Fuzzy C-Means

Introduction

- a fuzzy clustering algorithm – data points can belong to more then one clustering

- in hard clustering data is divided into clusters, where each data point can only belong to exactly one cluster. In fuzzy clustering, data points can potentially belong to multiple clusters.

- membership to a cluster → we assign membership grades to each data point

- membership grades = the degree to which data points belong to each cluster → data points at the edge may be in the cluster to a lesser degree

- FCM was developed by J.C. Dunn in 1973, and improved by J.C. Bezdek in 1981

Description

- FCM is similar to K-Means

- It aims to partition a finite collection of *n* elements into a collection of *c* fuzzy clusters with respect of a given criterion.

- Steps:

* Choose a number of clusters
* assign data points randomly to clusters
* compute centroid for each cluster
* for each data point compute coefficient of being in the clusters
* repeat until change is less then ε (epsilon, the sensitivity threshold)

Centroid

where m is the hyper-parameter for the clustering → determines how fuzzy the clustering will be (m high → fuzzyness high),

Objective function:

Degree of membership

Example

n=6

c=2

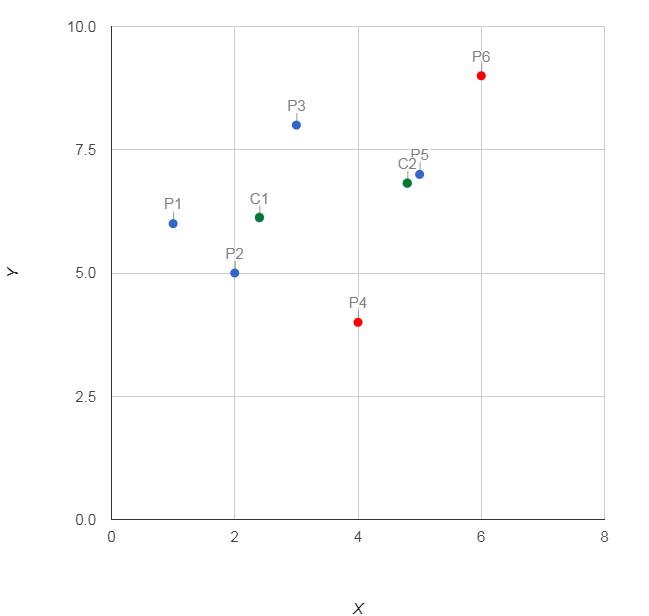
m=2

Random coefficients C1,C2

1.

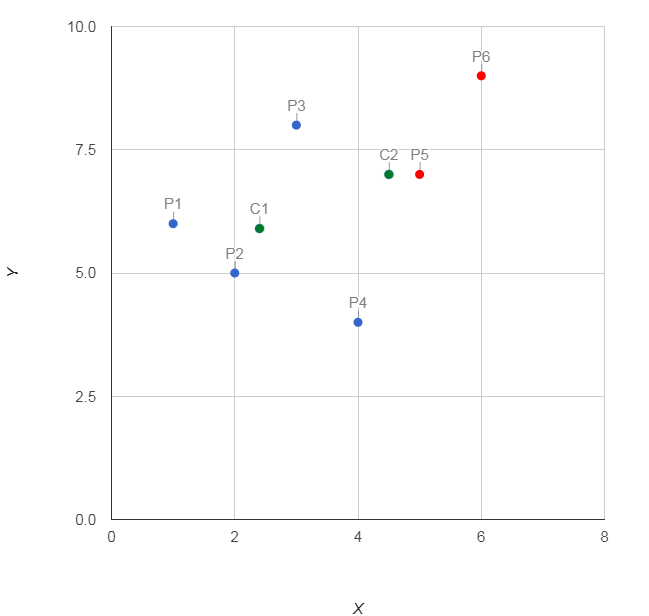
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | C1 | C2 | D(c1) | D(c2) |
| 1 | 6 | 0.8 | 0.2 | 1.40 | 3.88 |
| 2 | 5 | 0.9 | 0.1 | 1.17 | 3.32 |
| 3 | 8 | 0.7 | 0.3 | 1.99 | 2.16 |
| 4 | 4 | 0.3 | 0.7 | 2.64 | 2.91 |
| 5 | 7 | 0.5 | 0.5 | 2.75 | 0.28 |
| 6 | 9 | 0.2 | 0.8 | 4.62 | 2.50 |

Centroids C1 (2.4,6.1) , C2 (4.8,6.8)

2.

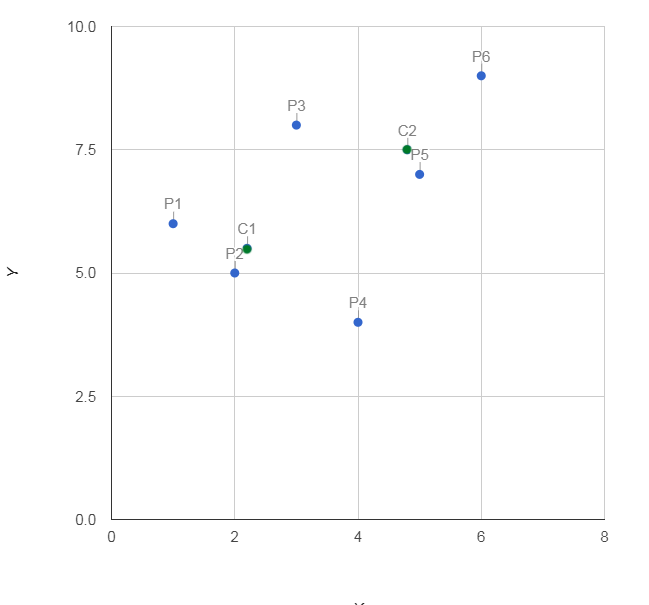
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | C1 | C2 | D(c1) | D(c2) |
| 1 | 6 | 0.7 | 0.3 | 1.45 | 3.61 |
| 2 | 5 | 0.6 | 0.4 | 1.05 | 3.19 |
| 3 | 8 | 0.5 | 0.5 | 2.13 | 1.75 |
| 4 | 4 | 0.5 | 0.5 | 2.49 | 3.07 |
| 5 | 7 | 0.1 | 0.9 | 2.76 | 0.54 |
| 6 | 9 | 0.3 | 0.7 | 4.68 | 2.50 |

Centroids C1 (2.4,5.9) , C2 (4.5,7)



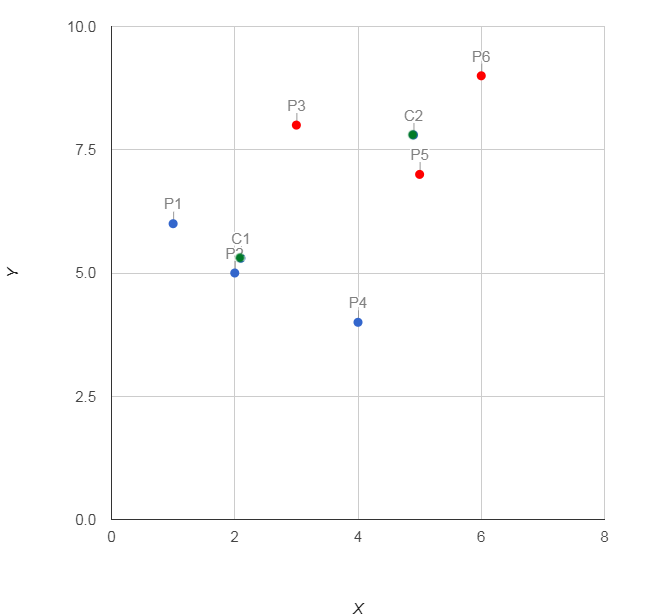
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | C1 | C2 | D(c1) | D(c2) |
| 1 | 6 | 0.9 | 0.1 | 1.27 | 4.11 |
| 2 | 5 | 0.9 | 0.1 | 0.53 | 3.78 |
| 3 | 8 | 0.4 | 0.6 | 2.64 | 1.89 |
| 4 | 4 | 0.6 | 0.4 | 2.37 | 3.61 |
| 5 | 7 | 0.0 | 1.0 | 3.21 | 0.54 |
| 6 | 9 | 0.2 | 0.8 | 5.19 | 1.90 |

Centroids C1 (2.2,5.5) , C2 (4.8,7.5)

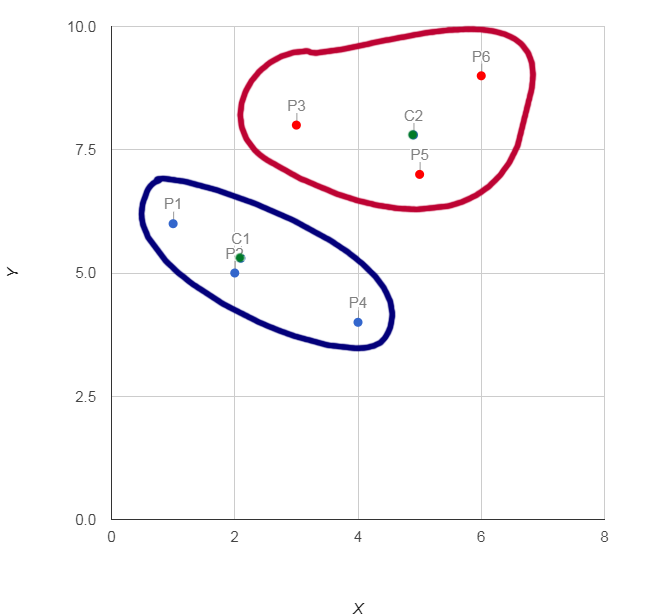


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | C1 | C2 | D(c1) | D(c2) |
| 1 | 6 | 0.9 | 0.1 | 1.32 | 4.28 |
| 2 | 5 | 1.0 | 0.0 | 0.34 | 4.00 |
| 3 | 8 | 0.3 | 0.7 | 2.83 | 1.92 |
| 4 | 4 | 0.7 | 0.3 | 2.28 | 3.86 |
| 5 | 7 | 0.0 | 1.0 | 3.33 | 0.76 |
| 6 | 9 | 0.1 | 0.9 | 5.35 | 1.66 |

Centroids C1 (2.1,5.3) , C2 (4.9,7.8)



Centroids remain stable



Usages

Pros

Cons