Mycompanion 43304

Assignment - 08

chusterina to orate a

* Ain > Implement is rean algorithm for clustering, to create a cluster on dataset (using Python)

* Objetive >1. To understand the concept of clustering
2: To implement is means clustering algorithm

* Theory >

K reans clustering aims to partition a observations into K clusters in which each observation belongs to the reinflow cluster with the rearest rean serving as prototype of the cluster

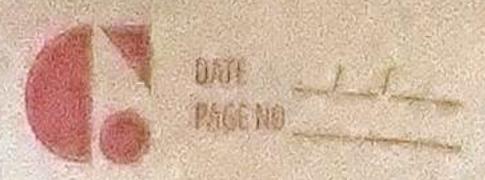
Working:

The K nears clustering algorithm attempts to split a given anonymous dataset, into a fixed number of cluster.

Intially & number of so called centroids are chosen. They are picked up randonly in the initial stage such that all centroids are unique. These centroids are used to train the KNN classifier.

The resulting classifier is used to classify the data and thereby produce an initial randomized set of clusters. Each centroid is the activities rean of the cluster it defines. The process of classification and centroid adjustment is repeated with the values of the centroids stabilizes.

The final centroids will be used to produce the final classification, clustering of the input data, effectively turning the set of initially anonymous data into a set of data, each with a class identity.



= inycompanion 13304 Advantages: 1 If variables are huge, then is near most of the times is computationally faster than hierarchial distant 2. Relatively simple to implement 3. Scales to large dataset 4. Easily adopts to new examples 5. Generalizes to plusters of different shapes and sizes · Disadvantages : 1. Choosing 15 manually 2 centraids can be dragged by outliers 3. Scaleng with number of dirension 4. Dependent on the initial values p 13 Conclusion > Thus in this assignment we bearned & about K years clustering algorithm and implemented it on a mandonly generated dataset.