**Feature Scaling**

**Introduction:**

Feature scaling is a method used to standardize the range of independent variables or features of data. In [data processing](https://en.wikipedia.org/wiki/Data_processing), it is also known as data normalization and is generally performed during the data preprocessing steps.

Feature scaling may be useful under certain circumstances (e.g. when variables span different ranges). There are several different versions of scaling, the most important of which are listed below. Scaling procedures may be applied to the full [data frame](http://www.statistics4u.info/fundstat_eng/cc_dmatrix.html), or to parts of the data only (e.g. column wise).

1. Normalization
2. Standardization
3. Mean Centering

**Why do we go for Feature scaling?**

If training an algorithm using different features and some of them are off the scale in their magnitude, then the results might be dominated by them. Since the range of values of raw data varies widely, in some [machine learning](https://en.wikipedia.org/wiki/Machine_learning) algorithms, objective functions will not work properly without [normalization](https://en.wikipedia.org/wiki/Normalization_%28statistics%29). For example, the majority of [classifiers](https://en.wikipedia.org/wiki/Statistical_classification) like KNN, K-means which uses Euclidean, Manhattan or cosine similarity to calculate the distance between two. If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance.

Another reason why feature scaling is applied is that [gradient descent](https://en.wikipedia.org/wiki/Gradient_descent) converges much faster with feature scaling than without it

**Scaling methods:**

1. Normalization:

Also called **Min-Max scaling.** It is the process of reducing unwanted variation either within or between variables. Normalization brings all of the variables into proportion with one another. It transforms data into a range between 0 and 1.



It is Sensitive to outliers

1. Standardization:

Also called Z-score or is the process where the features are rescaled so that they’ll have the properties of a standard normal distribution with μ=0 and σ=1, where μ is the mean (average) and σ is the standard deviation from the mean. Standard scores (also called z scores) of the samples are calculated as follows:

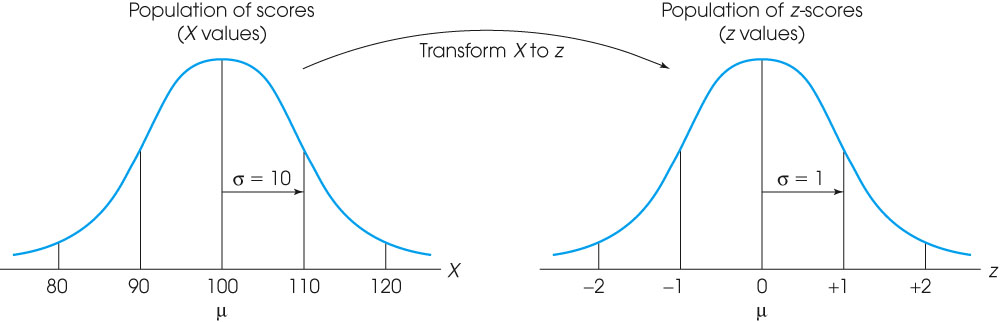


Where

U is the mean of population

2 is the standard deviation of the population

* Z represents the difference between raw score and population mean in the units of standard deviation.
* Z is negative when the raw score is below the mean and Z is positive when above mean
* Standardization works well if the data is uniformly distributed



## Which one should we use?

*“Standardization or Min-Max scaling?”* - There is no obvious answer to this question: it really depends on the application.

For example, in clustering analyses, standardization may be especially crucial in order to compare similarities between features based on certain distance measures. Another prominent example is the Principal Component Analysis, where we usually prefer standardization over Min-Max scaling, since we are interested in the components that maximize the variance (depending on the question and if the PCA computes the components via the correlation matrix instead of the covariance matrix; [but more about PCA in my previous article](http://sebastianraschka.com/Articles/2014_pca_step_by_step.html)).

However, this doesn’t mean that Min-Max scaling is not useful at all! A popular application is image processing, where pixel intensities have to be normalized to fit within a certain range (i.e., 0 to 255 for the RGB color range). Also, typical neural network algorithm requires data that on a 0-1 scale.

**Additional Information:**

Normalization/standardization is designed to achieve a similar goal, which is to create features that have similar ranges to each other. We want that so we can be sure we are capturing the true information in a feature and those we don’t over weigh a particular feature just because its values are much larger than other features. Normalization and standardization is pretty much the same thing and both relate to the issue of feature scaling.

Both of these techniques have their drawbacks. If you have outliers in your data set, normalizing your data will certainly scale the “normal” data to a very small interval. And generally, most of data sets have outliers. When using standardization, your new data aren’t bounded (unlike normalization).

**Interview Questions:**

1. What is the difference between normalization and standardization?
2. Which method should we use for scaling data?
3. Reason behind performing scaling on the data?