# K-Means Clustering Algorithms

# (Unsupervised K-Means)

# Steps:

# Identifying K numbers of Centroids

# (A centroid is the imaginary or real location representing the center of the cluster.)

# Allocating every data point to nearest cluster

# (A Cluster refers to a collection of data points aggregated together because of certain similarities.)

# Problem Solving

# Features:

# Price(P)

# Rating(R)

# (I think we should not use it as rating will be according to user requirement.)

# Quality(Q)

# Searching History(H)

# Product type(PT)

# Data Set 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H |
| P1 | 142 | 0 | 1(R) | 0 |
| P2 | 223 | 1 | 1(R) | 1 |
| P3 | 150 | 0 | 1(R) | 0 |
| P4 | 200 | 1 | 1(R) | 1 |
| P5 | 120 | 1 | 2(W) | 0 |

# K=2:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(142,0,1,0) | c2(223,1,1,1) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 0 | 81 | C1 |
| P2 | 223 | 1 | 1(R) | 1 | 81 | 0 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 8 | 73 | C1 |
| P4 | 200 | 1 | 1(R) | 1 | 58 | 23 | C2 |
| P5 | 120 | 1 | 2(W) | 0 | 22 | 103 | C1 |

# Formulation:

# c1 of P1=Squaeroot[(P – 142)^2+ (Q-0)^2+ (Pt-1)^2+(H-0)^2]

# c2 of P2= Squaeroot[(P – 223)^2+ (Q-1)^2+ (Pt-1)^2+(H-1)^2]

# .

# Updation:

# C1 = (137, 0, 1, 0)

# C2 = (212, 1, 1, 1)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | C1(137, 0, 1, 0) | C2(212, 1, 1, 1) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 5 | 70 | C1 |
| P2 | 223 | 1 | 1(R) | 1 | 86 | 11 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 13 | 62 | C1 |
| P4 | 200 | 1 | 1(R) | 1 | 63 | 12 | C2 |
| P5 | 120 | 1 | 2(W) | 0 | 17 | 92 | C1 |

# Values are related to same as previous so no need for more updation.

# .

# K=3:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | C1(142, 0, 1, 0) | C2(223, 1, 1, 1) | C3(120, 1, 2, 0) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 0 | 81 | 22 | C1 |
| P2 | 223 | 1 | 1(R) | 1 | 81 | 0 | 103 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 8 | 73 | 30 | C1 |
| P4 | 200 | 1 | 1(R) | 1 | 58 | 23 | 80 | C2 |
| P5 | 120 | 1 | 2(W) | 0 | 22 | 103 | 0 | C3 |

# Updation:

# C1= (137,0,1,0)

# C2= (211.5,1,1,1)

# C3= (120,1,2,0)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(137,0,1,0) | c2(211.5,1,1,1) | C3(120,1,2,0) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 0 | 69 | 22 | C1 |
| P2 | 223 | 1 | 1(R) | 1 | 81 | 11.5 | 103 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 8 | 61 | 30 | C1 |
| P4 | 200 | 1 | 1(R) | 1 | 58 | 11.5 | 80 | C2 |
| P5 | 120 | 1 | 2(W) | 0 | 22 | 91.5 | 0 | C3 |

# Values are related to same as previous so no need for more updation.

# K=4:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(142,0,1,0) | c2(223,1,1,1) | C3(120,1,2,0) | C4(60,0,2,0) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 0 | 81 | 22 | 82 | C1 |
| P2 | 223 | 1 | 1(R) | 1 | 81 | 0 | 103 | 163 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 8 | 73 | 30 | 90 | C1 |
| P4w | 200 | 1 | 1(R) | 1 | 58 | 23 | 80 | 140 | C2 |
| P5 | 120 | 1 | 2(W) | 0 | 22 | 103 | 0 | 60 | C3 |
| P5 | 180 | 0 | 1(R) | 1 | 38 | 43 | 60 | 120 | C1 |
| P6 | 100 | 1 | 2(W) | 1 | 42 | 123 | 20 | 40 | C3 |
| P7 | 60 | 0 | 2(W) | 0 | 82 | 163 | 60 | 0 | C4 |

# Updation:

# C1= (137,0,1,0)

# C2=(211.5,1,1,1)

# C3=(120,1,2,0)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(137,0,1,0) | c2(211.51,1,1) | C3(120,1,2,0) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 0 | 81 | 22 | C1 |
| P2 | 223 | 1 | 1(R) | 1 | 81 | 0 | 103 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 8 | 73 | 30 | C1 |
| P4 | 200 | 1 | 1(R) | 1 | 58 | 23 | 80 | C2 |
| P5 | 120 | 1 | 2(W) | 0 | 22 | 103 | 0 | C1 |

# Values are related to same as previous so no need for more updation.

# Data Set2:

# (With History 1 && k=2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(120,0,2,1) | c2(142,0,1,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 22 | 0 | C2 |
| P2 | 223 | 1 | 2(W) | 1 | 103 | 81 | C2 |
| P3 | 150 | 0 | 1(R) | 1 | 30 | 8 | C2 |
| P4 | 200 | 1 | 1(R) | 1 | 80 | 58 | C2 |
| P5 | 120 | 0 | 2(W) | 1 | 0 | 22 | C1 |
| P6 | 175 | 1 | 1(R) | 1 | 55 | 33 | C2 |
| P7 | 200 | 1 | 2(W) | 1 | 80 | 58 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 30 | 52 | C1 |
| P9 | 150 | 0 | 2(W) | 1 | 30 | 8 | C2 |

# Updation:

# C1=(105,0,1.5,1)

# C2=(177,0.5,1.4,1)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(105,0,1.5,1) | c2(177,0.5,1.4,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 37 | 35 | C2 |
| P2 | 223 | 1 | 2(W) | 1 | 118 | 46 | C2 |
| P3 | 150 | 0 | 1(R) | 1 | 45 | 27 | C2 |
| P4 | 200 | 1 | 1(R) | 1 | 95 | 23 | C2 |
| P5 | 120 | 0 | 2(W) | 1 | 15 | 57 | C1 |
| P6 | 175 | 1 | 1(R) | 1 | 70 | 2 | C2 |
| P7 | 200 | 1 | 2(W) | 1 | 95 | 23 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 15 | 87 | C1 |
| P9 | 150 | 0 | 2(W) | 1 | 45 | 27 | C2 |

# (With History 0 && k=2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(120,0,2,0) | c2(142,0,1,0) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 22 | 0 | C2 |
| P2 | 223 | 1 | 2(W) | 0 | 103 | 81 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 30 | 8 | C2 |
| P4 | 200 | 1 | 1(R) | 0 | 80 | 58 | C2 |
| P5 | 120 | 0 | 2(W) | 0 | 0 | 22 | C1 |
| P6 | 175 | 1 | 1(R) | 0 | 55 | 33 | C2 |
| P7 | 200 | 1 | 2(W) | 0 | 80 | 58 | C2 |
| P8 | 90 | 0 | 1(R) | 0 | 30 | 52 | C1 |
| P9 | 150 | 0 | 2(W) | 0 | 30 | 8 | C2 |

# Updation:

# C1=(105,0,1.5,0)

# C2=(177,0.5,1.4,0)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(105,0,1.5,0) | c2(177,0.5,1.4,0) | Related to |
| P1 | 142 | 0 | 1(R) | 0 | 37 | 35 | C2 |
| P2 | 223 | 1 | 2(W) | 0 | 118 | 46 | C2 |
| P3 | 150 | 0 | 1(R) | 0 | 45 | 27 | C2 |
| P4 | 200 | 1 | 1(R) | 0 | 95 | 23 | C2 |
| P5 | 120 | 0 | 2(W) | 0 | 15 | 57 | C1 |
| P6 | 175 | 1 | 1(R) | 0 | 70 | 2.1 | C2 |
| P7 | 200 | 1 | 2(W) | 0 | 95 | 23 | C2 |
| P8 | 90 | 0 | 1(R) | 0 | 15 | 87 | C1 |
| P9 | 150 | 0 | 2(W) | 0 | 45 | 27 | C2 |

# Data Set3:

# (With History 1 && k=2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(200,1,1,1) | c2(200,1,2,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 58 | 58 | - |
| P2 | 223 | 1 | 2(W) | 1 | 23 | 23 | - |
| P3 | 150 | 0 | 1(R) | 1 | 50 | 50 | - |
| P4 | 200 | 1 | 1(R) | 1 | 0 | 1 | C1 |
| P5 | 120 | 0 | 2(W) | 1 | 80 | 80 | - |
| P6 | 175 | 1 | 1(R) | 1 | 25 | 25 | - |
| P7 | 200 | 1 | 2(W) | 1 | 1 | 0 | C1 |
| P8 | 90 | 0 | 1(R) | 1 | 110 | 110 | - |
| P9 | 150 | 0 | 2(W) | 1 | 50 | 50 | - |

# Solving the problem of Centroid Values by increasing the value of K : K=3

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(200,1,1,1) | c2(200,1,2,1) | C3(150,0,2,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 58 | 58 | 10 | C3 |
| P2 | 223 | 1 | 2(W) | 1 | 23 | 23 | 73 | - |
| P3 | 150 | 0 | 1(R) | 1 | 50 | 50 | 1 | C3 |
| P4 | 200 | 1 | 1(R) | 1 | 0 | 1 | 50 | C1 |
| P5 | 120 | 0 | 2(W) | 1 | 80 | 80 | 30 | C3 |
| P6 | 175 | 1 | 1(R) | 1 | 25 | 25 | 25 | - |
| P7 | 200 | 1 | 2(W) | 1 | 1 | 0 | 50 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 110 | 110 | 60 | C3 |
| P9 | 150 | 0 | 2(W) | 1 | 50 | 50 | 0 | C3 |

# Solving the problem of Centroid Values by increasing the value of K: K=4

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(200,1,1,1) | c2(200,1,2,1) | C3(150,0,2,1) | C4(142,0,1,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 58 | 58 | 10 | 0 | C4 |
| P2 | 223 | 1 | 2(W) | 1 | 23 | 23 | 73 | 81 | -c2 |
| P3 | 150 | 0 | 1(R) | 1 | 50 | 50 | 1 | 8 | C3 |
| P4 | 200 | 1 | 1(R) | 1 | 0 | 1 | 50 | 58 | C1 |
| P5 | 120 | 0 | 2(W) | 1 | 80 | 80 | 30 | 22 | C4 |
| P6 | 175 | 1 | 1(R) | 1 | 25 | 25 | 25 | 33 | -c1 |
| P7 | 200 | 1 | 2(W) | 1 | 1 | 0 | 50 | 58 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 110 | 110 | 60 | 52 | C4 |
| P9 | 150 | 0 | 2(W) | 1 | 50 | 50 | 0 | 8 | C3 |

# Updation:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(187.5,1,1,1) | c2(211.5,1,2,1) | C3(150,0,1.5,1) | C4(81.3,0,2,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 45.5 | 69.5 | 8 | 60.7 | C3 |
| P2 | 223 | 1 | 2(W) | 1 | 35.5 | 11.5 | 73 | 141.7 | C2 |
| P3 | 150 | 0 | 1(R) | 1 | 37.5 | 61.5 | 0 | 68.7 | C3 |
| P4 | 200 | 1 | 1(R) | 1 | 12.5 | 11.5 | 50 | 118.7 | C2 |
| P5 | 120 | 0 | 2(W) | 1 | 67.5 | 91.5 | 30 | 38.7 | C3 |
| P6 | 175 | 1 | 1(R) | 1 | 12.5 | 36.5 | 25 | 93.7 | C1 |
| P7 | 200 | 1 | 2(W) | 1 | 12.5 | 11.5 | 50 | 118.7 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 97.5 | 121.5 | 60 | 8.7 | C4 |
| P9 | 150 | 0 | 2(W) | 1 | 37.5 | 61.5 | 0 | 68.7 | C3 |

# Updation:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | c1(187.5,1,1,1) | c2(207.6,1,1.6,1) | C3(140,0,1.5,1) | C4(90,0,1,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 45.5 | 65.6 | 2 | 52 | C3 |
| P2 | 223 | 1 | 2(W) | 1 | 37.5 | 15.4 | 83 | 133 | C2 |
| P3 | 150 | 0 | 1(R) | 1 | 37.5 | 57.6 | 10 | 60 | C3 |
| P4 | 200 | 1 | 1(R) | 1 | 12.5 | 7.6 | 60 | 110 | C2 |
| P5 | 120 | 0 | 2(W) | 1 | 67.5 | 87.6 | 20 | 30 | C3 |
| P6 | 175 | 1 | 1(R) | 1 | 12.5 | 32.6 | 35 | 85 | C1 |
| P7 | 200 | 1 | 2(W) | 1 | 12.5 | 7.6 | 60 | 110 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 97.5 | 117.6 | 50 | 0 | C4 |
| P9 | 150 | 0 | 2(W) | 1 | 37.5 | 57.6 | 10 | 60 | C3 |

# // While keeping k = 4, we get better results.

# With Additional feature Rating:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | R | c1(200,1,1,1,4) | c2(200,1,2,1,3) | C3(150,0,2,1,4) | C4(142,0,1,1,1) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 1 | 58 | 58 | 8.6 | 0 | C4 |
| P2 | 223 | 1 | 2(W) | 1 | 2 | 23.1 | 23 | 73 | 81 | C2 |
| P3 | 150 | 0 | 1(R) | 1 | 3 | 50 | 50 | 1.4 | 8.2 | C3 |
| P4 | 200 | 1 | 1(R) | 1 | 4 | 0 | 1.4 | 50 | 58 | C1 |
| P5 | 120 | 0 | 2(W) | 1 | 5 | 80 | 80 | 30 | 22.3 | C4 |
| P6 | 175 | 1 | 1(R) | 1 | 4 | 25 | 25 | 25 | 33.1 | -c1 |
| P7 | 200 | 1 | 2(W) | 1 | 3 | 1.4 | 0 | 50 | 58 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 2 | 110 | 110 | 60 | 52 | C4 |
| P9 | 150 | 0 | 2(W) | 1 | 4 | 50 | 50 | 0 | 8.6 | C3 |

# Updation:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No of point | P | Q | PT | H | R | c1(187.5,1,1,1,4) | c2(211.5,1,2,3) | C3(150,0,1.5,1,3.25) | C4(81.3,0,2,1,2) | Related to |
| P1 | 142 | 0 | 1(R) | 1 | 1 | 45.6 | 71.5 | 8.3 | 60 | C3 |
| P2 | 223 | 1 | 2(W) | 1 | 2 | 35.5 | 11.6 | 73 | 141.7 | C2 |
| P3 | 150 | 0 | 1(R) | 1 | 3 | 37.5 | 61.5 | 0.5 | 68.7 | C3 |
| P4 | 200 | 1 | 1(R) | 1 | 4 | 12.5 | 11.5 | 50 | 118 | C2 |
| P5 | 120 | 0 | 2(W) | 1 | 5 | 67.5 | 91.5 | 30 | 38 | C3 |
| P6 | 175 | 1 | 1(R) | 1 | 4 | 12.5 | 36.5 | 25 | 93.7 | C1 |
| P7 | 200 | 1 | 2(W) | 1 | 3 | 12.5 | 11.5 | 50 | 118.7 | C2 |
| P8 | 90 | 0 | 1(R) | 1 | 2 | 97.5 | 121.5 | 60 | 8.7 | C4 |
| P9 | 150 | 0 | 2(W) | 1 | 4 | 37.5 | 61.5 | 0.9 | 68.7 | C3 |

# With Rating we have to update result 2 times.

# Without Rating we have to update result 3 times.

# As the result is same no need for more iterations.