

## Machine Learning Experiment 7

Aim: To perform k-means clustering.

Theory:

It is an unsupervised machine learning algorithm.  
• It is used to group similar data points together into clusters, based on their similarity to each other.

The k-means algorithm works by first selecting a number of clusters, represented by  $k$ , and then initializing the center of those clusters randomly. The algorithm then iteratively assigns each data point to the closest cluster centre, and updates the center of each cluster to be the mean of all points assigned to it. This process continues till the cluster centers no longer move, or a maximum number of iterations is reached.

Commonly used functions for k-means.

i) `sklearn.cluster.KMeans`: It creates a `KMeans` object that is used to perform k-means clustering. The important parameters for this function are 'n\_clusters' which defines the number of clusters to form → in it which specifies the method for initialization of the cluster centres.

ii) `predict(x)`:

Predicts the closest cluster for each

input data + point in 'x'.

ii) fit(x): fit the kMeans model to the input data 'x'.

iv) inertia\_ : attribute retains the sum of squared distances of samples to their closest cluster centers. It can't be used to determine the quality of clustering.

v) transform(x): transforms input data 'x' to the distance space. For each data points in 'x' it returns the distance between they point with each of cluster centers.

vi) fit\_transform(x): fit the kMeans model to the input data 'x' & returns the transformed data.

Conclusion: Hence the kMeans was successfully implemented using python.