Machine learning-Assignment6

Student id:700674089

Student Name: Jahnavi Danda

CRN: CS571013466

**GitHub Repository**

https://github.com/jxd40890/Machine-Learning-Assignment6.git

**Assignment6 video URL**

https://drive.google.com/drive/folders/1td\_cRMb92iMcXVBFnURkirX4C8r-M1\_V?usp=sharing

Question1

1) (Provide only mathematical solutions for this question) Six points with the following attributes are given, calculate and find out clustering representations and dendrogram using Single, complete, and average link proximity functions in hierarchical clustering technique.

**Answer**

**Single Link Proximity:**

In **Single Linkage,**the distance between two clusters is the minimum distance between members of the two clusters

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **p1** | | **p2** | | **p3** | | **p4** | | **p5** | **p6** |
| **p1** | | 0 | | 0.2357 | | 0.2218 | | 0.3688 | | 0.3421 | 0.2347 |
| **p2** | | 0.2357 | | 0 | | 0.1483 | | 0.2042 | | 0.1388 | 0.254 |
| **p3** | | 0.2218 | | 0.1483 | | 0 | | 0.1513 | | 0.2843 | 0.11 |
| **p4** | | 0.3688 | | 0.2042 | | 0.1513 | | 0 | | 0.2932 | 0.2216 |
| **p5** | | 0.3421 | | 0.1388 | | 0.2843 | | 0.2932 | | 0 | 0.3921 |
| **p6** | | 0.2347 | | 0.254 | | 0.11 | | 0.2216 | | 0.3921 | 0 |
|  | |  | |  | |  | |  | |  |  |
| smallest distance from above data is | | | | | | | | | | 0.11 |  |
| so p3 and p6 forms first cluster | | | | | | | | | | |
|  | **p1** | | **p2** | | **p36** | | **p4** | | **p5** | |
| **p1** | 0 | | 0.2357 | | 0.2218 | | 0.3688 | | 0.3421 | |
| **p2** | 0.2357 | | 0 | | 0.1483 | | 0.2042 | | 0.1388 | |
| **p36** | 0.2218 | | 0.1483 | | 0 | | 0.1513 | | 0.2843 | |
| **p4** | 0.3688 | | 0.2042 | | 0.1513 | | 0 | | 0.2932 | |
| **p5** | 0.3421 | | 0.1388 | | 0.2843 | | 0.2932 | | 0 | |
|  |  | |  | |  | |  | |  | |
| smallest distance from above data is | | | | | | | | | 0.1388 | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| so p2 and p5 forms 2nd cluster | | | | | |
|  | **p1** | **p25** | **p36** | **p4** |  |
| **p1** | 0 | 0.2357 | 0.2218 | 0.3688 |  |
| **p25** | 0.2357 | 0 | 0.1483 | 0.2042 |  |
| **p36** | 0.2218 | 0.1483 | 0 | 0.1513 |  |
| **p4** | 0.3688 | 0.2042 | 0.1513 | 0 |  |
|  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.1483 |
| so p25 and p36 forms 3rdcluster | | | | | |
|  |  |  |  |  |  |
|  | **p1** | **p(25)(36)** | **p4** |  |  |
| **p1** | 0 | 0.2218 | 0.3688 |  |  |
| **p(25)(36)** | 0.2218 | 0 | 0.1513 |  |  |
| **p4** | 0.3688 | 0.1513 | 0 |  |  |
|  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.153 |
| so p(25)(36)and p4 forms 4thcluster | | | | | |
|  | **p1** | **p4(25)(36)** |  |  |  |
| **p1** | 0 | 0.2218 |  |  |  |
| **p4(25)(36)** | 0.2218 | 0 |  |  |  |

Chart

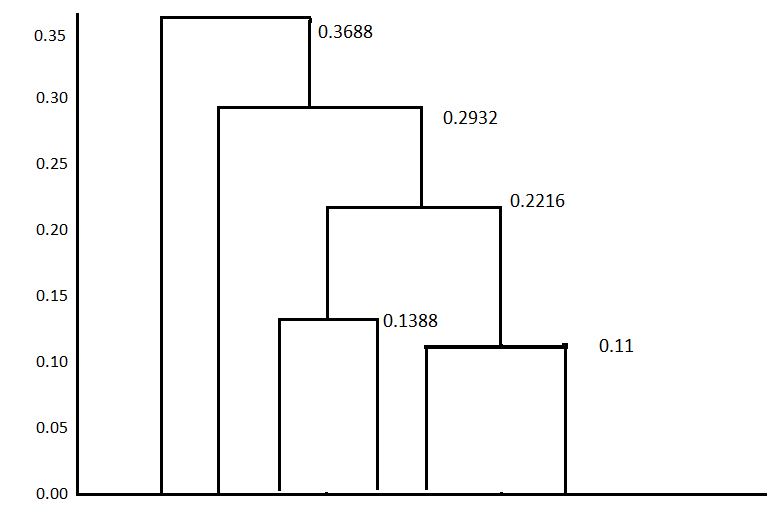
Description automatically generated

**1 4 2 5 3 6**

**Complete Link Proximity:**

In **Complete Linkage**, the distance between two clusters is the maximum distance between members of the two clusters

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **p1** | **p2** | **p3** | **p4** | **p5** | **p6** |
| **p1** | 0 | 0.2357 | 0.2218 | 0.3688 | 0.3421 | 0.2347 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **p2** | 0.2357 | 0 | 0.1483 | 0.2042 | 0.1388 | 0.254 |
| **p3** | 0.2218 | 0.1483 | 0 | 0.1513 | 0.2843 | 0.11 |
| **p4** | 0.3688 | 0.2042 | 0.1513 | 0 | 0.2932 | 0.2216 |
| **p5** | 0.3421 | 0.1388 | 0.2843 | 0.2932 | 0 | 0.3921 |
| **p6** | 0.2347 | 0.254 | 0.11 | 0.2216 | 0.3921 | 0 |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.11 |  |
| so p3 and p6 forms first cluster | | | | | |  |
|  | **p1** | **p2** | **p36** | **p4** | **p5** |  |
| **p1** | 0 | 0.2357 | 0.2347 | 0.3688 | 0.3421 |  |
| **p2** | 0.2357 | 0 | 0.254 | 0.2042 | 0.1388 |  |
| **p36** | 0.2347 | 0.254 | 0 | 0.2216 | 0.3921 |  |
| **p4** | 0.3688 | 0.2042 | 0.2216 | 0 | 0.2932 |  |
| **p5** | 0.3421 | 0.1388 | 0.3921 | 0.2932 | 0 |  |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.1388 |  |
| so p2 and p5 forms 2nd cluster | | | | | |  |
|  | **p1** | **p25** | **p36** | **p4** |  |  |
| **p1** | 0 | 0.3421 | 0.2347 | 0.3688 |  |  |
| **p25** | 0.3421 | 0 | 0.3921 | 0.2932 |  |  |
| **p36** | 0.2347 | 0.3921 | 0 | 0.2216 |  |  |
| **p4** | 0.3688 | 0.2932 | 0.2216 | 0 |  |  |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.2216 |  |
| so p25 and p36 forms 3rdcluster | | | | | |  |
|  | **p1** | **p(25)(36)** | **p4** |  |  |  |
| **p1** | 0 | 0.3421 | 0.3688 |  |  |  |
| **p(25)(36)** | 0.3421 | 0 | 0.2932 |  |  |  |
| **p4** | 0.3688 | 0.2932 | 0 |  |  |  |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.2932 |  |
| so p(25)(36)and p1 forms 4thcluster | | | | | |  |
|  | **p1(25)(36)** | **p4** |  |  |  |  |
| **p1(25)(36)** | 0 | 0.1483 |  |  |  |  |
| **p4** | 0.3688 | 0 |  |  |  |  |



**4 1 2 5 3 6**

**Average Link Proximity:**

In **Average Linkage,**the distance between two clusters is the average of all distances between members of the two clusters

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **p1** | **p2** | **p3** | **p4** | **p5** | **p6** |
| **p1** | 0 | 0.2357 | 0.2218 | 0.3688 | 0.3421 | 0.2347 |
| **p2** | 0.2357 | 0 | 0.1483 | 0.2042 | 0.1388 | 0.254 |
| **p3** | 0.2218 | 0.1483 | 0 | 0.1513 | 0.2843 | 0.11 |
| **p4** | 0.3688 | 0.2042 | 0.1513 | 0 | 0.2932 | 0.2216 |
| **p5** | 0.3421 | 0.1388 | 0.2843 | 0.2932 | 0 | 0.3921 |
| **p6** | 0.2347 | 0.254 | 0.11 | 0.2216 | 0.3921 | 0 |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.11 |  |
| so p3 and p6 forms first cluster | | | | | |  |
|  | **p1** | **p2** | **p36** | **p4** | **p5** |  |
| **p1** | 0 | 0.2357 | 0.22825 | 0.3688 | 0.3421 |  |
| **p2** | 0.2357 | 0 | 0.20115 | 0.2042 | 0.1388 |  |
| **p36** | 0.22825 | 0.20115 | 0 | 0.18645 | 0.3382 |  |
| **p4** | 0.3688 | 0.2042 | 0.18645 | 0 | 0.2932 |  |
| **p5** | 0.3421 | 0.1388 | 0.3382 | 0.2932 | 0 |  |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.1388 |  |
| so p2 and p5 forms 2nd cluster | | | | | |  |
|  | **p1** | **p25** | **p36** | **p4** |  |  |
| **p1** | 0 | 0.2889 | 0.2347 | 0.3688 |  |  |
| **p25** | 0.2889 | 0 | 0.269675 | 0.2487 |  |  |
| **p36** | 0.2347 | 0.269675 | 0 | 0.18645 |  |  |
| **p4** | 0.3688 | 0.2487 | 0.18645 | 0 |  |  |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.18645 |  |
| so p25 and p36 forms 3rdcluster | | | | | |  |
|  | **p1** | **p(25)(36)** | **p4** |  |  |  |
| **p1** | 0 | 0.2618 | 0.3688 |  |  |  |
| **p(25)(36)** | 0.2618 | 0 | 0.217575 |  |  |  |
| **p4** | 0.3688 | 0.217575 | 0 |  |  |  |
|  |  |  |  |  |  |  |
| smallest distance from above data is | | | | | 0.217575 |  |
| so p(25)(36)and p1 forms 4thcluster | | | | | |  |
|  | **p1(25)(36)** | **p4** |  |  |  |  |
| **p1(25)(36)** | 0 | 0.3153 |  |  |  |  |
| **p4** | 0.3153 | 0 |  |  |  |  |

Chart, box and whisker chart

Description automatically generated

**4 1 2 5 3 6**

Question2: -

2) Use CC\_GENERAL.csv given in the folder and apply:

a) Preprocess the data by removing the categorical column and filling the missing values.

b) Apply StandardScaler() and normalize() functions to scale and normalize raw input data.

c) Use PCA with K=2 to reduce the input dimensions to two features.

d) Apply Agglomerative Clustering with k=2,3,4 and 5 on reduced features and visualize result for each k value using scatter plot.

e) Evaluate different variations using Silhouette Scores and Visualize results with a bar chart.

**Answer**

