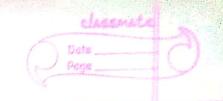
Classwite Date Dage

Assignment No. 2

Groblem Statement: Consider a Suitable dataset, for clustering of data instances in diff. groups , opply diff. dustering techniques (min. 2) · Visualie the dusterns using Suitable tools. * Objective: i) understand various dustoring types & how to implement the same. ii) use Python liberaries & appropriate date Sets to Penform clustering & visualize the same. Outcome: Understood K-means, Hierachia clustering & Centerin it on dataset coas data by brand. * Theony: Occusion on cluster analysis is a task of grouping a set of objects in such a way that objects in the same group (dustern) are more similar to each other than to those in each other. O clustering is the main task of explorationy data mining & a common

Classmate technique for statistical data analysis, and is used in many fields. @ K-mours dustoring is a type of unsupervised learning, which is used she you unlabeled data (i.e. data without defined categories on groups). The goal of this algorithm is to Rind groups in the data, with the rumber of groups represented by variable k. the algorithm works itteratively to assign each date Point to one of K groups based on the feature frovided data Points one clustered based on feature similarity. the result of this algorithm is-> the centraid of k-dustos can be used to label new date => labels from the training data lead data Point is assigned to single cluster @ Rather than defining groups, between locking at the data, clustering allows you to find & analyze the groups that have formed enganically. O Hieranchical clustering is an algorithm that groups similar abjects into groups icalled cluster. the end Point is a set of cluster where each duster is distinct from each other, and the object



wither each cluster are broadly similar

O Given a set of Mitems to be dustoned and an MXM distance methix, the bosic Bross of hierarchical dustoring is:

-> cossign each item to its own cluster,

So now there one M clusters; let the

dist. between the cluster equal the dist.

(Similornities) between the items they contain.

-> find the closest (most similar) Patr of

clusters and merge them into a single elustr

-> compute the distances (similarities) between

the new cluster & each of the old cluster

-> Repead above 2 steps until all clusters

one clustered into a single cluster of

size M.

The default distance measured in the Euclidean dist. which is the Sq. root of the Sum of the differences.

Other agglomenative clustering approach, there one 4 beside methods, anothis method being one of them. It says that the dist between two clusters. A & B, is how much the sum of squares will increase when they one menged

6	classmate Date Page
	O The dataset used was 'mall-custorer' which contains various ferameters.
	these entries were closted using the algorithm.
*	Condusion:
	K-means and hierarchical (agglometral
<u>.</u>	dustering techiques were understood,
1. Owner	clustering techiques were understood, Successfuly implemented, Is the sieg output
<u> </u>	was obtained.
1-4-5 1	English to the terms of the ter
100000	
BITH!	1 mai , it is in the second in the
72	
19.11	
and the second	World with the second of the s
grand and the second se	
	de moderne de la companya de la comp