

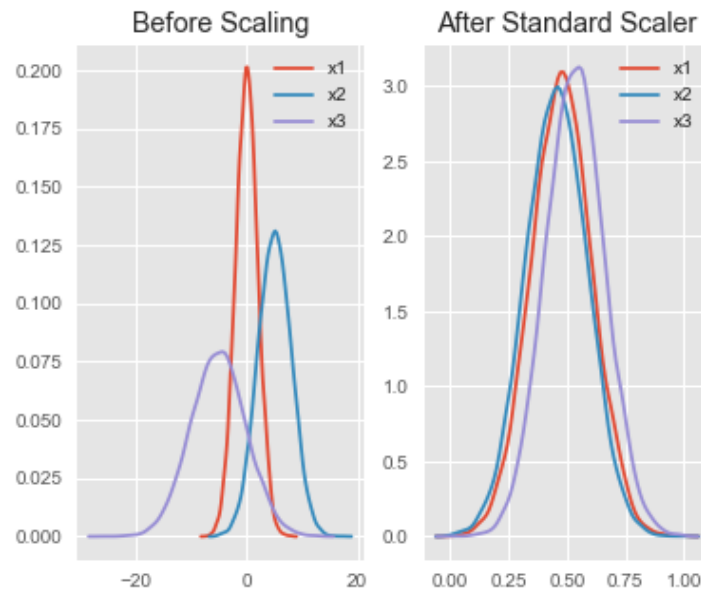
# Standard Scaling and Normalization

## Introduction

**Data preprocessing is a crucial step in machine learning and data analysis.** Two common techniques used to prepare data for modeling are **standard scaling and normalization**. These techniques help in ensuring that features are on the same scale, making algorithms perform better and more accurately. In this article, we will dive into the concepts of standard scaling and normalization, providing simple explanations and real-life examples to illustrate their significance.

## Standard Scaling:

Standard scaling, also known as **Z-score normalization**, involves **transforming features to have a mean of 0 and a standard deviation of 1**. This technique is particularly useful when dealing with features that have different units or ranges. By standard scaling, we center the data around the mean and ensure that it follows a standard normal distribution.



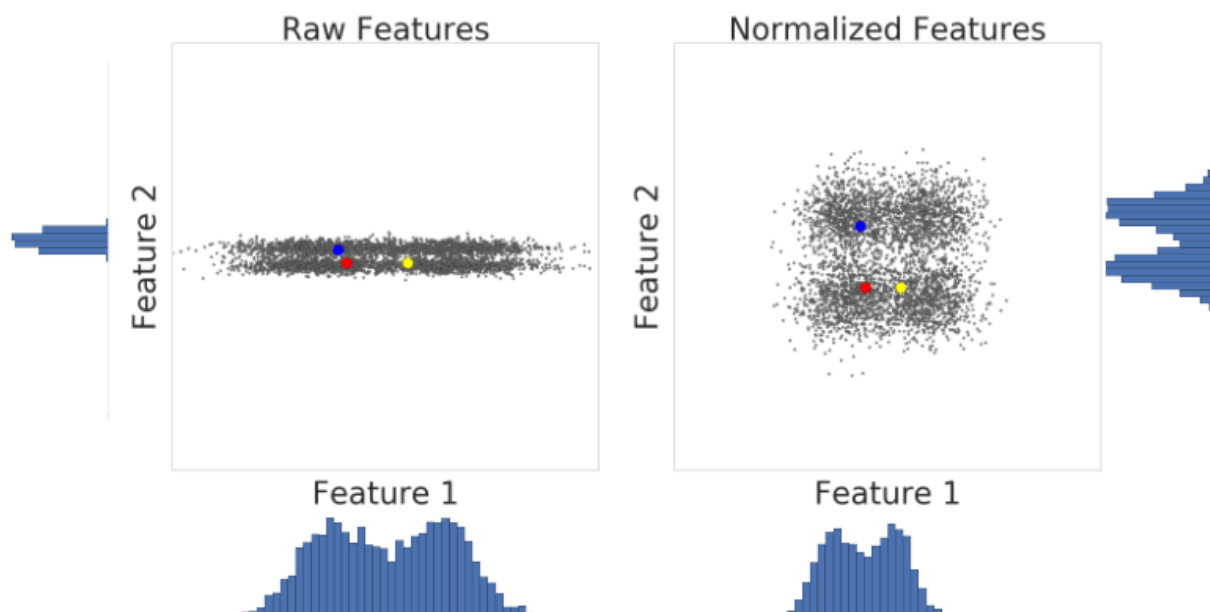
## Example:

Consider a dataset of house prices, where the 'area' feature represents the square footage of houses, and the 'price' feature represents the sale prices. The 'area' feature has values ranging from hundreds to thousands, while the 'price' feature is in the tens of thousands to millions. To standard scale the 'area' and 'price' features:

- Calculate the mean and standard deviation of the 'area' and 'price' features.
- For each value in the 'area' and 'price' columns, subtract the mean and divide by the standard deviation.

## Normalization

**Normalization involves scaling features to a specific range, typically between 0 and 1.** This technique is beneficial when you want to maintain the relative relationships between features and prevent any single feature from dominating the others due to its larger values.



## Example:

Let's consider a dataset of student exam scores, with 'math score' and 'history score' features. The 'math score' ranges from 0 to 100, while the 'history score' ranges from 0 to 150. To normalize these scores:

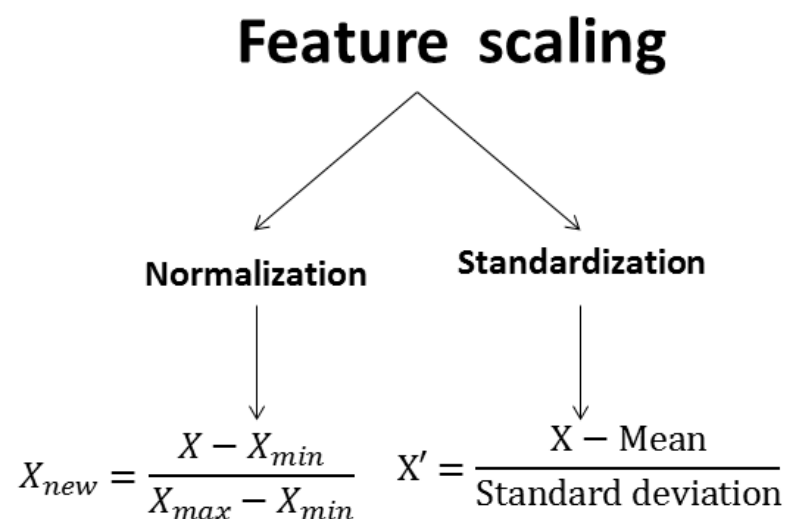
- Find the minimum and maximum values of 'math score' and 'history score'.
- For each value in the 'math score' and 'history score' columns, subtract the minimum value and divide it by the range (maximum - minimum).

## Comparing Standard Scaling and Normalization:

To better understand the differences, let's go back to our house prices and student exam scores examples.

- In the house prices example, standard scaling would bring the 'area' and 'price' features to a similar scale, allowing machine learning algorithms to treat them equally during model training.
- On the other hand, normalization would ensure that 'math score' and 'history score' are confined to the same range, preventing one from overshadowing the other.

## Formula:



## **Conclusion:**

Standard scaling and normalization are essential data preprocessing techniques that contribute to the success of machine learning models and data analysis tasks. By ensuring that features are on the same scale, we eliminate biases introduced by the inherent differences in units and ranges. Whether you choose standard scaling or normalization depends on your specific dataset and goals. Armed with a better understanding of these techniques and real-life examples, you can confidently prepare your data for more accurate and effective analyses.