Feature Scaling

M.SAXENA

Feature scaling, also known as data normalization or standardization, is a preprocessing technique used to bring all features or variables of a dataset to a similar scale or range. It is particularly important when working with machine learning algorithms that rely on distance-based calculations or when comparing variables with different units or scales.

M.SAXENA

Goals of feature scaling

M.SAXENA

Mitigating the impact of different scales

Variables in a dataset often have different units and ranges. Some variables might have larger values or wider ranges compared to others. When working with distance-based algorithms (e.g., k-nearest neighbors, clustering), these differences in scales can lead to biased results, as the algorithm may give more importance to variables with larger values or ranges.

M.SAXENA

Promoting faster convergence

Many optimization algorithms, such as gradient descent, converge faster when features are on a similar scale. Feature scaling can help in achieving faster convergence and reducing the computational burden during model training.

M.SAXENA

Techniques for feature scaling

M.SAXENA

Standardization

It transforms the data such that it has zero mean and unit variance. Each value is subtracted by the mean of the feature and divided by its standard deviation. This scales the data around the mean and is represented as z-scores. Standardization retains the shape of the distribution but centers it at zero.

M.SAXENA

Min-Max scaling

It transforms the data to a fixed range, typically between 0 and 1. Each value is subtracted by the minimum value of the feature and divided by the range (maximum value minus minimum value). Min-Max scaling preserves the original distribution and maps the data to a specific range.

M.SAXENA

Robust scaling

It is similar to standardization, but it uses robust statistics to handle outliers. Instead of using the mean and standard deviation, robust scaling subtracts the median and scales the data using the interquartile range.

M.SAXENA

By applying feature scaling, variables are put on a similar scale, reducing bias, improving the performance of certain algorithms, and enabling a fair comparison and interpretation of their effects on the model's output.

M.SAXENA

Follow me

M.Saxena