UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Fuzzy Clustering

Cristiano Leite de Castro
crislcastro@ufmg.br

Departamento de Engenharia Elétrica
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

Some of the contents and figures in this presentation are taken from “An Introduction to Statistical Learning with applications in R (Springer 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R. Tibshirani.

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

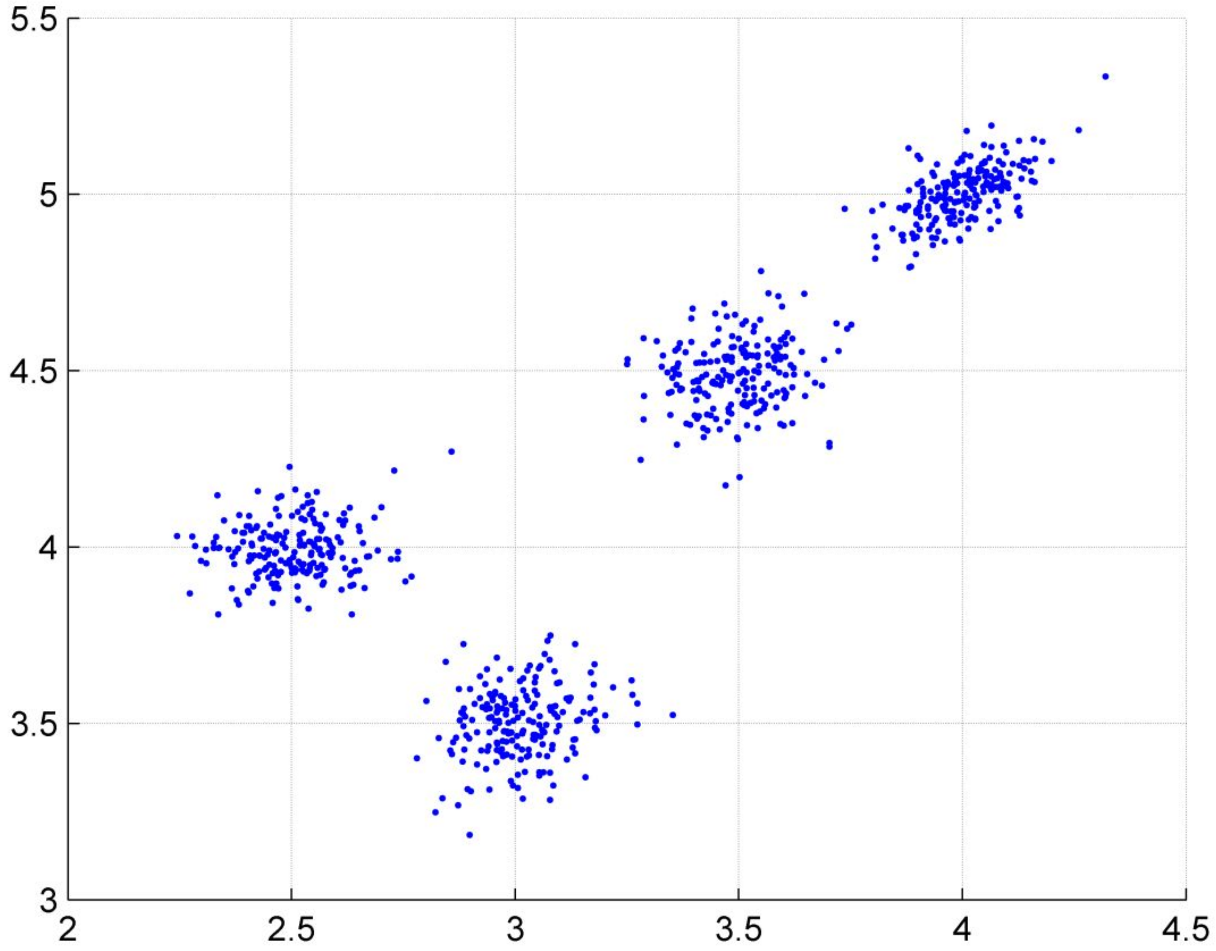
Clustering

- Clustering refers to a very broad set of techniques for finding subgroups, or clusters, in a data set;

- We seek to partition the data into distinct groups so that the observations within each group are quite similar to each other, while observations in different groups are quite different to each other;

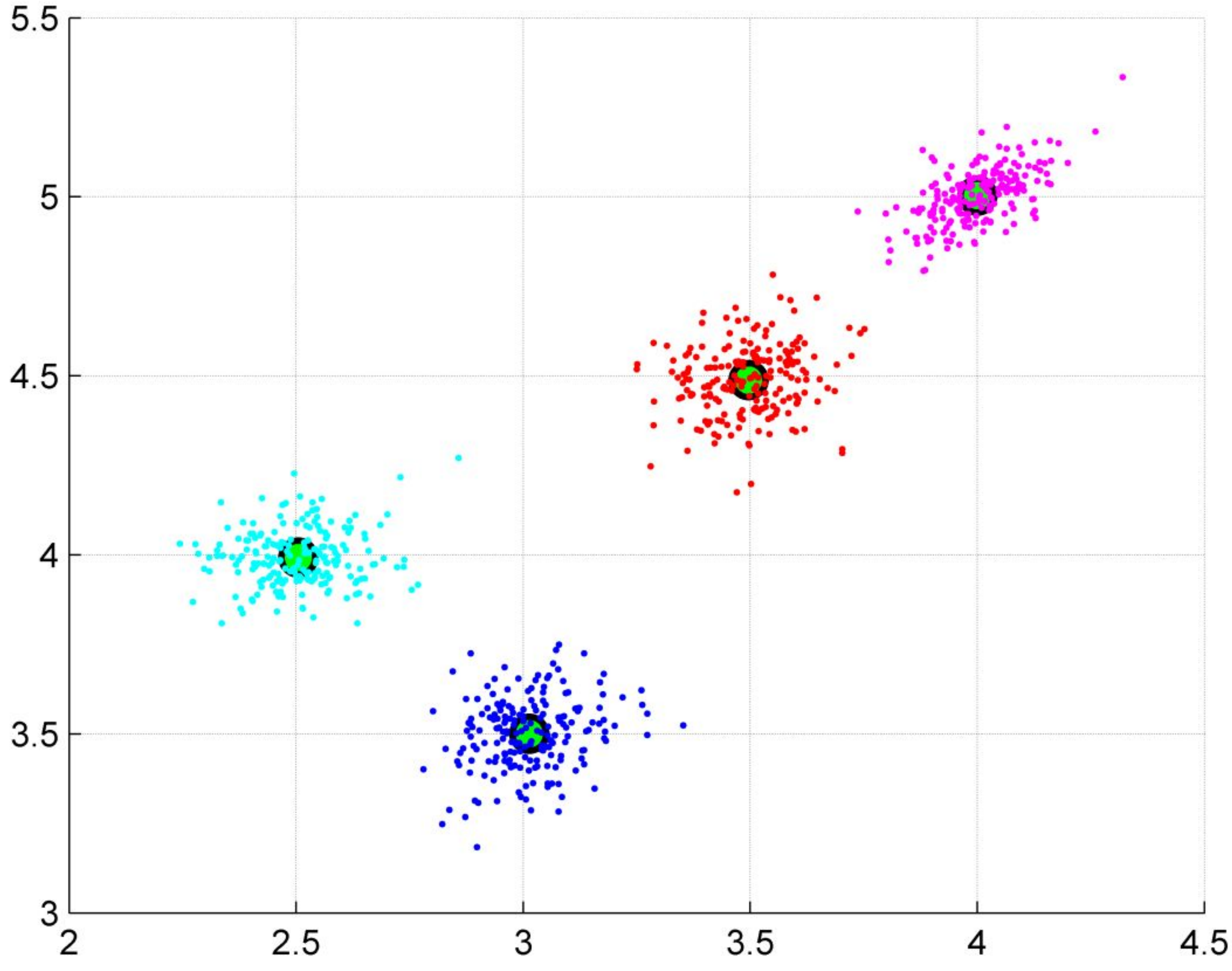
- To make this concrete, we must define what it means for two or more observations to be similar or different.

Clustering



UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Clustering



Clustering Application

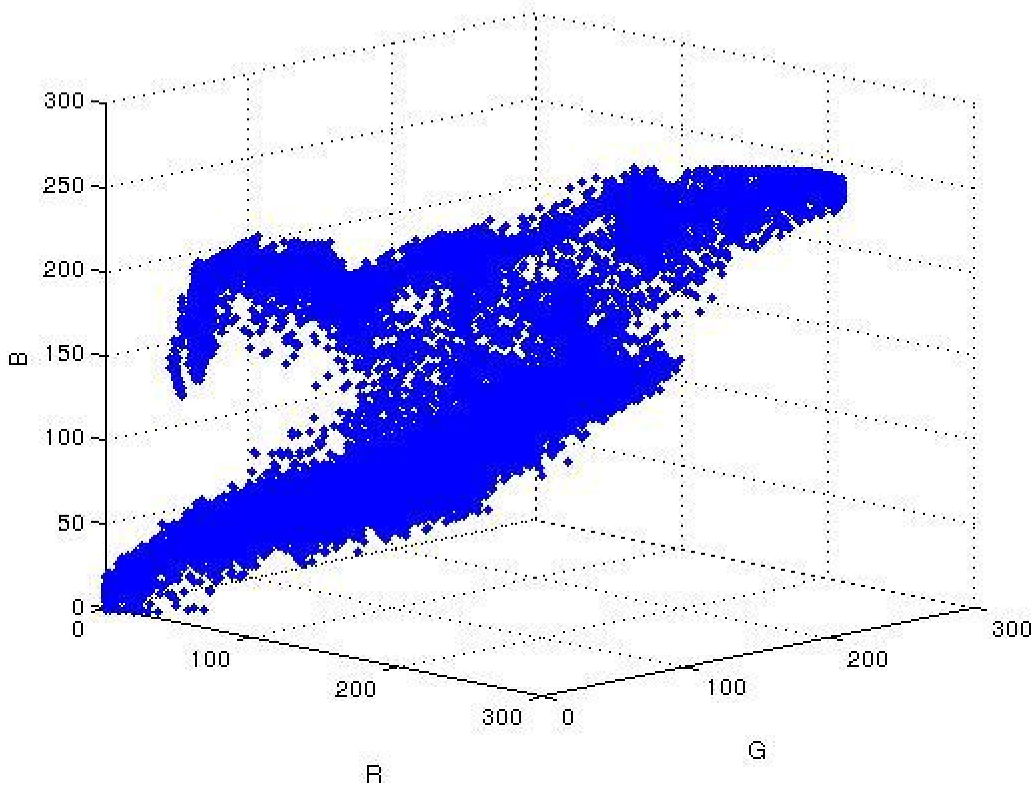
- Clustering for *Market Segmentation*:
  - patterns = a large number of people from a region.
  - features = a large number of measurements: median household income, occupation, distance from nearest urban area, etc.
  - goal: identify subgroups of people who might be more receptive to a particular form of advertising, or more likely to purchase a particular product.
  - The task of performing market segmentation amounts to clustering the people in the data set.
  - It is also suitable for recommender systems in the Internet.

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Clustering Application

- Image Segmentation by Region: clustering image pixels by their RGB values;

Original Image



UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Clustering Application

- Image Segmentation by Region

Original Image

Result

- Partition for Ordinary Sets:
  - family of subsets of A

$$\pi(A) = \{A_i | i \in I, A_i \subseteq A\}$$

such that

$$A_i \cap A_j = \emptyset; \forall i, j \in I, i \neq j$$

and

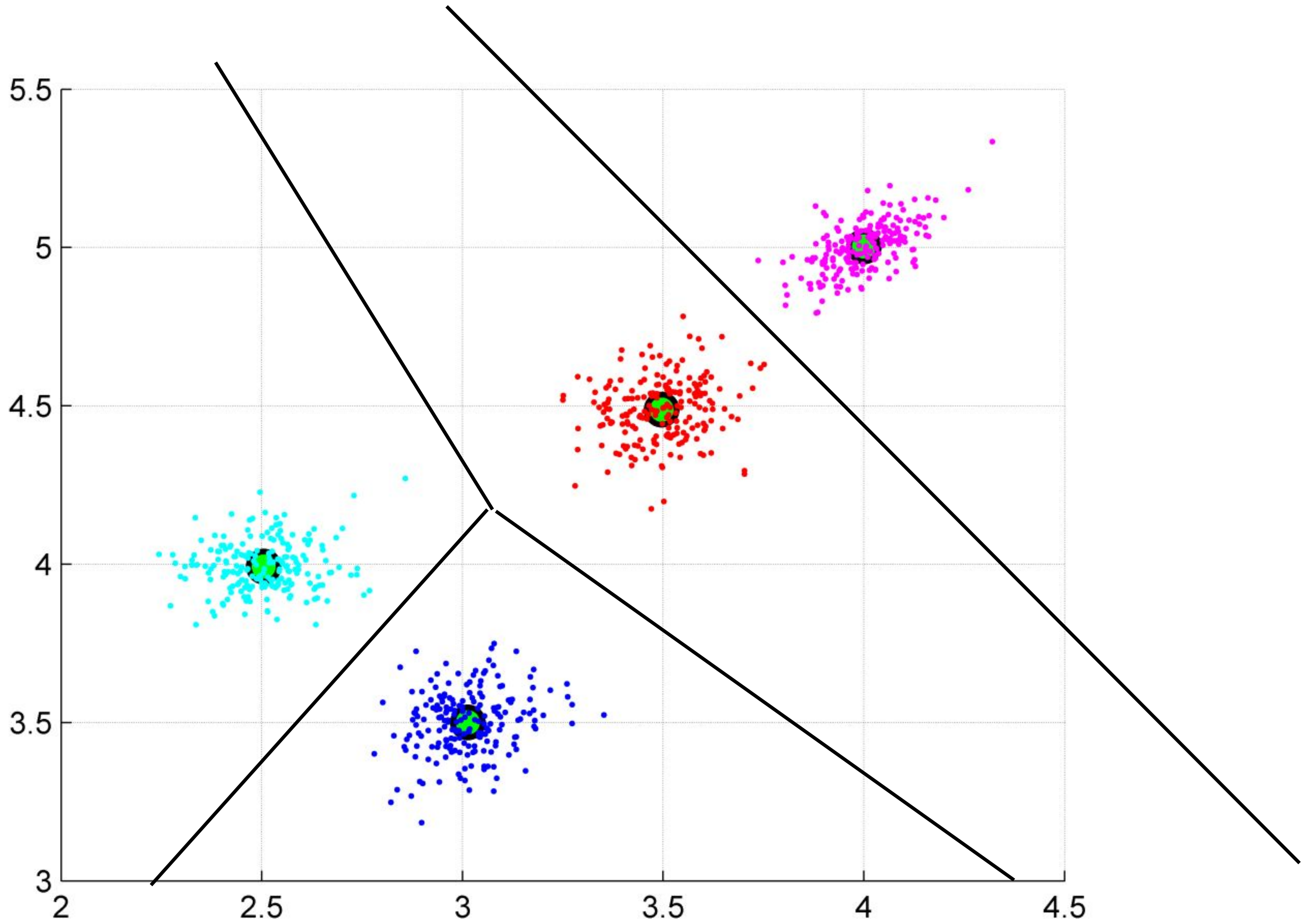
$$\bigcup_{i \in I} A_i = A$$

Partition

A = set of vehicles of Brazil

- possible partitions:

Partition



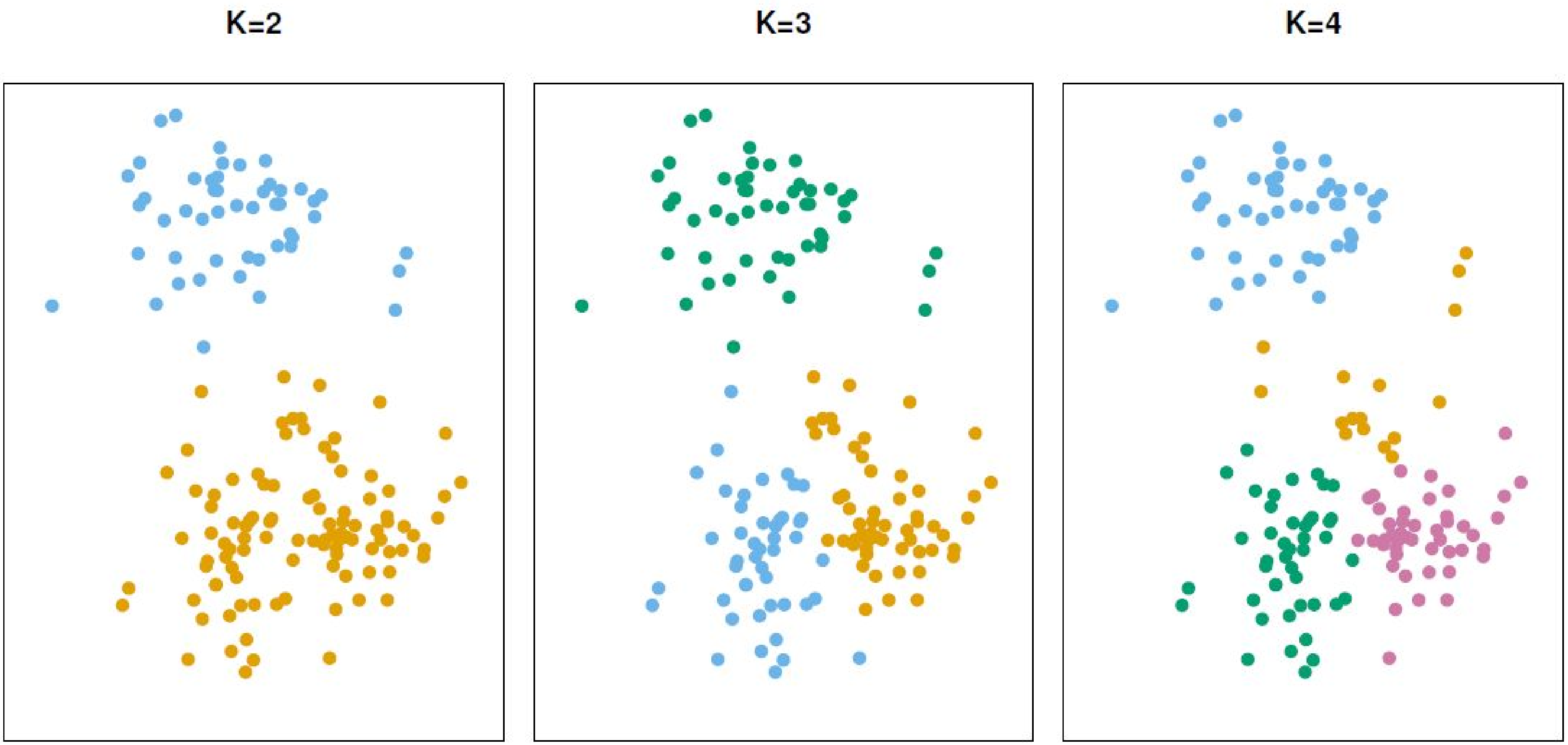
UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

K-Means Clustering

- most commonly used partitional algorithm;
- K-Means partitions the observations into a pre-specified number of clusters = K;
- K-means takes assumptions of ordinary (traditional) sets in order to create the clusters:
  - each observation belongs to a single cluster
  - the clusters are non-overlapping (disjunct sets)

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

K-Means Clustering



UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Details of K-Means Clustering

Let $X = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_i, \ldots, \vec{x}_n]^T$ be the matrix of observations

where each observation (row of X) has *p* features, i.e., $\vec{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{ip}]$

Let $C_1, \ldots, C_K$ denote sets containing the indices of the observations in each cluster. These sets satisfy two properties:

1. $C_1 \cup C_2 \cup \ldots \cup C_K = \{1, \ldots, n\}$. In other words, each observation belongs to at least one of the K clusters.
2. $C_k \cap C_{k'} = \emptyset$ for all $k \neq k'$. In other words, the clusters are non-overlapping: no observation belongs to more than one cluster.

For instance, if the ith observation is in the kth cluster, then $i \in C_k$.

Details of K-Means Clustering

- The idea behind K-means clustering is that a *good* clustering is one for which the *within-cluster variation* is as small as possible.
- The within-cluster variation for cluster C_k is a measure $\mathrm{WCV}(C_k)$ of the amount by which the observations within a cluster differ from each other.
- Hence we want to solve the problem

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} \mathrm{WCV}(C_k) \right\}. \tag{2}$$

- In words, this formula says that we want to partition the observations into K clusters such that the total within-cluster variation, summed over all K clusters, is as small as possible.

How to define within-cluster variation?

- Typically we use Euclidean distance

$$\mathrm{WCV}(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2, \qquad (3)$$

where $|C_k|$ denotes the number of observations in the kth cluster.

- Combining (2) and (3) gives the optimization problem that defines K-means clustering,

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\}. \qquad (4)$$

K-Means Clustering Algorithm

1. Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.
2. Iterate until the cluster assignments stop changing:
   2.1 For each of the K clusters, compute the cluster *centroid*. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster.
   2.2 Assign each observation to the cluster whose centroid is closest (where *closest* is defined using Euclidean distance).

K-Means Clustering Algorithm

- Obtaining the centroid for the k-th Cluster

$$\vec{c}_k = \frac{1}{|C_k|} \sum_{i \in C_k} \vec{x}_i$$

- The partitioned groups (clusters) are typically defined by a *n* x *K* binary membership matrix *U*, where the element

  $u_{i,k} = 1$ if the *i*-th observation $\vec{x}_i$ belongs to cluster C_k

  and

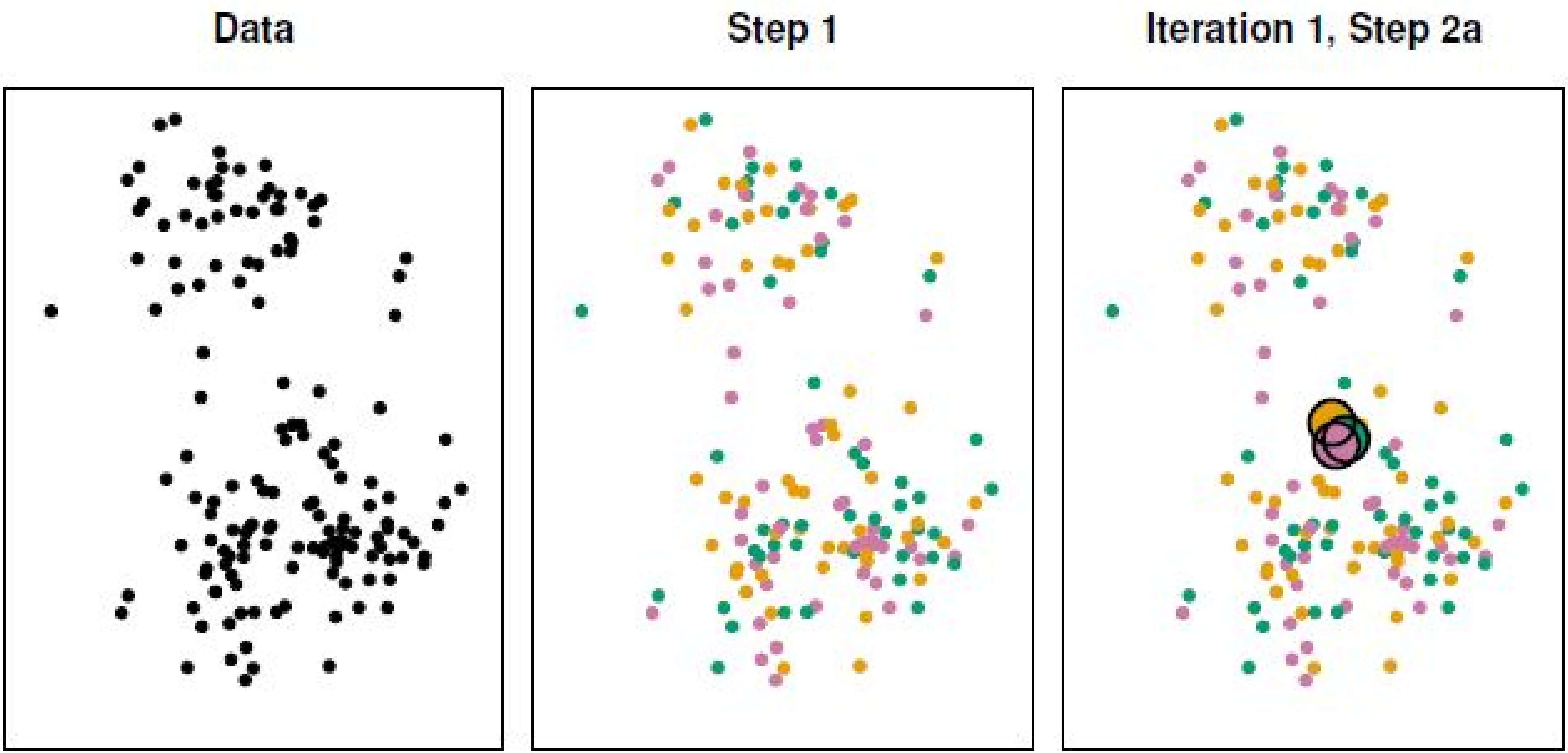
  $u_{i,k} = 0$ otherwise.

1. Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.
2. Iterate until the cluster assignments stop changing:
   2.1 For each of the K clusters, compute the cluster *centroid*. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster.
   2.2 Assign each observation to the cluster whose centroid is closest (where *closest* is defined using Euclidean distance).

Example

$$\vec{c}_k = \frac{1}{|C_k|} \sum_{i \in C_k} \vec{x}_i$$

Update matrix *U*

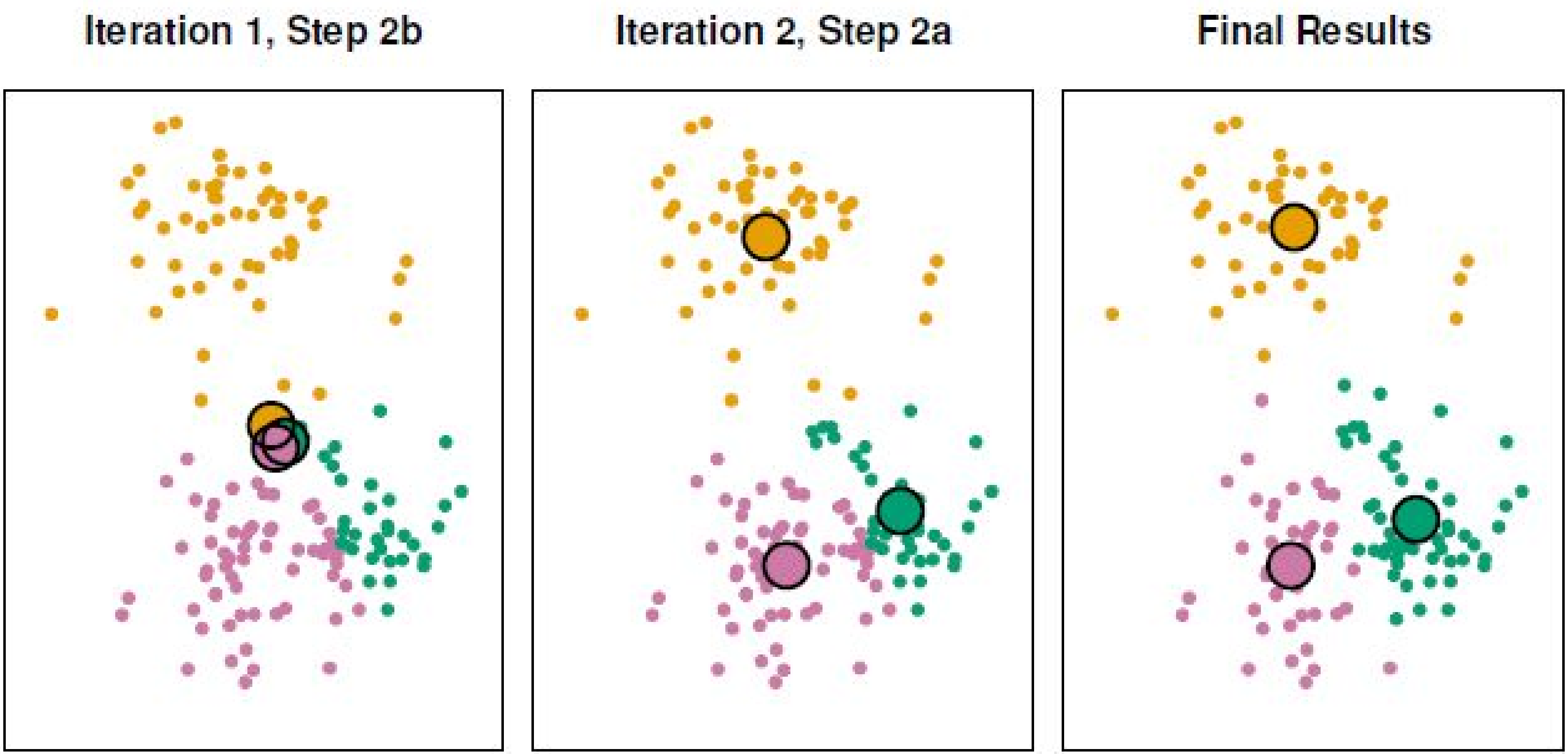


1. Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.
2. Iterate until the cluster assignments stop changing:
   2.1 For each of the K clusters, compute the cluster *centroid*. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster.
   2.2 Assign each observation to the cluster whose centroid is closest (where *closest* is defined using Euclidean distance).

Example

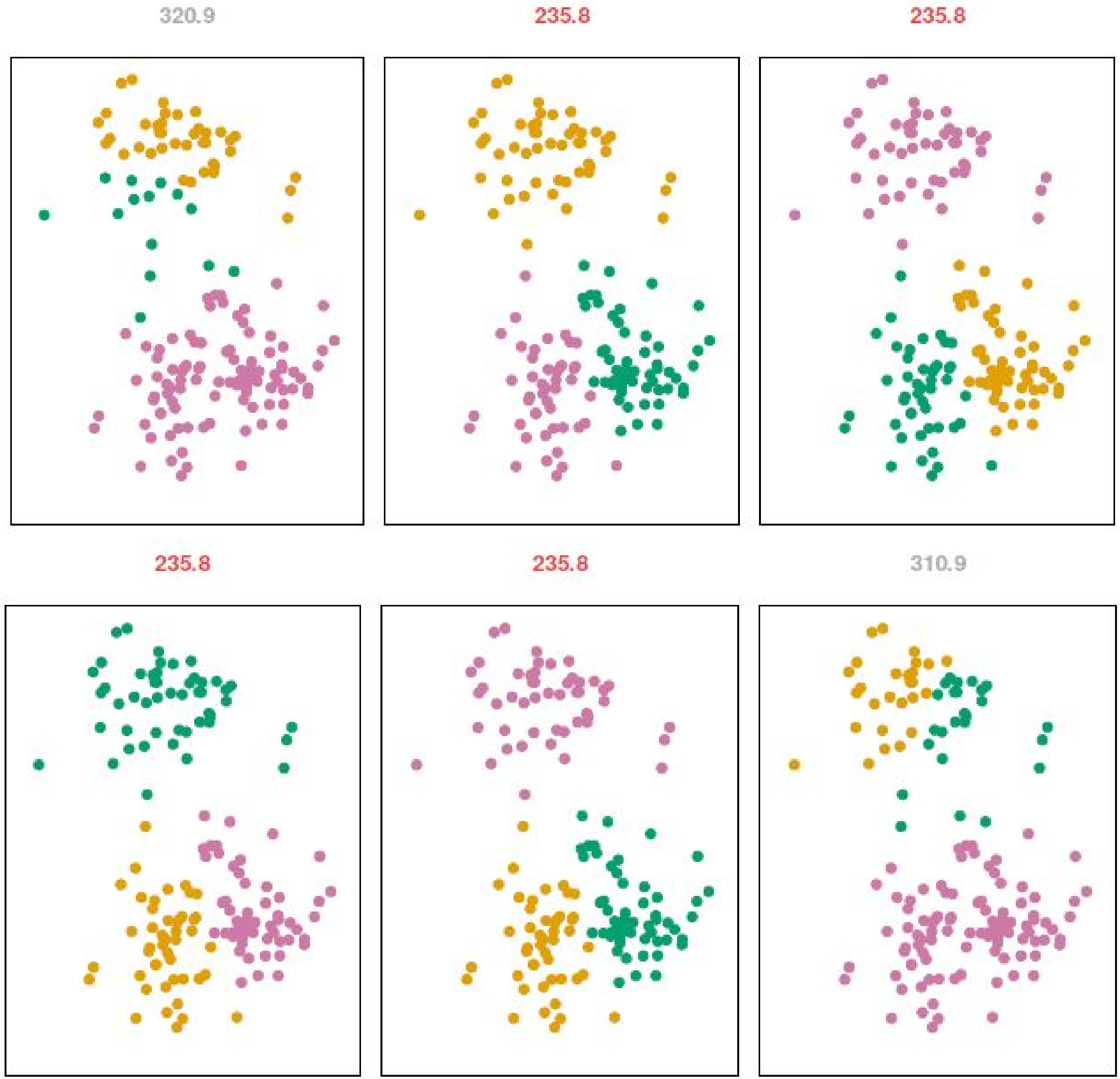
$$\vec{c}_k = \frac{1}{|C_k|} \sum_{i \in C_k} \vec{x}_i$$

Update matrix *U*



Different Starting Values

- The objective function (4) is non-convex so that the iterative K-Means algorithm can provide a local minimum rather than a global minimum.
- Therefore, the results obtained will depend on the initial (random) cluster assignment of each pattern in Step 1 of the algorithm.
- For this reason, it is important to run the algorithm multiple times from different random initial configurations - then one selects the *best* solution, i.e., that for which the objective (4) is smallest.


320.9
235.8
235.8
235.8
235.8
310.9

UNIVERSIDADE FEDERAL DE MINAS GERAIS

MATLAB Example

Fuzzy Clustering

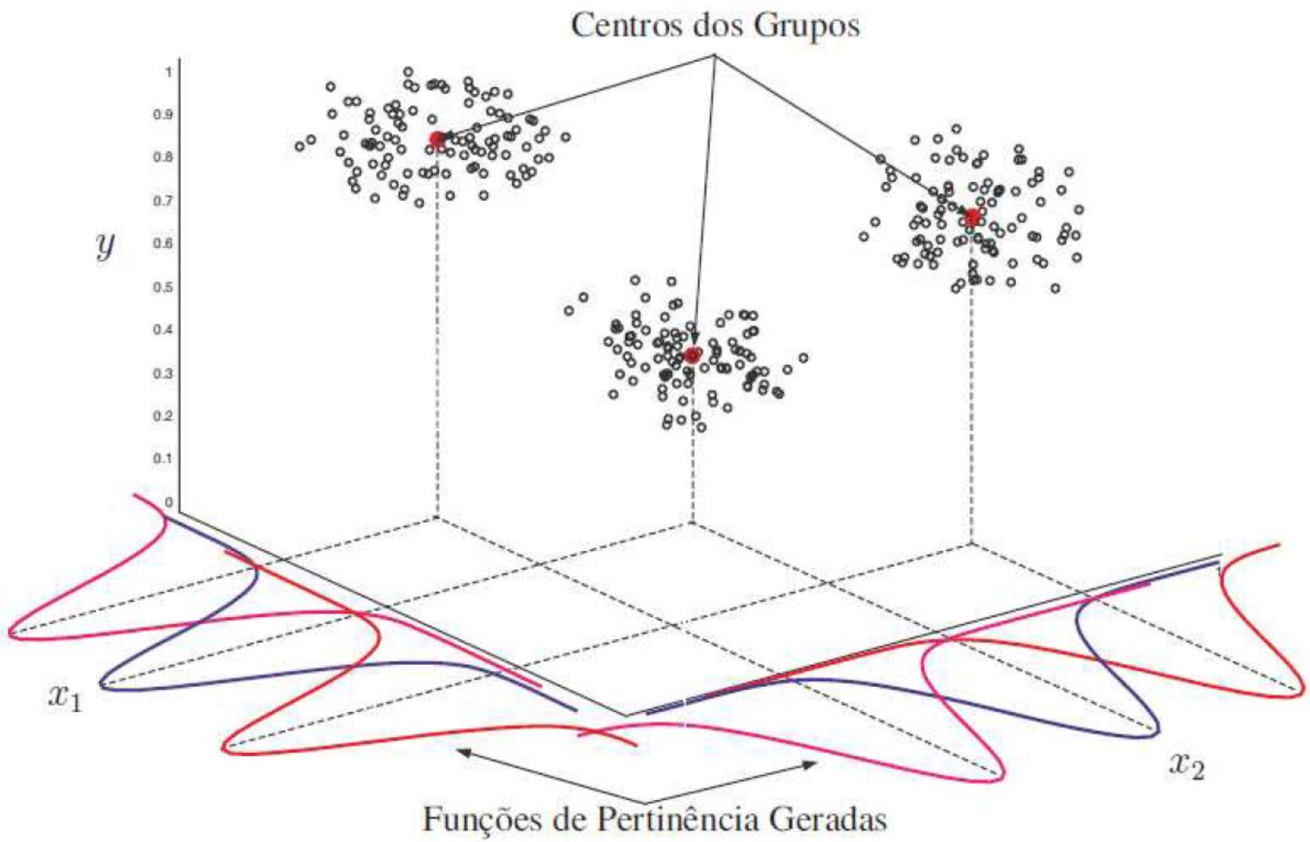
- Traditional Clustering (K-Means)
  - each observation belongs to a single cluster
  - the clusters are non-overlapping (disjunct sets)
- Fuzzy Clustering (Fuzzy K-Means)
  - extends the notion of assigning of each observation to clusters by using the concept of membership degree;
  - each observation belongs to a cluster according to a membership degree.

UFMG

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Motivation

- representation of imprecise (subjective) concepts;
- structure definition for a Fuzzy Model like ANFIS;



Fuzzy K-Means

- To accommodate fuzzy partitioning, the *n* x *K* membership matrix *U* is alllowed to have elements with the degree of belongingness specified by membership grades between 0 and 1

  $u_{i,k} =$ membership degree of the i-th observation $\vec{x}_i$ to the cluster C_k

- Normalization: the summation of membership degrees for a given observation $\vec{x}_i$ always be equal to unity:

$$\sum_{k=1}^{K} u_{i,k} = 1, \ \forall i = 1, \ldots, n$$

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Fuzzy K-Means Algorithm

1. Randomly assign membership degrees to each of the observations. This is equivalent to the Membership Matrix *U* initialization step.
2. Iterate until the cluster assignments stop changing:
   a. for each one of the K clusters, compute the cluster centroid:

$$\vec{c}_k = \frac{\sum_{i=1}^{n} u_{i,k}^{m} \vec{x}_i}{\sum_{i=1}^{n} u_{i,k}^{m}}$$

   b. update the Membership Matrix U by calculating the values according to

$$u_{i,k} = \frac{1}{\left(\frac{(\vec{x}_i - \vec{c}_k)^2}{\sum_{t=1}^{K} (\vec{x}_i - \vec{c}_t)^2}\right)^{2/(m-1)}}$$

m is often defined as 2;

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Fuzzy K-Means

- Centroid calculation:

$$c_k = \frac{\sum_{i=1}^{n} u_{i,k}^{m} \vec{x}_i}{\sum_{i=1}^{n} u_{i,k}^{m}}$$

- average of all observations weighted by their membership degree to the cluster C_k ;

m is often defined as 2;

Fuzzy K-Means

- Membership degree of the *i*-th observation to the cluster C_k:

$$u_{i,k} = \frac{1}{\left(\frac{(\vec{x}_i - \vec{c}_k)^2}{\sum_{t=1}^{K}(\vec{x}_i - \vec{c}_t)^2}\right)^{2/(m-1)}}$$

- the closer the observation $\vec{x}_i$ is to the centroid $\vec{c}_k$, the higher is its $u_{i,k}$.
- whether $\vec{x}_i$ coincides with $\vec{c}_k$, then

$$u_{i,k} = 1 \text{ for } C_k \text{ and } u_{i,t} = 0 \text{ for all } \; C_t \neq C_k$$

m is often defined as 2;