|  |  |  |
| --- | --- | --- |
| **Individual** | **Variable 1** | **Variable 2** |
| 1 | 1.0 | 1.0 |
| 2 | 1.5 | 2.0 |
| 3 | 3.0 | 4.0 |
| 4 | 5.0 | 7.0 |
| 5 | 3.5 | 5.0 |
| 6 | 4.5 | 5.0 |
| 7 | 3.5 | 4.5 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Data point** | **Centroid 1** | **Centroid 2** | **Belongs to which group** |
| **(1,1)** | **1.57** | **7.21** | **Group-1** |
| **(1.5,2)** | **0.47** | **6.1** | **Group-1** |
| **(3,4)** | **2.04** | **1.78** | **Group-2** |
| **(5,7)** | **5.64** | **1.84** | **Group-2** |
| **(3.5,5)** | **3.15** | **0.73** | **Group-2** |
| **(4,5,5)** | **3.78** | **0.54** | **Group-2** |
| **(3.5,4.5)** | **2.74** | **1.08** | **Group-2** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Data point** | **Centroid 1** | **Centroid 2:** | **Belongs to which group** |
| **(1,1)** | **0.56** | **7.21** | **Group-1** |
| **(1.5,2)** | **0.56** | **6.1** | **Group-1** |
| **(3,4)** | **3.05** | **1.42** | **Group-2** |
| **(5,7)** | **6.66** | **2.20** | **Group-2** |
| **(3.5,5)** | **4.16** | **0.42** | **Group-2** |
| **(4,5,5)** | **4.78** | **0.61** | **Group-2** |
| **(3.5,4.5)** | **3.75** | **0.72** | **Group-2** |

,(3.9,5.1)

Final two clusters

|  |  |  |
| --- | --- | --- |
| 1 | 1 | 2 |
| 2 | 2 | 3 |
| 3 | 3 | 4 |
| 4 | 4 | 5 |
| 5 | 5 | 6 |
| 6 | 6 | 7 |
| 7 | 7 | 8 |
| 8 | 8 | 9 |
| 9 | 9 | 10 |
| 10 | 10 | 11 |
| 11 | 11 | 12 |
| 12 | 12 | 13 |

how to select (use) clustering model K -Mean in Realtime project ?

1. We did not completed K-means
2. I given one assignment=== How to choose k
3. Python code

which scenario we will use the model

when to chose the model ?

Logistic DT NB KNN =====

Classification:7algorithms

When we will user cluster algorithms

When we will use K-means

How to select K in a particular problem

K=2 k=3 k=4 (Assignment)

When we use cluster algorithms

We don’t have any target column

Finance

Retail

Marketing

Health

Autom

group similar regions together, document clustering based on topics