Weekly Report 2 - K-Means

Ganji Varshitha

AI20BTECH11009

Introduction

K-Means is an unsupervised learning algorithm which performs partitioned clustering. Clustering helps us to understand the structure of the data by grouping it into distinct sub-groups.

Algorithm

It is a parametric method where we need to specify the number of clusters K, which we want to divide the data into.

It is a simple and iterative algorithm described as follows:

> Initialize K random points from the dataset as centroids
>
> **repeat**
>
> > Form K clusters by assigning all points to the closest centroid
> >
> > Recompute the centroid of each cluster
>
> **until** The centroids don't change

It can be seen as Expectation-Maximisation problem where the E-step is assigning the data points to the closest cluster and the M-step is computing the centroid of each cluster. It minimises the intra-cluster sum of squared distance from its centroid and keeps the clusters distant from each other.