# A Simple example showing the implementation of k-means algorithm (using K=2)

| 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 |

**Step 1**:

Initialization: Randomly we choose following two centroids (k=2) for two clusters.

In this case the 2 centroid are: m1=(1.0,1.0) and m2=(5.0,7.0).

| 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 |

| | Individual | Mean Vector |
|---|---|---|
| Group 1 | 1 | (1.0, 1.0) |
| Group 2 | 4 | (5.0, 7.0) |

**Step 2:**

- Thus, we obtain two clusters containing:

  {1,2,3} and {4,5,6,7}.
- Their new centroids are:

$$m_1 = \left(\frac{1}{3}(1.0+1.5+3.0), \frac{1}{3}(1.0+2.0+4.0)\right) = (1.83, 2.33)$$

$$m_2 = \left(\frac{1}{4}(5.0+3.5+4.5+3.5), \frac{1}{4}(7.0+5.0+5.0+4.5)\right)$$

$$= (4.12, 5.38)$$

| Individual | Centroid 1 | Centroid 2 |
|---|---|---|
| 1 | 0 | 7.21 |
| 2 (1.5, 2.0) | 1.12 | 6.10 |
| 3 | 3.61 | 3.61 |
| 4 | 7.21 | 0 |
| 5 | 4.72 | 2.5 |
| 6 | 5.31 | 2.06 |
| 7 | 4.30 | 2.92 |

$$d(m_1, 2) = \sqrt{|1.0-1.5|^2 + |1.0-2.0|^2} = 1.12$$

$$d(m_2, 2) = \sqrt{|5.0-1.5|^2 + |7.0-2.0|^2} = 6.10$$

**Step 3:**

- Now using these centroids we compute the Euclidean distance of each object, as shown in table.

- Therefore, the new clusters are:
  {1,2} and {**3**,4,5,6,7}

- Next centroids are: m1=(1.25,1.5) and m2 = (3.9,5.1)

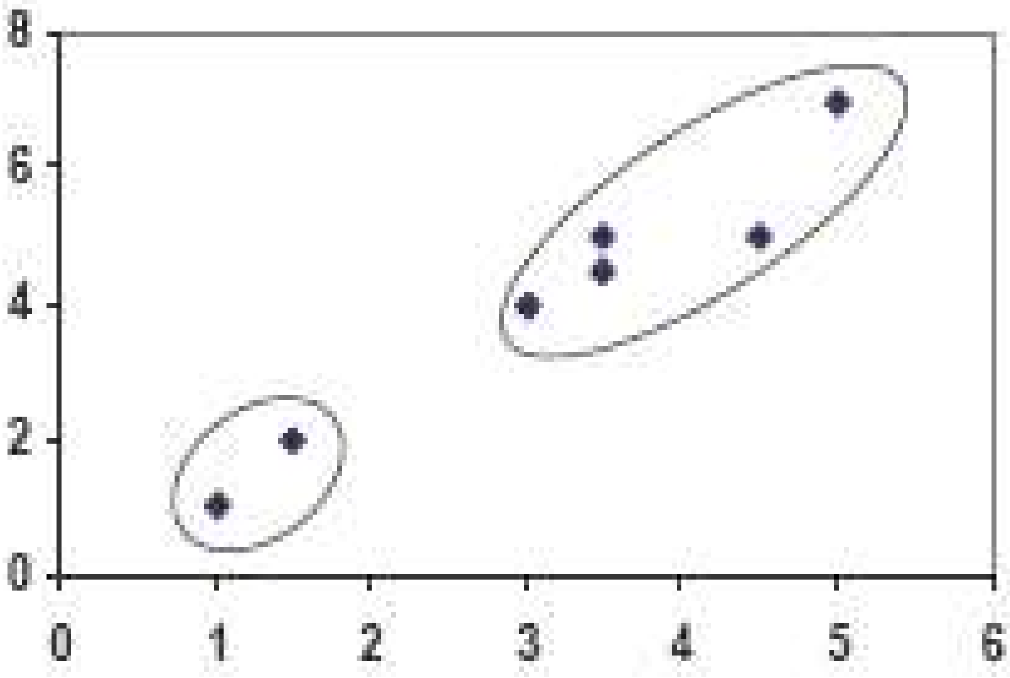
| Individual | Centroid 1 | Centroid 2 |
|---|---|---|
| 1 | 1.57 | 5.38 |
| 2 | 0.47 | 4.28 |
| 3 | 2.04 | 1.78 |
| 4 | 5.64 | 1.84 |
| 5 | 3.15 | 0.73 |
| 6 | 3.78 | 0.54 |
| 7 | 2.74 | 1.08 |

- Step 4 :
  The clusters obtained are:
  {1,2} and {3,4,5,6,7}

- Therefore, there is no change in the cluster.
- Thus, the algorithm comes to a halt here and final result consist of 2 clusters {1,2} and {3,4,5,6,7}.

| Individual | Centroid 1 | Centroid 2 |
|---|---|---|
| 1 | 0.56 | 5.02 |
| 2 | 0.56 | 3.92 |
| 3 | 3.05 | 1.42 |
| 4 | 6.66 | 2.20 |
| 5 | 4.16 | 0.41 |
| 6 | 4.78 | 0.61 |
| 7 | 3.75 | 0.72 |

# PLOT



# Clustering Evaluation-Example

| 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 |

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} d^2(m_i, x)$$

$$d(x, y) = \sum_{i=1}^{p} |x_i - y_i|$$

Therefore, the new clusters are:

Cluster 1: {1,2}

Cluster 2: {**3**,4,5,6,7}

Centroids are:

m1 = (1.25,1.5)

m2 = (3.9,5.1)

**Calculate the SSE using <u>Manhattan Distance</u>**

**SSE=[dis²(m1,x1)+dis²(m1,x2)] +**
**[dis²(m2,x3)+dis²(m2,x4) + dis²(m2,x5)+ dis²(m2,x6) + dis²(m2,x7)]**

dis²(m1,x1) = (|1.25-1|)+|1.5-1|) ² = 0.5625
dis²(m1,x2) = (|1.25-1.5|)+|1.5-2|) ² = 0.5625

dis²(m2,x3) = (|3.9-3.0|)+|5.1-4.0|) ² = 4.0
dis²(m2,x4) = (|3.9-5.0|)+|5.1-7.0|) ² = 9.0
dis²(m2,x5) = (|3.9-3.5|)+|5.1-5.0|) ² = 0.25
dis²(m2,x6) = (|3.9-4.5)+|5.1-5.0|) ² = 0.49
dis²(m2,x7) = (|3.9-3.5|)+|5.1-4.5|) ² = 1.0

**SSE= (**0.5625+
0.5625+4.0+9.0+0.25+0.49+1.0**)**
**= 15.865**

## (with K=3)

| Individual | $m_1 = 1$ | $m_2 = 2$ | $m_3 = 3$ | cluster |
|---|---|---|---|---|
| 1 | 0 | 1.11 | 3.61 | 1 |
| 2 | 1.12 | 0 | 2.5 | 2 |
| 3 | 3.61 | 2.5 | 0 | 3 |
| 4 | 7.21 | 6.10 | 3.61 | 3 |
| 5 | 4.72 | 3.61 | 1.12 | 3 |
| 6 | 5.31 | 4.24 | 1.80 | 3 |
| 7 | 4.30 | 3.20 | 0.71 | 3 |

$C_3$

clustering with initial centroids (1, 2, 3)

**Step 1**

| Individual | $m_1$ (1.0, 1.0) | $m_2$ (1.5, 2.0) | $m_3$ (3.9,5.1) | cluster |
|---|---|---|---|---|
| 1 | 0 | 1.11 | 5.02 | 1 |
| 2 | 1.12 | 0 | 3.92 | 2 |
| 3 | 3.61 | 2.5 | 1.42 | 3 |
| 4 | 7.21 | 6.10 | 2.20 | 3 |
| 5 | 4.72 | 3.61 | 0.41 | 3 |
| 6 | 5.31 | 4.24 | 0.61 | 3 |
| 7 | 4.30 | 3.20 | 0.72 | 3 |

**Step 2**

# Example

- Exercise:

Calculate the SSE when K=3?

# PLOT