Units + Magnitude

***Note: The idea is to eliminate the units of the numeric columns and bring it in to the common scale.***

**Types:**

1. **Min Max Scaling.**
2. **Mean Normalization.**
3. **Max Absolute Scaling.**
4. **Robust Scaling.**

**2.Normalization:**

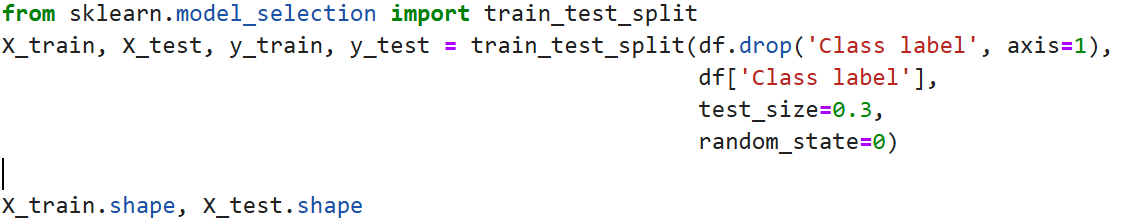
**Min Max Scaling:**

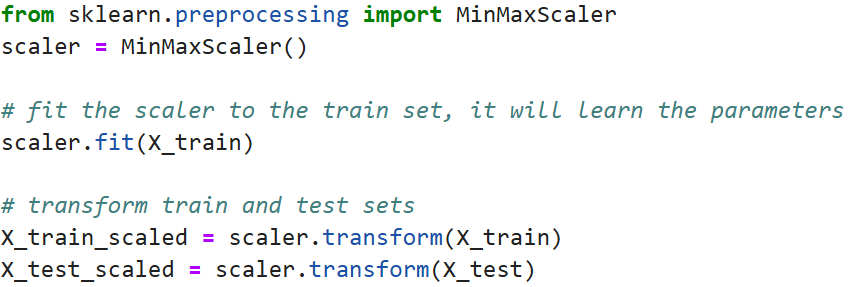
This technique scales the values of a feature to a range between **0** and **1.** This is done by subtracting the minimum value of the feature from each value, and then dividing by the range of the feature. Transformed valued outcome (***MIN = 0***, and ***MAX = 1)***.

**Mathematical Intuition (Formula):**



**Code Snippets:**



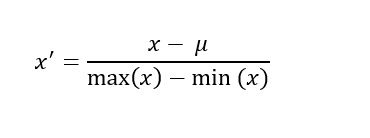


**2.Normalization:**

**Mean Normalization:**

It is a Normalization technique that transforms value of numeric columns in dataset and Centers the data on the mean and scales it to a range of **[-1, 1].**

**Mathematical Intuition (Formula):**



***Note: It is used when we want mean centralized data (Less Common)***

**2.Normalization:**

**Max Absolute Scaling:**

It is a normalization technique that rescales value of numeric columns in dataset between **-1 and 1** by dividing every observation by its maximum absolute value. We can apply the maximum absolute scaling in Pandas using the .max () and .abs () methods.

***Note: It is used when there is sparse data (Excessive Zero) less Common.***

**Robust Scaling:**

It is technique used to standardize input variable in the presence of outliers. It uses Median and interquartile range (IQR) to scale raw input values.

**Mathematical Intuition (Formula):**

IQR = 75th percentile  – 25th percentile

X robust = X – median/ IQR