

Apprentissage Non-Supervisé

L3 FLEX, Machine Learning, Université Paul Sabatier

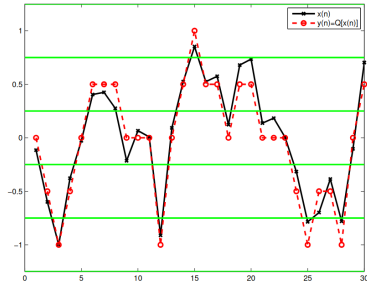
Romain Thoreau
`romain.thoreau@cnes.fr`

Mars 2025

Quantification : algorithme de Lloyd-Max

Clustering : algorithme des K-moyennes

Quantification



(a) Quantification d'un signal.



(b) Quantification d'une image.

Figure: Exemples de quantification. Source :

https://cagnazzo.wp.imt.fr/files/2013/05/compression_01_quantification.pdf

Quantifieur

On considère une variable aléatoire X , dont les réalisations sont dans \mathbb{R}^n . On note f_X la densité de probabilité de X qui décrit la vraisemblance qu'une réalisation de X se trouve dans une certaine région de \mathbb{R}^n .

Un quantifieur est une fonction

$$q : \mathbb{R}^n \longrightarrow \mathcal{C}$$

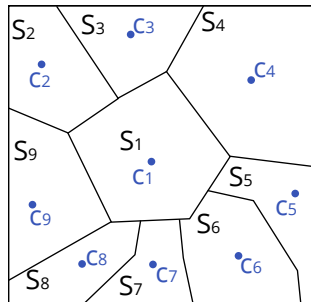
où $\mathcal{C} = \{c_i\}_{i=1}^K$ désigne un *dictionnaire* ou un *alphabet* qui contient K vecteurs de dimension n .

Le quantifieur est défini par

$$q(x) = c_i \text{ si } x \in S_i$$

où $\{S_i\}_{i=1}^K$ est une partition de \mathbb{R}^n :

$$\bigcup_{i=1}^K S_i = \mathbb{R}^n, \quad S_i \cap S_j = \emptyset \text{ pour tout } i \neq j.$$



Un quantifieur compresse l'information

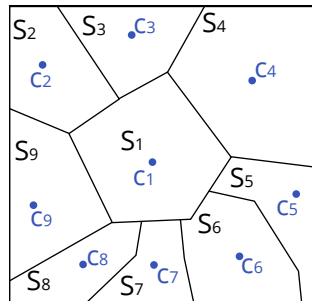
Soit un quantifieur $q : \mathbb{R}^n \longrightarrow \mathcal{C}$ associé à la partition $\{S_1, \dots, S_K\}$ et un dictionnaire $\{c_1, \dots, c_K\}$.

La variable X est *résumée* aux centres du dictionnaire.

Ainsi, le quantifieur résulte en une distorsion de \mathbb{R}^n , notée $D(q)$:

$$D(q) = \int_{\mathbb{R}^n} d(x, q(x)) f_X(x) dx$$

où d désigne une distance dans \mathbb{R}^n , et f_X désigne la densité de probabilité de X .



Minimisation de la distorsion

Conditions nécessaires pour minimiser la distorsion

- 1. Étant donné un dictionnaire $\{c_1, \dots, c_K\}$, la partition $\{S_1, \dots, S_K\}$ doit satisfaire :

$$\forall i, x \in S_i, j \neq i, d(x, c_i) \leq d(x, c_j)$$

pour minimiser la distorsion. Une partition qui respecte cette condition, appelée la **condition du plus proche voisin**, s'appelle une **partition de Voronoï**.

- 2. Étant donnée une partition $\{S_1, \dots, S_K\}$, un dictionnaire $\{c_1, \dots, c_K\}$ doit satisfaire :

$$\forall i, \int_{S_i} d(x, c_i) f_X(x) dx = \min_u \int_{S_i} d(x, u) f_X(x) dx$$

pour minimiser la distorsion. Pour la distance euclidienne, on a :

$$c_i = \arg \min_u \int_{S_i} d(x, u) f_X(x) dx = \frac{\int_{S_i} x f_X(x) dx}{\int_{S_i} f_X(x) dx}$$

Algorithme de Lloyd-Max

Algorithme de Lloyd-Max

- Initialiser aléatoirement le dictionnaire,
- Mettre à jour la partition pour satisfaire la condition 1,
- Mettre à jour le dictionnaire pour satisfaire la condition 2,
- Itérer jusqu'à la convergence.

Algorithme de Lloyd-Max

Algorithme de Lloyd-Max pour une séquence d'observations = algorithme des K-moyennes

On dispose de données $\{x_1, \dots, x_M\}$ à quantifier.

- Initialiser aléatoirement le dictionnaire,
- Mettre à jour la partition pour satisfaire la condition 1 :
assigner chaque point x_i à une région S_j tel que $j = \arg \min_k d(x_i, c_k)$.
- Mettre à jour le dictionnaire pour satisfaire la condition 2 :
$$c_j \leftarrow \frac{1}{\sum_{i=1}^M \gamma_{ij}} \sum_{i=1}^M \gamma_{ij} x_i \text{ où } \gamma_{ij} = \begin{cases} 1 & \text{si } x_i \in S_j \\ 0 & \text{sinon} \end{cases}$$
- Itérer jusqu'à la convergence, e.g.,

$$\frac{D^t - D^{t+1}}{D^t} \leq \epsilon$$

où t désigne l'itération de l'algorithme.

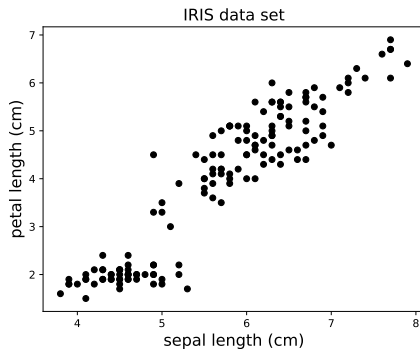
Le quantifieur obtenu est appelé un **quantifieur des plus proches voisins** d'ordre K.

Quantification : algorithme de Lloyd-Max

Clustering : algorithme des K-moyennes

Algorithme de clustering

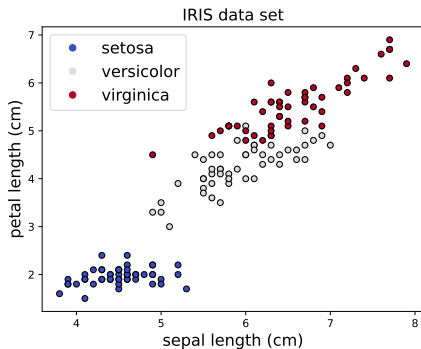
- ▶ Le principe des algorithmes de clustering est de grouper les données en plusieurs catégories.
- ▶ On ne dispose d'aucunes annotations mais on suppose que des classes / catégories / groupes structurent la distribution des données.



- ▶ En combien de catégories regrouperiez-vous ces données ?

Algorithme de clustering

- ▶ Le principe des algorithmes de clustering est de grouper les données en plusieurs catégories.
- ▶ On ne dispose d'aucunes annotations mais on suppose que des classes / catégories / groupes structurent la distribution des données.



- ▶ En combien de catégories regrouperiez-vous ces données ?

Applications d'algorithmes de clustering en imagerie

Original image (96,615 colors)



Quantized image (64 colors, K-Means)



Figure: Postérisation d'images avec l'algorithme des k-moyennes. Source : Sklearn

Applications d'algorithmes de clustering en imagerie

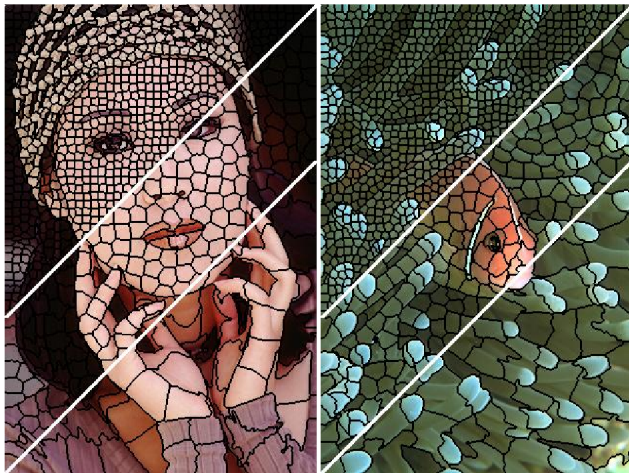


Figure: Images segmentées en super-pixels avec SLIC [Achanta et al., 2012]. Source : [Achanta et al., 2012]

Applications d'algorithmes clustering dans des domaines variés

- ▶ An evaluation of **document clustering** and topic modelling in two [online social networks: Twitter and Reddit](#) [Curiskis et al., 2020]
- ▶ Dynamic **clustering** of residential [electricity consumption time series](#) data based on Hausdorff distance [Benitez et al., 2016]
- ▶ Utility of the k-means **clustering** algorithm in differentiating apparent diffusion coefficient values of [benign and malignant neck pathologies](#) [Srinivasan et al., 2010]

Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

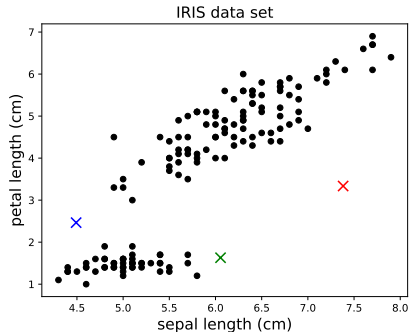
Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$



Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

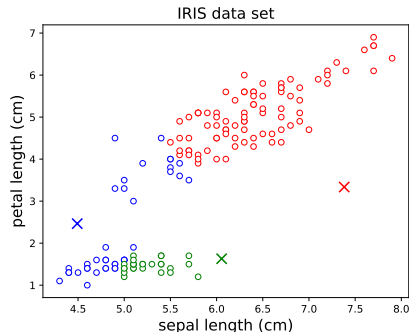
$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

fin tant que



où $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ est une distance (e.g. distance euclidienne)

Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

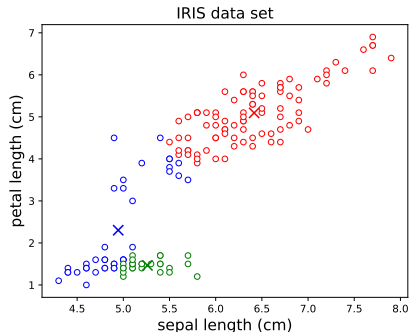
▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

où $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ est une distance (e.g. distance euclidienne) et $\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$ est appelée

la responsabilité de la classe k pour la donnée i .



Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

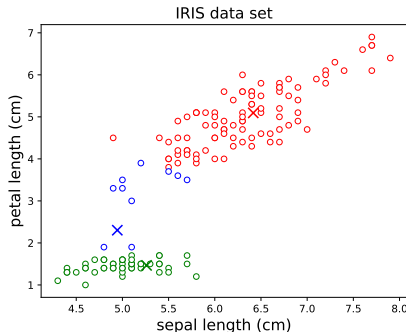
▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

où $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ est une distance (e.g. distance euclidienne) et $\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$ est appelée

la responsabilité de la classe k pour la donnée i .



Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

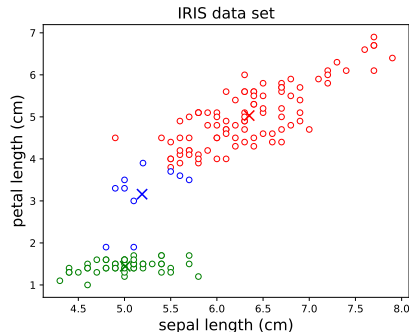
▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que



où $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ est une distance (e.g. distance euclidienne) et $\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$ est appelée

la responsabilité de la classe k pour la donnée i .

Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

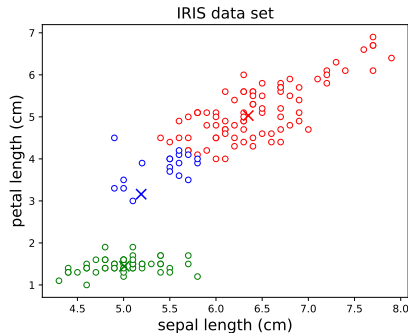
▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

où $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ est une distance (e.g. distance euclidienne) et $\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$ est appelée

la responsabilité de la classe k pour la donnée i .



Algorithme des K-moyennes

On considère N données non annotées $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}, \mathbf{x}^{(i)} \in \mathcal{X}$.

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

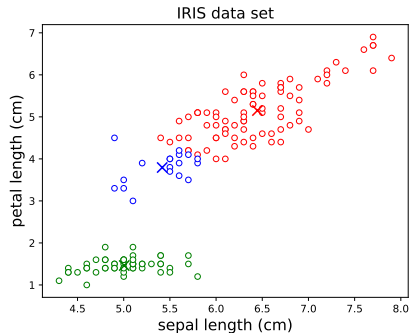
▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

où $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ est une distance (e.g. distance euclidienne) et $\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$ est appelée

la responsabilité de la classe k pour la donnée i .



Algorithme des K-moyennes

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

► L'algorithme des K-moyennes est un algorithme d'optimisation !

$$\min_{\mathbf{c}, y^{(1)}, \dots, y^{(N)}} \sum_{i=1}^N \sum_{k=1}^K \gamma_k^{(i)} d(\mathbf{x}^{(i)}, \mathbf{c}_k)$$

$$s.c. \quad \forall i \quad \sum_{k=1}^K \gamma_k^{(i)} = 1$$

$$\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$$

Algorithme des K-moyennes

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

► L'algorithme des K-moyennes est un algorithme d'optimisation !

$$\min_{\mathbf{c}, y^{(1)}, \dots, y^{(N)}} \sum_{i=1}^N \sum_{k=1}^K \gamma_k^{(i)} d(\mathbf{x}^{(i)}, \mathbf{c}_k)$$

$$s.c. \quad \forall i \quad \sum_{k=1}^K \gamma_k^{(i)} = 1$$

$$\gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases}$$

Algorithme des K-moyennes

Algorithme des K-moyennes

▷ On fixe le nombre de classes

$K \leftarrow 3$

▷ Initialisation aléatoires des centres des classes

$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \leftarrow \text{Initialisation}(K)$

tant que les centres \mathbf{c}_i "changent" **faire**

▷ Attribution des classes

$y^{(i)} \leftarrow \arg \min_k d(\mathbf{x}^{(i)}, \mathbf{c}_k)$

▷ Mise à jour des centres

$\mathbf{c}_k \leftarrow \frac{1}{\sum_{i=1}^N \gamma_k^{(i)}} \sum_{i=1}^N \gamma_k^{(i)} \mathbf{x}^{(i)}$

fin tant que

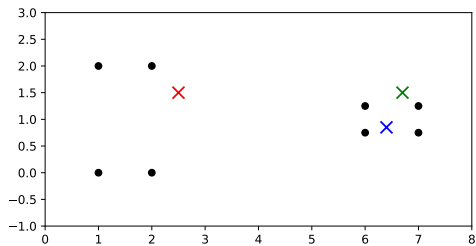
► Chaque étape, 1) l'attribution des classes et 2) la mise à jour des centres, diminue la fonction objective, pour 1) les centres et 2) les classes fixés.

► Mais la fonction objective n'est pas convexe, le résultat dépend de l'initialisation des centres. On converge probablement vers un minimum local.

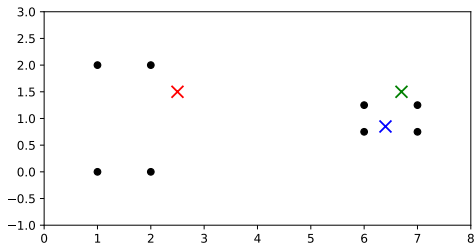
► L'algorithme des K-moyennes est un algorithme d'optimisation !

$$\begin{aligned} \min_{\mathbf{c}, y^{(1)}, \dots, y^{(N)}} \quad & \sum_{i=1}^N \sum_{k=1}^K \gamma_k^{(i)} d(\mathbf{x}^{(i)}, \mathbf{c}_k) \\ \text{s.c. } \forall i \quad & \sum_{k=1}^K \gamma_k^{(i)} = 1 \\ & \gamma_k^{(i)} = \begin{cases} 1 & \text{si } k = y^{(i)} \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

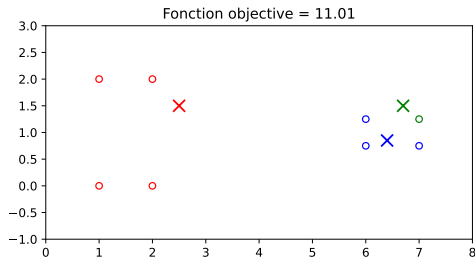
Robustesse à l'initialisation des centres

(a) $t = 0$

Robustesse à l'initialisation des centres

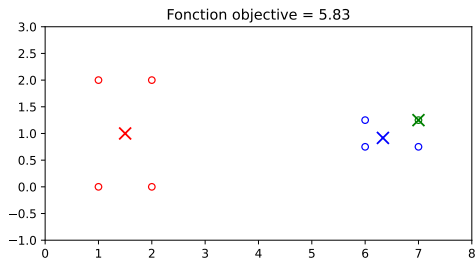


(a) $t = 0$

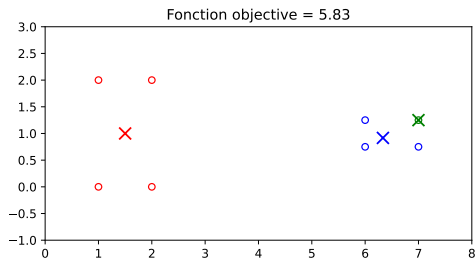
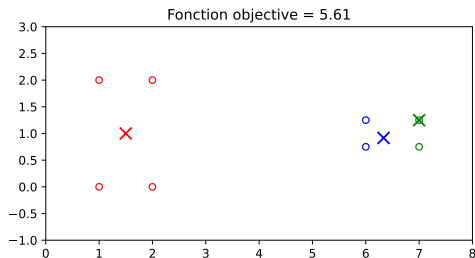


(b) $t = 1$

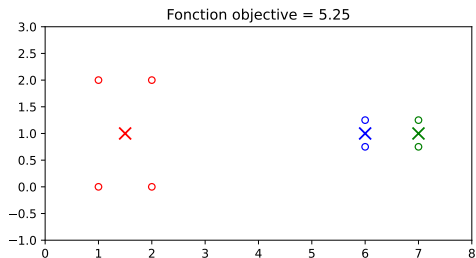
Robustesse à l'initialisation des centres

(a) $t = 2$

Robustesse à l'initialisation des centres

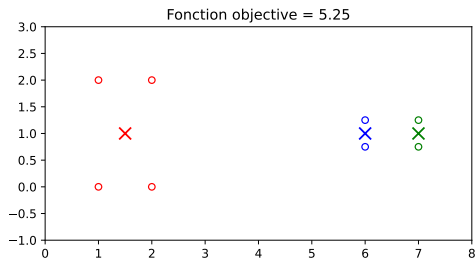
(a) $t = 2$ (b) $t = 3$

Robustesse à l'initialisation des centres

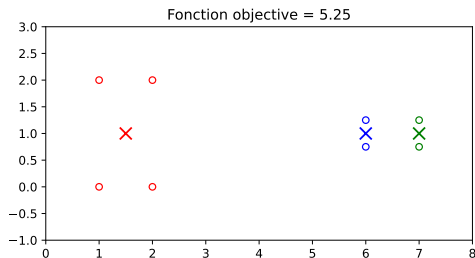


(a) $t = 4$

Robustesse à l'initialisation des centres

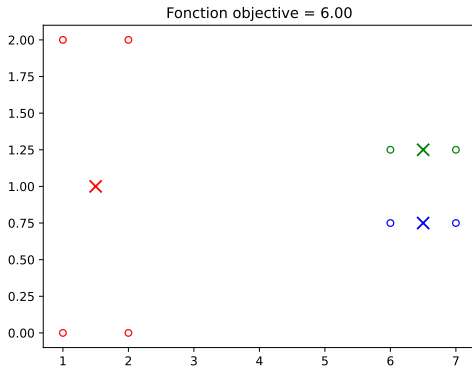


(a) $t = 4$

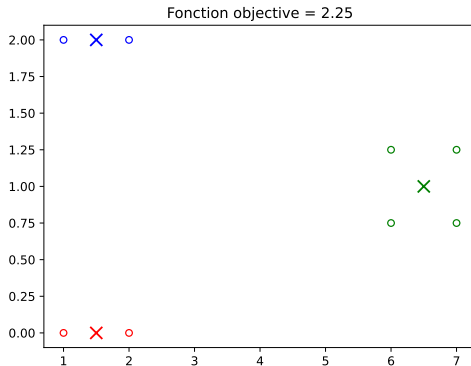


(b) $t = 5$

Robustesse à l'initialisation des centres




(a) "Mauvais" optimum local



(b) Optimum global

Inconvénients de la méthode des K-moyennes

- ▶ Attribution des classes sans prendre en compte l'incertitude pour des données proches de deux clusters,
- ▶ A l'inférence, peu robuste à des données n'appartenant à aucun cluster,
- ▶ Isotropie de la distance : ne prend pas en compte les différences de variance suivant les différentes dimensions,

- 
- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012).
Slic superpixels compared to state-of-the-art superpixel methods.
IEEE transactions on pattern analysis and machine intelligence, 34(11):2274–2282.
- 
- Benitez, I., Díez, J.-L., Quijano, A., and Delgado, I. (2016).
Dynamic clustering of residential electricity consumption time series data based on hausdorff distance.
Electric Power Systems Research, 140:517–526.
- 
- Curiskis, S. A., Drake, B., Osborn, T. R., and Kennedy, P. J. (2020).
An evaluation of document clustering and topic modelling in two online social networks: Twitter and reddit.
Information Processing & Management, 57(2):102034.
- 
- Deselaers, T., Heigold, G., and Ney, H. (2010).
Object classification by fusing svms and gaussian mixtures.
Pattern Recognition, 43(7):2476–2484.
- 
- Liang, C., Wang, W., Miao, J., and Yang, Y. (2022).
Gmmseg: Gaussian mixture based generative semantic segmentation models.
Advances in Neural Information Processing Systems, 35:31360–31375.
- 
- Liu, T.-C., Kalugin, P. N., Wilding, J. L., and Bodmer, W. F. (2022).
Gmmchi: gene expression clustering using gaussian mixture modeling.
BMC bioinformatics, 23(1):457.
- 
- Srinivasan, A., Galban, C., Johnson, T., Chenevert, T., Ross, B., and Mukherji, S. (2010).
Utility of the k-means clustering algorithm in differentiating apparent diffusion coefficient values of benign and malignant neck pathologies.
American journal of neuroradiology, 31(4):736–740.



Zissman, M. A. (1993).

Automatic language identification using gaussian mixture and hidden markov models.

In *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 399–402. IEEE.