**K-means Clustering for 1 Dimensional Data**

Dataset {2,4,10,12,3,20,30,11,25}

K=2

Randomly picking the centroid value of the cluster from given dataset.

e.g., C1=2, C2=4

Calculating the distance of values from C1 and C2

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **C1 (2)** | **C2 (4)** | **Assigned Cluster base on nearest value** |
| 2 | **0** | 2 | **C1** |
| 4 | 2 | **0** | **C2** |
| 10 | 8 | **6** | **C2** |
| 12 | 10 | **8** | **C2** |
| 3 | **1** | 1 | **C1** |
| 20 | 18 | **16** | **C2** |
| 30 | 28 | **26** | **C2** |
| 11 | 9 | **7** | **C2** |
| 25 | 23 | **21** | **C2** |

Updated centroid by taking mean of the dataset from the randomly picked values.

C1 = (2+3)/2 = 2.5 C2 = (4+10+12+20+30+11+25)/7 = 16

|  |  |  |
| --- | --- | --- |
| **Centroid** | **C1** | **C2** |
| Old | 2 | 4 |
| New | 2.5 | 16 |

Calculate the mean of new assigned centroid values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **C1 (2.5)** | **C2 (16)** | **Assigned Cluster base on nearest value** |
| 2 | **0.5** | 14 | **C1** |
| 4 | **1.5** | 12 | **C1** |
| 10 | 7.5 | **6** | **C2** |
| 12 | 9.5 | **4** | **C2** |
| 3 | **0.5** | 13 | **C1** |
| 20 | 17.5 | **4** | **C2** |
| 30 | 27.5 | **14** | **C2** |
| 11 | 8.5 | **5** | **C2** |
| 25 | 22.5 | **9** | **C2** |

**Updated centroid.**

C1 = (2+3+4)/3 = **3** C2 = (10+12+20+30+11+25)/6 = **18**

|  |  |  |
| --- | --- | --- |
| **Centroid** | **C1** | **C2** |
| Old | 2.5 | 16 |
| New | 3 | 18 |

Calculate the mean of new assigned centroid values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **C1 (3)** | **C2 (18)** | **Assigned Cluster base on nearest value** |
| 2 | **1** | 16 | **C1** |
| 4 | **1** | 14 | **C1** |
| 10 | **7** | 8 | **C1** |
| 12 | 9 | **6** | **C2** |
| 3 | **0** | 15 | **C1** |
| 20 | 17 | **2** | **C2** |
| 30 | 27 | **12** | **C2** |
| 11 | 8 | **7** | **C2** |
| 25 | 22 | **7** | **C2** |

**Updated centroid.**

C1 = (2+3+4+10)/4 = **4.75** C2 = (12+20+30+11+25)/5 = **19.6**

|  |  |  |
| --- | --- | --- |
| **Centroid** | **C1** | **C2** |
| Old | 3 | 18 |
| New | 4.75 | 19.6 |

Calculate the mean of new assigned centroid values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **C1 (4.75)** | **C2 (19.6)** | **Assigned Cluster base on nearest value** |
| 2 | **2.75** | 17.6 | **C1** |
| 4 | **0.75** | 15.6 | **C1** |
| 10 | **5.25** | 9.6 | **C1** |
| 12 | **7.25** | 7.6 | **C1** |
| 3 | **1.75** | 16.6 | **C1** |
| 20 | 15.25 | **0.4** | **C2** |
| 30 | 25.5 | **10.4** | **C2** |
| 11 | **6.25** | 8.6 | **C1** |
| 25 | 20.25 | **5.4** | **C2** |

**Updated centroid.**

C1 = (2+3+4+10+11+12)/6 = **7** C2 = (20+30+25)/3 = **25**

|  |  |  |
| --- | --- | --- |
| **Centroid** | **C1** | **C2** |
| Old | 4.75 | 19.6 |
| New | 7 | 25 |

Calculate the mean of new assigned centroid values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **C1 (7)** | **C2 (25)** | **Assigned Cluster base on nearest value** |
| 2 | **5** | 23 | **C1** |
| 4 | **3** | 21 | **C1** |
| 10 | **3** | 15 | **C1** |
| 12 | 5 | **13** | **C1** |
| 3 | **4** | 22 | **C1** |
| 20 | 13 | **5** | **C2** |
| 30 | 23 | **5** | **C2** |
| 11 | 4 | **14** | **C1** |
| 25 | 18 | **0** | **C2** |

**Updated centroid.**

C1 = (2+3+4+10+11+12)/6 = **7** C2 = (20+30+25)/3 = **25**

|  |  |  |
| --- | --- | --- |
| **Centroid** | **C1** | **C2** |
| Old | 7 | 25 |
| New | 7 | 25 |

When there is no movement of points from one cluster to another. That is centroid of the previous iteration and current iteration is same. This leads to convergence of K Means.

**K-means Clustering for 1 Dimensional Data**

Dataset {2,3,6,8,9,12,15,18,22}

K=3

Randomly picking the centroid value of the cluster from given dataset.

e.g., C1=2, C2=8, C3=15