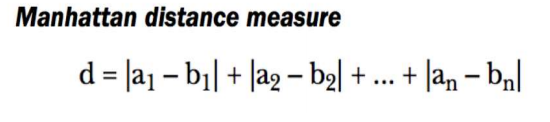
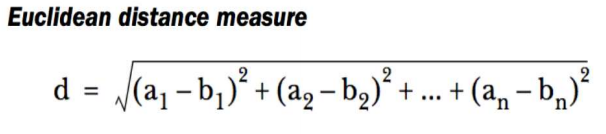
1. [75 pts] Given two initial cluster centers at (1, 1) and (2, 1), use *k*-means algorithm to decide the clusters based on Euclidean distance measure with the termination condition of 3 iterations. Please show the intermediate results in each iteration.
2. Manhattan distance is given by the following equation

Hence Manhattan distance between point (1,2) and (9,8) is |1-9| + |2-8| 🡺 **14**

1. Using centroids (1,1) and (2,1) first we calculate Euclidean distances of all the data points from the initial centroids

Euclidean distance is given by



Hence distances for each point from initial centroids were calculated as below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X1** | **X2** | **Dist from C1 (1,1)** | **Dist from C2 (2,1)** | **Cluster assignment** |
| 1 | 1 | 0 | 1 | 1 |
| 2 | 1 | 1 | 0 | 2 |
| 1 | 2 | 1 | 1.414213562 | 1 |
| 2 | 2 | 1.414213562 | 1 | 2 |
| 3 | 3 | 2.828427125 | 2.236067977 | 2 |
| 8 | 8 | 9.899494937 | 9.219544457 | 2 |
| 8 | 9 | 10.63014581 | 10 | 2 |
| 9 | 8 | 10.63014581 | 9.899494937 | 2 |
| 9 | 9 | 11.3137085 | 10.63014581 | 2 |

Based on the mean distance we rearrange the clusters and calculate mean centeroid, for cluster 1 we get (1, 1.5) and cluster 2 centeroid will be (5.858, 5.714) we get the following eucledian distances and rearrange the clusters as follows

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X1** | **X2** | **Dist from C1 (1,1.5)** | **Dist from C2 (5.858,5.714)** | **Cluster assignment** |
| 1 | 1 | 0.5 | 6.769191975 | 1 |
| 1 | 2 | 0.5 | 6.115060098 | 1 |
| 2 | 1 | 1.118033989 | 6.091466162 | 1 |
| 2 | 2 | 1.118033989 | 5.355180669 | 1 |
| 3 | 3 | 2.5 | 3.941314502 | 1 |
| 8 | 8 | 9.552486587 | 3.132724054 | 2 |
| 8 | 9 | 10.25914226 | 3.922494105 | 2 |
| 9 | 8 | 10.30776406 | 3.885609347 | 2 |
| 9 | 9 | 10.9658561 | 4.54642277 | 2 |

Updated centeroids for the clusters are (1.8, 1.8) and (8.5,8.5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X1** | **X2** | **Dist from C1 (1.8,1.8)** | **Dist from C2 (8.5,8.5)** | **Cluster assignment** |
| 1 | 1 | 1.13137085 | 10.60660172 | 1 |
| 1 | 2 | 0.824621125 | 9.924716621 | 1 |
| 2 | 1 | 0.824621125 | 9.924716621 | 1 |
| 2 | 2 | 0.282842712 | 9.192388155 | 1 |
| 3 | 3 | 1.697056275 | 7.778174593 | 1 |
| 8 | 8 | 8.768124087 | 0.707106781 | 2 |
| 8 | 9 | 9.501578816 | 0.707106781 | 2 |
| 9 | 8 | 9.501578816 | 0.707106781 | 2 |
| 9 | 9 | 10.18233765 | 0.707106781 | 2 |

Hence after three iterations we acertain that cluster one has [(1,1), (1,2),(2,1),(2,2),(3,3)] and cluster 2 contains [(8,8),(8,9),(9,8),(9,9)]