**K-Means Clustering**

Divide the following data into 4 clusters.

|  |  |
| --- | --- |
| **X** | **Y** |
| 0.1605 | 0.9284 |
| 0.1873 | 0.9635 |
| 0.1930 | 0.8984 |
| 0.2031 | 0.1558 |
| 0.2150 | 0.9390 |
| 0.3035 | 0.3009 |
| 0.3084 | 0.1278 |
| 0.3585 | 0.7390 |
| 0.4458 | 0.6998 |
| 0.5726 | 0.8922 |
| 0.5800 | 0.2989 |
| 0.6188 | 0.5232 |
| 0.7123 | 0.6643 |
| 0.7818 | 0.6516 |
| 0.7846 | 0.9092 |
| 0.7894 | 0.3764 |
| 0.9021 | 0.1139 |
| 0.9549 | 0.9701 |
| 0.9950 | 0.9231 |

Step 1: Plot in Graph and choose 4 points as centroid (Marked in Red)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | X | Y |
|  | Centroid 1 | 0.3035 | 0.3009 |
| Centroid 2 | 0.9549 | 0.9701 |
| Centroid 3 | 0.2150 | 0.9390 |
| Centroid 4 | 0.7123 | 0.6643 |

Step 2: Measure distance of each point from the 4 centroids and assign each point to a cluster that has the minimum distance.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **X** | **Y** | **D1** | **D2** | **D3** | **D4** | **Cluster** |
| 0.1605 | 0.9284 | 0.6436 | 0.7955 | 0.0556 | 0.6118 | C3 |
| 0.1873 | 0.9635 | 0.6727 | 0.7676 | 0.0369 | 0.6042 | C3 |
| 0.1930 | 0.8984 | 0.6077 | 0.7652 | 0.0461 | 0.5696 | C3 |
| 0.2031 | 0.1558 | 0.1764 | 1.1083 | 0.7833 | 0.7196 | C1 |
| 0.2150 | 0.9390 | 0.6442 | 0.7405 | 0.0000 | 0.5681 | C3 |
| 0.3035 | 0.3009 | 0.0000 | 0.9339 | 0.6442 | 0.5469 | C1 |
| 0.3084 | 0.1278 | 0.1732 | 1.0619 | 0.8166 | 0.6716 | C1 |
| 0.3585 | 0.7390 | 0.4415 | 0.6396 | 0.2462 | 0.3616 | C3 |
| 0.4458 | 0.6998 | 0.4236 | 0.5764 | 0.3323 | 0.2689 | C4 |
| 0.5726 | 0.8922 | 0.6497 | 0.3902 | 0.3606 | 0.2674 | C4 |
| 0.5800 | 0.2989 | 0.2765 | 0.7688 | 0.7369 | 0.3886 | C1 |
| 0.6188 | 0.5232 | 0.3858 | 0.5592 | 0.5796 | 0.1692 | C4 |
| 0.7123 | 0.6643 | 0.5470 | 0.3903 | 0.5682 | 0.0000 | C4 |
| 0.7818 | 0.6516 | 0.5931 | 0.3625 | 0.6355 | 0.0707 | C4 |
| 0.7846 | 0.9092 | 0.7756 | 0.1809 | 0.5704 | 0.2554 | C2 |
| 0.7894 | 0.3764 | 0.4917 | 0.6164 | 0.8040 | 0.2981 | C4 |
| 0.9021 | 0.1139 | 0.6271 | 0.8578 | 1.0737 | 0.5822 | C4 |
| 0.9549 | 0.9701 | 0.9338 | 0.0000 | 0.7405 | 0.3903 | C2 |
| 0.9950 | 0.9231 | 0.9302 | 0.0618 | 0.7802 | 0.3833 | C2 |

Step 3: Re-calculate the centroids by finding the average of the points assigned to a cluster.

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| Centroid 1 | 0.3487 | 0.2208 |
| Centroid 2 | 0.9115 | 0.9341 |
| Centroid 3 | 0.2229 | 0.8937 |
| Centroid 4 | 0.6890 | 0.5602 |

Step 4: Repeat Step 2 and 3 until centroid points become constant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **X** | **Y** | **D1** | **D2** | **D3** | **D4** | **Cluster** |
| 0.1605 | 0.9284 | 0.7322 | 0.7511 | 0.0714 | 0.6441 | C3 |
| 0.1873 | 0.9635 | 0.7600 | 0.7248 | 0.0783 | 0.6437 | C3 |
| 0.1930 | 0.8984 | 0.6953 | 0.7194 | 0.0302 | 0.6003 | C3 |
| 0.2031 | 0.1558 | 0.1594 | 1.0524 | 0.7382 | 0.6322 | C1 |
| 0.2150 | 0.9390 | 0.7305 | 0.6965 | 0.0460 | 0.6067 | C3 |
| 0.3035 | 0.3009 | 0.0920 | 0.8778 | 0.5982 | 0.4646 | C1 |
| 0.3084 | 0.1278 | 0.1014 | 1.0070 | 0.7707 | 0.5761 | C1 |
| 0.3585 | 0.7390 | 0.5183 | 0.5864 | 0.2057 | 0.3758 | C3 |
| 0.4458 | 0.6998 | 0.4888 | 0.5213 | 0.2954 | 0.2805 | C4 |
| 0.5726 | 0.8922 | 0.7078 | 0.3415 | 0.3497 | 0.3519 | C2 |
| 0.5800 | 0.2989 | 0.2441 | 0.7165 | 0.6938 | 0.2831 | C1 |
| 0.6188 | 0.5232 | 0.4055 | 0.5045 | 0.5422 | 0.0794 | C4 |
| 0.7123 | 0.6643 | 0.5735 | 0.3353 | 0.5405 | 0.1067 | C4 |
| 0.7818 | 0.6516 | 0.6109 | 0.3109 | 0.6091 | 0.1302 | C4 |
| 0.7846 | 0.9092 | 0.8148 | 0.1293 | 0.5619 | 0.3619 | C2 |
| 0.7894 | 0.3764 | 0.4673 | 0.5709 | 0.7672 | 0.2095 | C4 |
| 0.9021 | 0.1139 | 0.5636 | 0.8202 | 1.0341 | 0.4945 | C4 |
| 0.9549 | 0.9701 | 0.9638 | 0.0563 | 0.7359 | 0.4886 | C2 |
| 0.9950 | 0.9231 | 0.9544 | 0.0842 | 0.7727 | 0.4747 | C2 |

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| Centroid 1 | 0.3487 | 0.2208 |
| Centroid 2 | 0.8268 | 0.9237 |
| Centroid 3 | 0.2229 | 0.8934 |
| Centroid 4 | 0.7084 | 0.5049 |

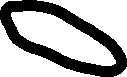
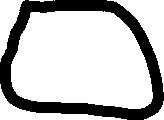
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **X** | **Y** | **D1** | **D2** | **D3** | **D4** | **Cluster** |
| 0.1605 | 0.9284 | 0.7322 | 0.6664 | 0.0714 | 0.6925 | C3 |
| 0.1873 | 0.9635 | 0.7600 | 0.6407 | 0.0783 | 0.6941 | C3 |
| 0.1930 | 0.8984 | 0.6953 | 0.6343 | 0.0302 | 0.6484 | C3 |
| 0.2031 | 0.1558 | 0.1594 | 0.9893 | 0.7382 | 0.6141 | C1 |
| 0.2150 | 0.9390 | 0.7305 | 0.6119 | 0.0460 | 0.6571 | C3 |
| 0.3035 | 0.3009 | 0.0920 | 0.8134 | 0.5982 | 0.4534 | C1 |
| 0.3084 | 0.1278 | 0.1014 | 0.9499 | 0.7707 | 0.5498 | C1 |
| 0.3585 | 0.7390 | 0.5183 | 0.5034 | 0.2057 | 0.4210 | C3 |
| 0.4458 | 0.6998 | 0.4888 | 0.4419 | 0.2954 | 0.3271 | C3 |
| 0.5726 | 0.8922 | 0.7078 | 0.2561 | 0.3497 | 0.4105 | C2 |
| 0.5800 | 0.2989 | 0.2441 | 0.6718 | 0.6938 | 0.2428 | C4 |
| 0.6188 | 0.5232 | 0.4055 | 0.4513 | 0.5422 | 0.0915 | C4 |
| 0.7123 | 0.6643 | 0.5735 | 0.2835 | 0.5405 | 0.1595 | C4 |
| 0.7818 | 0.6516 | 0.6109 | 0.2758 | 0.6091 | 0.1640 | C4 |
| 0.7846 | 0.9092 | 0.8148 | 0.0446 | 0.5619 | 0.4114 | C2 |
| 0.7894 | 0.3764 | 0.4673 | 0.5486 | 0.7672 | 0.1519 | C4 |
| 0.9021 | 0.1139 | 0.5636 | 0.8133 | 1.0341 | 0.4363 | C4 |
| 0.9549 | 0.9701 | 0.9638 | 0.1362 | 0.7359 | 0.5264 | C2 |
| 0.9950 | 0.9231 | 0.9544 | 0.1682 | 0.7727 | 0.5070 | C2 |

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| Centroid 1 | 0.2717 | 0.1948 |
| Centroid 2 | 0.8268 | 0.9237 |
| Centroid 3 | 0.2600 | 0.8614 |
| Centroid 4 | 0.7307 | 0.4381 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **X** | **Y** | **D1** | **D2** | **D3** | **D4** | **Cluster** |
| 0.1605 | 0.9284 | 0.7420 | 0.6664 | 0.1200 | 0.7520 | C3 |
| 0.1873 | 0.9635 | 0.7733 | 0.6407 | 0.1253 | 0.7558 | C3 |
| 0.1930 | 0.8984 | 0.7080 | 0.6343 | 0.0765 | 0.7078 | C3 |
| 0.2031 | 0.1558 | 0.0789 | 0.9893 | 0.7079 | 0.5984 | C1 |
| 0.2150 | 0.9390 | 0.7464 | 0.6119 | 0.0897 | 0.7189 | C3 |
| 0.3035 | 0.3009 | 0.1108 | 0.8134 | 0.5622 | 0.4487 | C1 |
| 0.3084 | 0.1278 | 0.0764 | 0.9499 | 0.7352 | 0.5241 | C1 |
| 0.3585 | 0.7390 | 0.5510 | 0.5034 | 0.1571 | 0.4786 | C3 |
| 0.4458 | 0.6998 | 0.5342 | 0.4419 | 0.2462 | 0.3869 | C3 |
| 0.5726 | 0.8922 | 0.7596 | 0.2561 | 0.3141 | 0.4809 | C2 |
| 0.5800 | 0.2989 | 0.3254 | 0.6718 | 0.6472 | 0.2052 | C4 |
| 0.6188 | 0.5232 | 0.4778 | 0.4513 | 0.4930 | 0.1406 | C4 |
| 0.7123 | 0.6643 | 0.6439 | 0.2835 | 0.4934 | 0.2270 | C4 |
| 0.7818 | 0.6516 | 0.6847 | 0.2758 | 0.5624 | 0.2195 | C4 |
| 0.7846 | 0.9092 | 0.8795 | 0.0446 | 0.5268 | 0.4742 | C2 |
| 0.7894 | 0.3764 | 0.5486 | 0.5486 | 0.7180 | 0.0852 | C4 |
| 0.9021 | 0.1139 | 0.6356 | 0.8133 | 0.9854 | 0.3667 | C4 |
| 0.9549 | 0.9701 | 1.0333 | 0.1362 | 0.7033 | 0.5773 | C2 |
| 0.9950 | 0.9231 | 1.0265 | 0.1682 | 0.7376 | 0.5524 | C2 |

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| Centroid 1 | 0.2717 | 0.1948 |
| Centroid 2 | 0.8268 | 0.9237 |
| Centroid 3 | 0.2600 | 0.8614 |
| Centroid 4 | 0.7307 | 0.4381 |

Step 5: Plot the points and mark the clusters.



**Hierarchical Clustering**

Draw a dendrogram using hierarchical clustering for the following data:

|  |  |  |
| --- | --- | --- |
| **Point** | **X** | **Y** |
| A | 0.5726 | 0.8922 |
| B | 0.1873 | 0.9635 |
| C | 0.3035 | 0.3009 |
| D | 0.9549 | 0.9701 |
| E | 0.3084 | 0.1278 |
| F | 0.6188 | 0.5232 |
| G | 0.5800 | 0.2989 |
| H | 0.3585 | 0.7390 |
| I | 0.1605 | 0.9284 |
| J | 0.1930 | 0.8984 |

Step 1: Measure distance from one point to another and find the points with minimum distance.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G | H | I | J |
| A | 0.000 | 0.392 | 0.650 | 0.390 | 0.809 | 0.372 | 0.593 | 0.263 | 0.414 | 0.380 |
| B | 0.392 | 0.000 | 0.673 | 0.768 | 0.844 | 0.551 | 0.772 | 0.282 | 0.044 | 0.065 |
| C | 0.650 | 0.673 | 0.000 | 0.934 | 0.173 | 0.386 | 0.276 | 0.442 | 0.644 | 0.608 |
| D | 0.390 | 0.768 | 0.934 | 0.000 | 1.062 | 0.559 | 0.769 | 0.639 | 0.795 | 0.765 |
| E | 0.809 | 0.844 | 0.173 | 1.062 | 0.000 | 0.503 | 0.321 | 0.613 | 0.814 | 0.779 |
| F | 0.372 | 0.551 | 0.386 | 0.559 | 0.503 | 0.000 | 0.228 | 0.338 | 0.612 | 0.568 |
| G | 0.593 | 0.772 | 0.276 | 0.769 | 0.321 | 0.228 | 0.000 | 0.493 | 0.756 | 0.714 |
| H | 0.263 | 0.282 | 0.442 | 0.639 | 0.613 | 0.338 | 0.493 | 0.000 | 0.274 | 0.230 |
| I | 0.414 | 0.044 | 0.644 | 0.795 | 0.814 | 0.612 | 0.756 | 0.274 | 0.000 | 0.044 |
| J | 0.380 | 0.065 | 0.608 | 0.765 | 0.779 | 0.568 | 0.714 | 0.230 | 0.044 | 0.000 |

Step 2: Merge the points with minimum distance into one cluster.

The distance between B, I and I, J is the minimum (0.044). So, they will be merged to create one cluster.

Step 3: Repeat Step 1 and Step 2 for N-1 times or K times where, N = Number of Points and K = pre-defined number of clusters.

[Note: The points merged in the previous step will be considered as a single point and distance with other points will be the minimum distance among those points]

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | C | D | E | F | G | H | BIJ |
| A | 0.000 | 0.650 | 0.390 | 0.809 | 0.372 | 0.593 | 0.263 | 0.380 |
| C | 0.650 | 0.000 | 0.934 | 0.173 | 0.386 | 0.276 | 0.442 | 0.608 |
| D | 0.390 | 0.934 | 0.000 | 1.062 | 0.559 | 0.769 | 0.639 | 0.765 |
| E | 0.809 | 0.173 | 1.062 | 0.000 | 0.503 | 0.321 | 0.613 | 0.779 |
| F | 0.372 | 0.386 | 0.559 | 0.503 | 0.000 | 0.228 | 0.338 | 0.551 |
| G | 0.593 | 0.276 | 0.769 | 0.321 | 0.228 | 0.000 | 0.493 | 0.714 |
| H | 0.263 | 0.442 | 0.639 | 0.613 | 0.338 | 0.493 | 0.000 | 0.230 |
| BIJ | 0.380 | 0.608 | 0.765 | 0.779 | 0.551 | 0.714 | 0.230 | 0.000 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | CE | D | F | G | H | BIJ |
| A | 0.000 | 0.650 | 0.390 | 0.372 | 0.593 | 0.263 | 0.380 |
| CE | 0.650 | 0.000 | 0.934 | 0.386 | 0.276 | 0.442 | 0.608 |
| D | 0.390 | 0.934 | 0.000 | 0.559 | 0.769 | 0.639 | 0.765 |
| F | 0.372 | 0.386 | 0.559 | 0.000 | 0.228 | 0.338 | 0.551 |
| G | 0.593 | 0.276 | 0.769 | 0.228 | 0.000 | 0.493 | 0.714 |
| H | 0.263 | 0.442 | 0.639 | 0.338 | 0.493 | 0.000 | 0.230 |
| BIJ | 0.380 | 0.608 | 0.765 | 0.551 | 0.714 | 0.230 | 0.000 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A | CE | D | FG | H | BIJ |
| A | 0.000 | 0.650 | 0.390 | 0.372 | 0.263 | 0.380 |
| CE | 0.650 | 0.000 | 0.934 | 0.276 | 0.442 | 0.608 |
| D | 0.390 | 0.934 | 0.000 | 0.559 | 0.639 | 0.765 |
| FG | 0.372 | 0.276 | 0.559 | 0.000 | 0.338 | 0.551 |
| H | 0.263 | 0.442 | 0.639 | 0.338 | 0.000 | 0.230 |
| BIJ | 0.380 | 0.608 | 0.765 | 0.551 | 0.230 | 0.000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | CE | D | FG | HBIJ |
| A | 0.000 | 0.650 | 0.390 | 0.372 | 0.263 |
| CE | 0.650 | 0.000 | 0.934 | 0.276 | 0.442 |
| D | 0.390 | 0.934 | 0.000 | 0.559 | 0.639 |
| FG | 0.372 | 0.276 | 0.559 | 0.000 | 0.338 |
| HBIJ | 0.263 | 0.442 | 0.639 | 0.338 | 0.000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | AHBIJ | CE | D | FG |
| AHBIJ | 0.000 | 0.442 | 0.390 | 0.338 |
| CE | 0.442 | 0.000 | 0.934 | 0.276 |
| D | 0.390 | 0.934 | 0.000 | 0.559 |
| FG | 0.338 | 0.276 | 0.559 | 0.000 |

A diagram of a clustering structure

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
|  | AHBIJ | CEFG | D |
| AHBIJ | 0.000 | 0.338 | 0.390 |
| CEFG | 0.338 | 0.000 | 0.559 |
| D | 0.390 | 0.559 | 0.000 |

|  |  |  |
| --- | --- | --- |
|  | AHBIJCEFG | D |
| AHBIJCEFG | 0.000 | 0.390 |
| D | 0.390 | 0.000 |

Step 4: Draw the dendrogram.