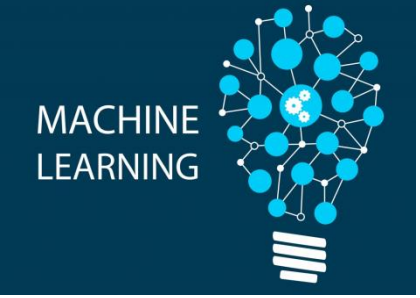
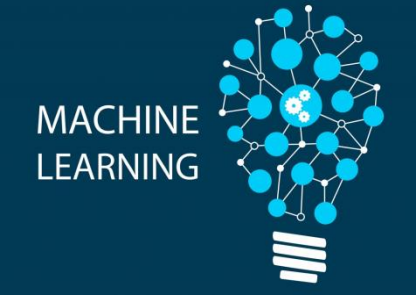


Feature Scaling

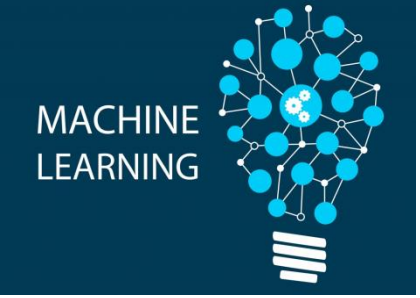
Diptangshu Banik

Twitter - @dipbanik



Overview

- Lot of ML models are based on Euclidean distance(squared roots of the sum of the squared coordinates for two data points).
- Age and Salary are of very different scale. So they have to be converted into the same scale.
- Standardization = $\frac{x - \text{mean}(x)}{\text{Standard deviation}(x)}$
- Normalization = $\frac{x = \min(x)}{\max(x) - \min(x)}$



Code

```
>>> from sklearn.preprocessing import StandardScalar

>>> sc_X = StandardScaler()

>>> X_train = sc_X.fit_transform(X_train)

>>> X_test = sc_X.transform(X_test)
```