# 6 - Clustering problems

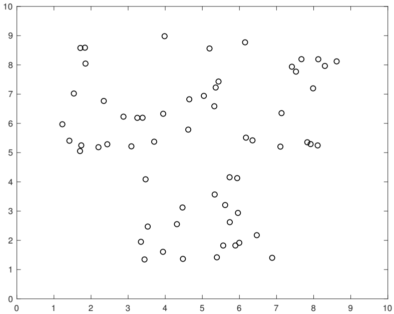
Mauro Passacantando

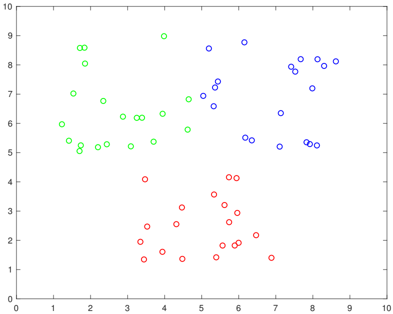
Department of Computer Science, University of Pisa
mauro.passacantando@unipi.it

Optimization Methods and Game Theory
Master of Science in Artificial Intelligence and Data Engineering
University of Pisa – A.Y. 2020/21

## Clustering problem – definition

Given a set $S$ of patterns and an integer number $k$, find a partition of $S$ in $k$ subsets $S_1, \dots, S_k$ (clusters) that are homogeneous and well separated.




Clustering problem is central in unsupervisioned machine learning.

## Clustering problem – optimization model

Assume that patterns are vectors $p_1, \dots, p_\ell \in \mathbb{R}^n$.

Consider a distance $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ between vectors in $\mathbb{R}^n$.
In the following $d(x, y) = \|x - y\|_2^2$ or $d(x, y) = \|x - y\|_1$.

For each cluster $S_j$ introduce a centroid $x_j \in \mathbb{R}^n$ (unknown).
Define clusters so that each pattern is associated to the closest centroid.

We aim to find $k$ centroids in order to minimize the sum of the distances between each pattern and the closest centroid:

$$\begin{cases} \min \displaystyle\sum_{i=1}^{\ell} \min_{j=1,\dots,k} d(p_i, x_j) \\ x_j \in \mathbb{R}^n \qquad \forall\, j = 1, \dots, k \end{cases}$$

## Clustering problem – optimization model with $\|\cdot\|_2$

Consider the distance $d(x, y) = \|x - y\|_2^2$.
The optimization problem to solve is

$$\begin{cases} \min \ \sum_{i=1}^{\ell} \min_{j=1,\dots,k} \|p_i - x_j\|_2^2 \\ x_j \in \mathbb{R}^n \qquad \forall\ j = 1, \dots, k \end{cases}$$

If $k = 1$ (one cluster), then it is a convex quadratic programming problem without constraints:

$$\begin{cases} \min \ \sum_{i=1}^{\ell} \|p_i - x\|_2^2 = \sum_{i=1}^{\ell} (x - p_i)^{\mathsf{T}}(x - p_i) \\ x \in \mathbb{R}^n \end{cases} \tag{1}$$

The global optimum is the stationary point:

$$2\ell x - 2\sum_{i=1}^{\ell} p_i = 0 \quad \Longleftrightarrow \quad x = \frac{\sum_{i=1}^{\ell} p_i}{\ell} \quad \text{(mean or baricenter)}$$

## Clustering problem – optimization model with $\|\cdot\|_2$

If $k > 1$ (at least two clusters), then the problem is nonconvex and nonsmooth:

$$\begin{cases} \min\limits_{x} \sum\limits_{i=1}^{\ell} \min\limits_{j=1,\dots,k} \|p_i - x_j\|_2^2 \\ x_j \in \mathbb{R}^n \qquad \forall\, j = 1,\dots,k \end{cases} \tag{2}$$

### Theorem

Problem (2) is equivalent to the following nonconvex smooth problem:

$$\begin{cases} \min\limits_{x,\alpha} \sum\limits_{i=1}^{\ell} \sum\limits_{j=1}^{k} \alpha_{ij} \|p_i - x_j\|_2^2 \\ \sum\limits_{j=1}^{k} \alpha_{ij} = 1 \quad \forall\, i = 1,\dots,\ell \\ \alpha_{ij} \geq 0 \qquad \forall\, i = 1,\dots,\ell,\ j = 1,\dots,k \\ x_j \in \mathbb{R}^n \qquad \forall\, j = 1,\dots,k. \end{cases} \tag{3}$$

**Proof.** Notice that $\min\limits_{j=1,\dots,k} \{a_j\} = \min \left\{ \sum\limits_{j=1}^{k} \alpha_j a_j :\ \sum\limits_{j=1}^{k} \alpha_j = 1,\ \alpha \geq 0 \right\}$.

## Clustering problem – $k$-means algorithm

The $k$-means algorithm is based on the following properties of problem (3):

- If $x_j$ are fixed, then (3) is decomposable into $\ell$ very simple LP problems: for any $i = 1, \dots, \ell$, the optimal solution is

$$\alpha^*_{ij} = \begin{cases} 1 & \text{if } j \text{ is the first index s.t. } \|p_i - x_j\|_2 = \min\limits_{h=1,\dots,k} \|p_i - x_h\|_2 \\ & \quad (x_j \text{ is the first closest centroid to } p_i), \\ 0 & \text{otherwise.} \end{cases}$$

- If $\alpha_{ij}$ are fixed, then (3) is decomposable into $k$ very simple convex QP problems similar to (1): for any $j = 1, \dots, k$, the optimal solution is

$$x^*_j = \frac{\sum\limits_{i=1}^{\ell} \alpha_{ij} p_i}{\sum\limits_{i=1}^{\ell} \alpha_{ij}} \qquad \text{(mean of patterns).}$$

## Clustering problem – $k$-means algorithm

The $k$-means algorithm consists in an alternating minimization of
$f(x,\alpha) = \sum_{i=1}^{\ell}\sum_{j=1}^{k} \alpha_{ij}\|p_i - x_j\|_2^2$ with respect to the two blocks of variables $x$ and $\alpha$.

**0.** (Inizialization) Set $t = 0$, choose centroids $x_1^0, \dots, x_k^0 \in \mathbb{R}^n$ and assign patterns to clusters: for any $i = 1, \dots, \ell$

$$\alpha_{ij}^0 = \begin{cases} 1 & \text{if } j \text{ is the first index s.t. } \|p_i - x_j^0\|_2 = \min\limits_{h=1,\dots,k} \|p_i - x_h^0\|_2 \\ 0 & \text{otherwise.} \end{cases}$$

**1.** (Update centroids) For each $j = 1, \dots, k$ compute the mean

$$x_j^{t+1} = \left(\sum_{i=1}^{\ell} \alpha_{ij}^t p_i\right) \Big/ \left(\sum_{i=1}^{\ell} \alpha_{ij}^t\right).$$

**2.** (Update clusters) For any $i = 1, \dots, \ell$ compute

$$\alpha_{ij}^{t+1} = \begin{cases} 1 & \text{if } j \text{ is the first index s.t. } \|p_i - x_j^{t+1}\|_2 = \min\limits_{h=1,\dots,k} \|p_i - x_h^{t+1}\|_2 \\ 0 & \text{otherwise.} \end{cases}$$

**3.** (Stopping criterion) If $f(x^{t+1}, \alpha^{t+1}) = f(x^t, \alpha^t)$ then STOP
else $t = t + 1$, go to Step 1.

## Clustering problem – $k$-means algorithm

**Theorem**

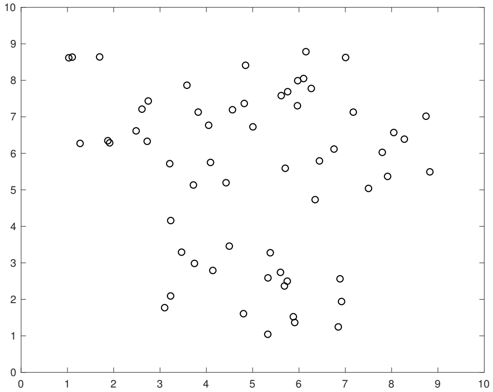
The $k$-means algorithm stops after a finite number of iterations at a solution $(x^*, \alpha^*)$ of the KKT system of problem (3) such that

$$
\begin{aligned}
& f(x^*, \alpha^*) \le f(x^*, \alpha), && \forall\ \alpha \ge 0 \ \text{ s.t. } \sum_{j=1}^{k} \alpha_{ij} = 1 \quad \forall\ i = 1, \dots, \ell, \\
& f(x^*, \alpha^*) \le f(x, \alpha^*), && \forall\ x \in \mathbb{R}^{kn}.
\end{aligned}
$$

**Remark.** The $k$-means algorithm does not guarantee to find a global optimum.

## Clustering problem – $k$-means algorithm

**Exercise 6.1.** Consider the $k$-means algorithm, with $k = 3$, for the set of patterns given in the file `6-1.txt`.




**a)** Run the algorithm starting from centroids $x_1 = (5, 7)$, $x_2 = (6, 3)$, $x_3 = (4, 3)$.

**b)** Run the algorithm starting from centroids $x_1 = (5, 7)$, $x_2 = (6, 3)$, $x_3 = (4, 4)$.

**c)** Is it possible to improve the solutions obtained in a) and b)?

## Clustering problem – optimization model with $\|\cdot\|_1$

Consider now the distance $d(x,y) = \|x-y\|_1$.
The optimization problem to solve is

$$\begin{cases} \min \displaystyle\sum_{i=1}^{\ell} \min_{j=1,\dots,k} \|p_i - x_j\|_1 \\ x_j \in \mathbb{R}^n \qquad \forall\, j = 1,\dots,k \end{cases}$$

If $k = 1$ (one cluster), then it is a convex problem decomposable into $n$ convex problems of one variable:

$$\begin{cases} \min \displaystyle\sum_{i=1}^{\ell} \|p_i - x\|_1 = \sum_{i=1}^{\ell}\sum_{h=1}^{n} |x_h - (p_i)_h| = \sum_{h=1}^{n} \underbrace{\sum_{i=1}^{\ell} |x_h - (p_i)_h|}_{f_h(x_h)} \\ x \in \mathbb{R}^n \end{cases} \tag{4}$$

## Clustering problem – optimization model with $\|\cdot\|_1$

Given $\ell$ real numbers $a_1 < a_2 < \cdots < a_\ell$, what is the optimal solution of

$$\begin{cases} \min \ \displaystyle\sum_{i=1}^{\ell} |x - a_i| = f(x) \\ x \in \mathbb{R} \end{cases} \qquad ?$$

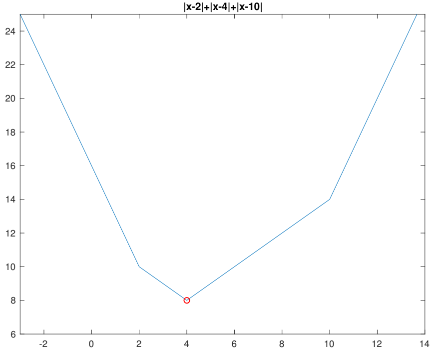
The objective function is convex and piecewise linear:

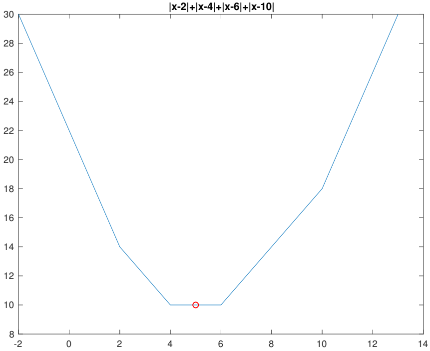
$$f(x) = \begin{cases} -\ell x + \sum_{i=1}^{\ell} a_i & \text{if } x < a_1 \\ (2-\ell)x + \sum_{i=2}^{\ell} a_i - a_1 & \text{if } x \in [a_1, a_2] \\ \dots & \dots \\ (2r-\ell)x + \sum_{i=r+1}^{\ell} a_i - \sum_{i=1}^{r} a_1 & \text{if } x \in [a_r, a_{r+1}] \\ \dots & \dots \\ (\ell-2)x + a_\ell - \sum_{i=1}^{\ell-1} a_i & \text{if } x \in [a_{\ell-1}, a_\ell] \\ \ell x - \sum_{i=1}^{\ell} a_i & \text{if } x > a_\ell \end{cases}$$

The global optimum is median$(a_1, \dots, a_\ell)$ $= \begin{cases} a_{(\ell+1)/2} & \text{if } \ell \text{ is odd,} \\ \dfrac{a_{\ell/2} + a_{1+\ell/2}}{2} & \text{if } \ell \text{ is even.} \end{cases}$

## Clustering problem – optimization model with $\|\cdot\|_1$

The global optimum is median$(a_1, \dots, a_\ell) = \begin{cases} a_{(\ell+1)/2} & \text{if } \ell \text{ is odd,} \\ \dfrac{a_{\ell/2} + a_{1+\ell/2}}{2} & \text{if } \ell \text{ is even.} \end{cases}$



## Clustering problem – optimization model with $\|\cdot\|_1$

If $k > 1$ (at least two clusters), then the problem is nonconvex and nonsmooth:

$$\begin{cases} \min\limits_{x} \ \sum\limits_{i=1}^{\ell} \ \min\limits_{j=1,\dots,k} \|p_i - x_j\|_1 \\ x_j \in \mathbb{R}^n \qquad \forall\ j = 1,\dots,k \end{cases} \tag{5}$$

**Theorem**

Problem (5) is equivalent to the following problem:

$$\begin{cases} \min\limits_{x,\alpha} \ \sum\limits_{i=1}^{\ell} \sum\limits_{j=1}^{k} \alpha_{ij} \|p_i - x_j\|_1 \\ \sum\limits_{j=1}^{k} \alpha_{ij} = 1 \quad \forall\ i = 1,\dots,\ell \\ \alpha_{ij} \geq 0 \qquad \forall\ i = 1,\dots,\ell,\ j = 1,\dots,k \\ x_j \in \mathbb{R}^n \qquad \forall\ j = 1,\dots,k. \end{cases} \tag{6}$$

## Clustering problem – optimization model with $\|\cdot\|_1$

**Theorem**

Problem (6) is, in turn, equivalent to the nonconvex bilinear problem:

$$
\begin{cases}
\displaystyle\min_{x,\alpha,u} \sum_{i=1}^{\ell} \sum_{j=1}^{k} \sum_{h=1}^{n} \alpha_{ij} u_{ijh} \\
u_{ijh} \geq (p_i)_h - (x_j)_h \qquad \forall\ i = 1,\dots,\ell,\ j = 1,\dots,k,\ \ h = 1,\dots,n \\
u_{ijh} \geq (x_j)_h - (p_i)_h \qquad \forall\ i = 1,\dots,\ell,\ j = 1,\dots,k,\ \ h = 1,\dots,n \\
\displaystyle\sum_{j=1}^{k} \alpha_{ij} = 1 \qquad \forall\ i = 1,\dots,\ell \\
\alpha_{ij} \geq 0 \qquad \forall\ i = 1,\dots,\ell,\ j = 1,\dots,k \\
x_j \in \mathbb{R}^n \qquad \forall\ j = 1,\dots,k.
\end{cases}
\tag{7}
$$

## Clustering problem – $k$-median algorithm

The $k$-median algorithm is based on the following properties of problem (6):

- If $x_j$ are fixed, then (6) is decomposable into $\ell$ very simple LP problems: for any $i = 1, \dots, \ell$, the optimal solution is

$$\alpha^*_{ij} = \begin{cases} 1 & \text{if } j \text{ is the first index s.t. } \|p_i - x_j\|_1 = \min\limits_{h=1,\dots,k} \|p_i - x_h\|_1 \\ & \quad (x_j \text{ is the first closest centroid to } p_i), \\ 0 & \text{otherwise.} \end{cases}$$

- If $\alpha_{ij} \in \{0,1\}$ are fixed, then (6) is decomposable into $k$ very simple convex problems similar to (4): for any $j = 1, \dots, k$, the optimal solution is

$$x^*_j = \text{median}(p_i : \ \alpha_{ij} = 1).$$

## Clustering problem – $k$-median algorithm

The $k$-median algorithm consists in an alternating minimization of

$f(x,\alpha) = \sum_{i=1}^{\ell} \sum_{j=1}^{k} \alpha_{ij} \|p_i - x_j\|_1$ with respect to the two blocks of variables $x$ and $\alpha$.

**0.** (Inizialization) Set $t = 0$, choose centroids $x_1^0, \dots, x_k^0 \in \mathbb{R}^n$ and assign patterns to clusters: for any $i = 1, \dots, \ell$

$$\alpha_{ij}^0 = \begin{cases} 1 & \text{if } j \text{ is the first index s.t. } \|p_i - x_j^0\|_1 = \min_{h=1,\dots,k} \|p_i - x_h^0\|_1 \\ 0 & \text{otherwise.} \end{cases}$$

**1.** (Update centroids) For each $j = 1, \dots, k$ compute

$$x_j^{t+1} = \text{median}(p_i : \ \alpha_{ij}^t = 1).$$

**2.** (Update clusters) For any $i = 1, \dots, \ell$ compute

$$\alpha_{ij}^{t+1} = \begin{cases} 1 & \text{if } j \text{ is the first index s.t. } \|p_i - x_j^{t+1}\|_1 = \min_{h=1,\dots,k} \|p_i - x_h^{t+1}\|_1 \\ 0 & \text{otherwise.} \end{cases}$$

**3.** (Stopping criterion) If $f(x^{t+1}, \alpha^{t+1}) = f(x^t, \alpha^t)$ then STOP

else $t = t + 1$, go to Step 1.

## Clustering problem – $k$-median algorithm

**Theorem**

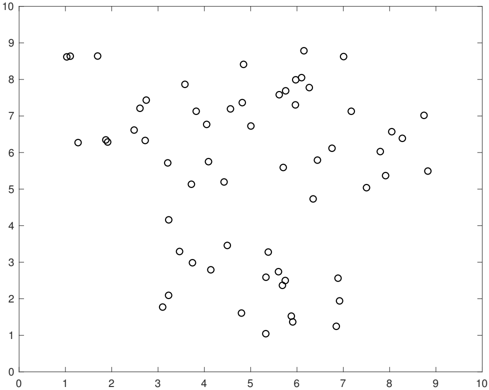
The $k$-median algorithm stops after a finite number of iterations at a stationary point $(x^*, \alpha^*)$ of problem (6) such that

$$
\begin{aligned}
f(x^*, \alpha^*) &\leq f(x^*, \alpha), && \forall\ \alpha \geq 0 \ \text{ s.t. } \sum_{j=1}^{k} \alpha_{ij} = 1 \quad \forall\ i = 1, \dots, \ell, \\
f(x^*, \alpha^*) &\leq f(x, \alpha^*), && \forall\ x \in \mathbb{R}^{kn}.
\end{aligned}
$$

**Remark.** The $k$-means algorithm does not guarantee to find a global optimum.

## Clustering problem – $k$-median algorithm

**Exercise 6.2.** Consider the $k$-median algorithm, with $k = 3$, for the set of patterns given in the file `6-1.txt`.




**a)** Run the algorithm starting from centroids $x_1 = (5, 7)$, $x_2 = (6, 3)$, $x_3 = (4, 3)$.
**b)** Run the algorithm starting from centroids $x_1 = (5, 7)$, $x_2 = (6, 3)$, $x_3 = (4, 4)$.
**c)** Is it possible to improve the solutions obtained in a) and b)?