STANDARDIZATION AND NORMALIZATION AND SCALING

Scaling - Feature Scaling is a technique to standardize the independent feature present in the data in a fixed range. It is mainly used to handle highly varying magnitudes or values or units.

- If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as lower, regardless of the unit of the value.

Example - If an algorithm is not using feature scaling method then it can consider the value of 3000 meter to be greater than 5 km but that's actually not true & and in this case, the algorithm will give wrong prediction. So we use Feature Scaling to bring all values to same magnitude and thus tackle the issue.

Normalization - Normalization is a scaling technique in which value are shifted and rescale, so that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

Formula, $$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

X_{max} → Maximum value of feature

X_{min} → Minimum value of feature

- When X is minimum, numerator will be 0 and hence X' is 0.
- When X is maximum, numerator is equal to dunominator ($X = X_{max}$) hence we X' is 1.

Standardization - standardization is another scaling technique when the value are centered around mean with a unit Standard deviation. This means that mean of the attributes becomes zero and the resultant distribution has a unit standard deviation.

Formula, $$X' = \frac{X - \mu}{\sigma}$$

- μ is the mean of the feature value.
- σ is the standard deviation of feature value.
- The value is not restricted to any range.
- Not affected by outlier as it does not have a strict range.

→ Normalization is good to use when distribution of data does not follow normal distribution. This can be useful in algorithm that do not assume any distribution like K Nearest neighbour and Neural networks.

→ Standerdization can be useful when data follows a normal distribution.

→ Algorithm DO NOT REQUIRE SCALING - CART, Random forest, Gradient boosting decision tree. They (Algorithm) always relies on some type of rules.

→ K-nearest neighbour with an Euclidean distance measure is sensitive to magnitude & hence should be scaled for all features to weigh in equally.

→ Scaling is critical, while performing PCA (Principal Component Analysis) - PCA tries to get feature with maximum variance & variance is high for high magnitude.