Faculté d'informatique
Departement IA & SD

# DATA MINING

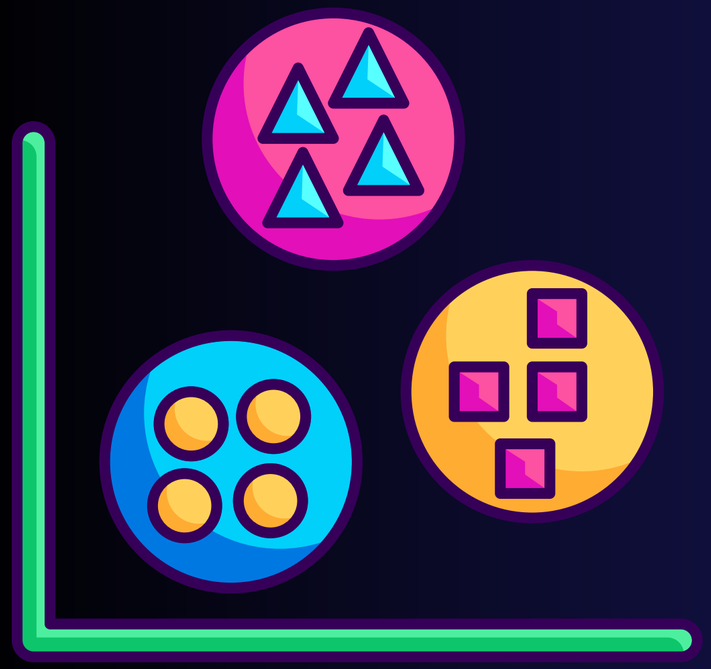
## TP2: Algorithme K-Means

BENADJAL.N
Phd student in computer science
LAB: MOVEP
Email: nesrinebenadjel63@gmail.com

INGSOFT4

2025/2026

# Introduction



Dans le TP précédent, nous allons découvrir les étapes à suivre dans la phase de prétraitement pour assurer que notre ensemble de données est prêt pour l'entraînement du modèle.

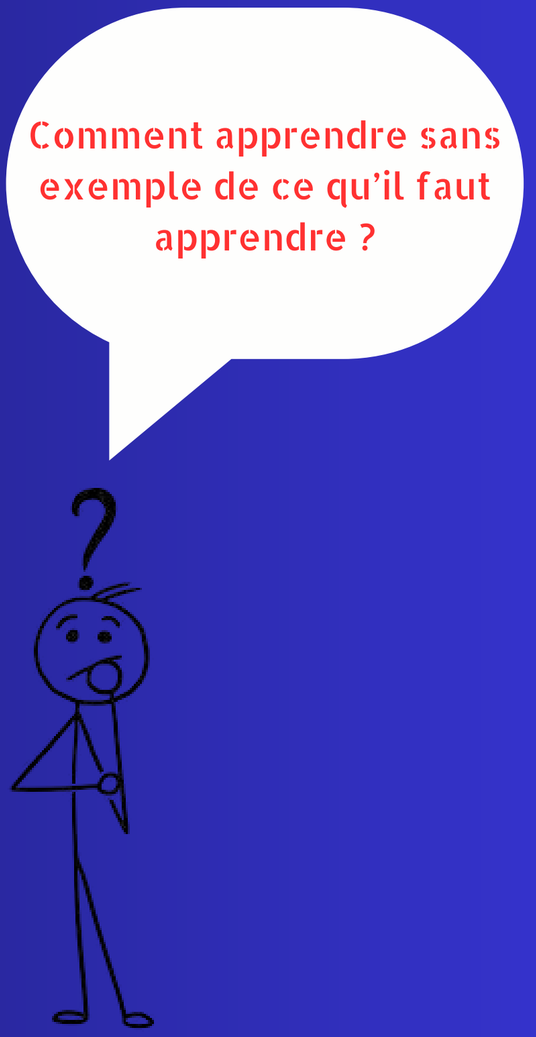
Dans ce TP, inshallah, nous allons nous concentrer sur la deuxième étape de l'apprentissage automatique la construction d'un modèle.

Nous allons commencer avec **l'apprentissage non supervisé** et exactement l'algorithme de **K-means.**

# Le problème de l'Apprentissage automatique

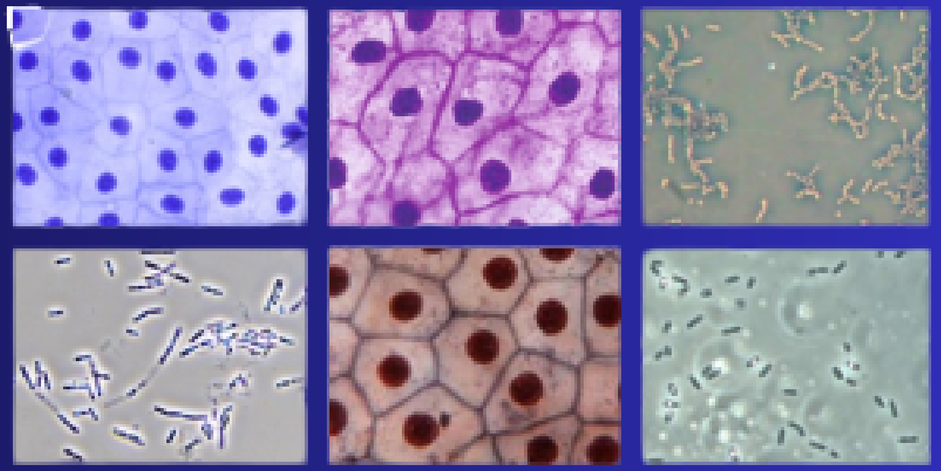
D'un certain point de vue, l'apprentissage automatique consiste à enseigner à la machine des choses que nous connaissons déjà, étant donné que nous construisons à l'avance un Dataset qui contient des questions $X$ et des réponses $y$.

- Que faire alors si vous disposez d'un Dataset sans valeur $y$ ?



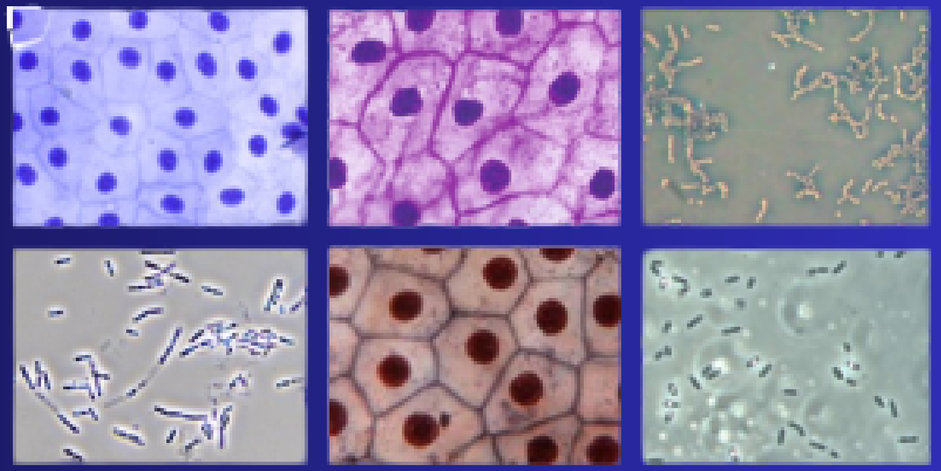
# Le problème de l'Apprentissage automatique

Regardez ces 6 photos. Pouvez-vous les regrouper en 2 familles selon leur ressemblance ?

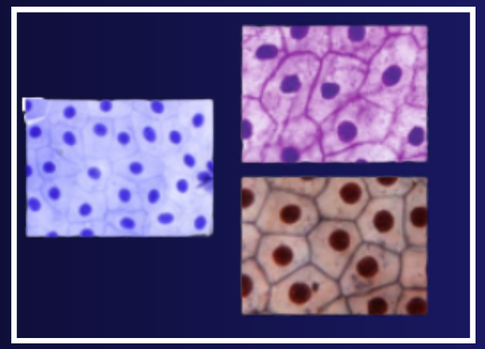


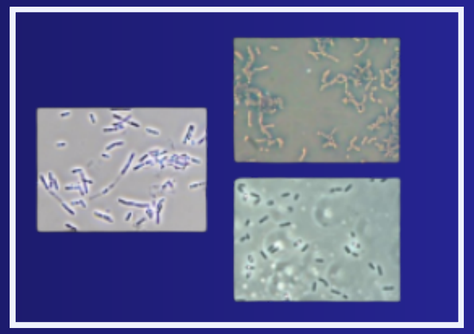
# Le problème de l'Apprentissage automatique

Regardez ces 6 photos. Pouvez-vous les regrouper en 2 familles selon leur ressemblance ?



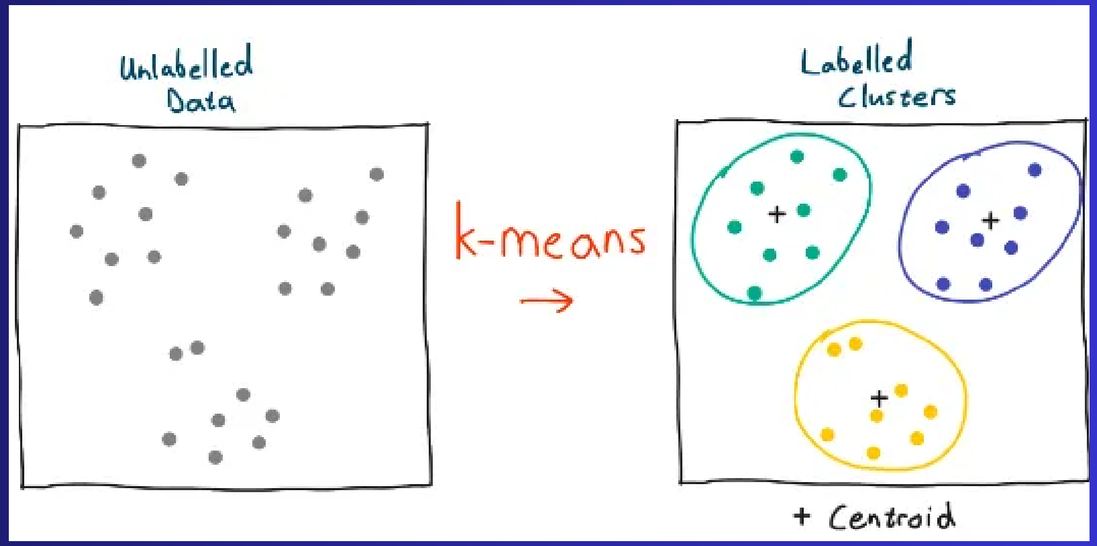
Bien sûr ! C'est même plutôt simple. Nul besoin de savoir s'il s'agit de cellules animales, de bactéries ou de protéines pour apprendre à classer ces images. Votre cerveau a en fait reconnu des structures communes dans les données que vous lui avez montrées.





# Algorithme de K-Mean Clustering

Le K-Mean est un algorithme de l'apprentissage automatique non supervisé, qui permet de regrouper les points de données en clusters en fonction de leur similarité intrinsèque.




L'algorithme fonctionne en 2 étapes répétées en boucle. On commence par détermination du nombre de clusters (k) en suite à rallier chaque instance au centre le plus proche. Après cette étape, nous avons K clusters l'étape 2 consiste à déplacer les centres au milieu de leur Cluster.

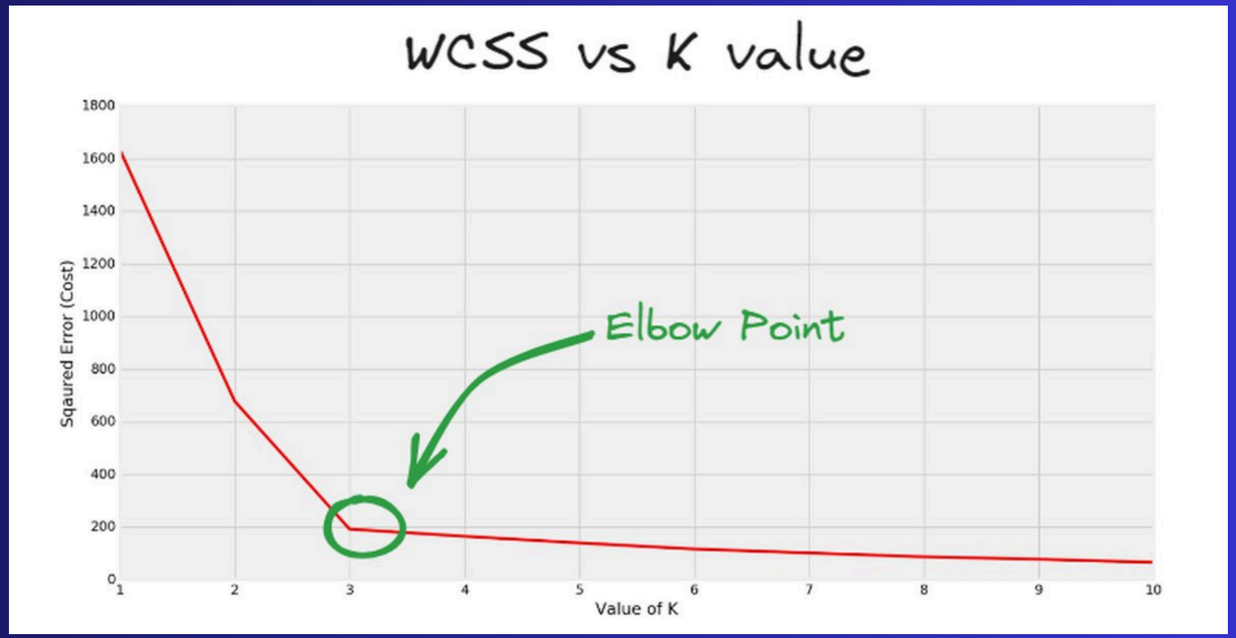
On répète ainsi les étapes 1 et 2 en boucle jusqu'à ce que les centres **ne bougent plus**.

# Détermination du nombre de clusters (k)

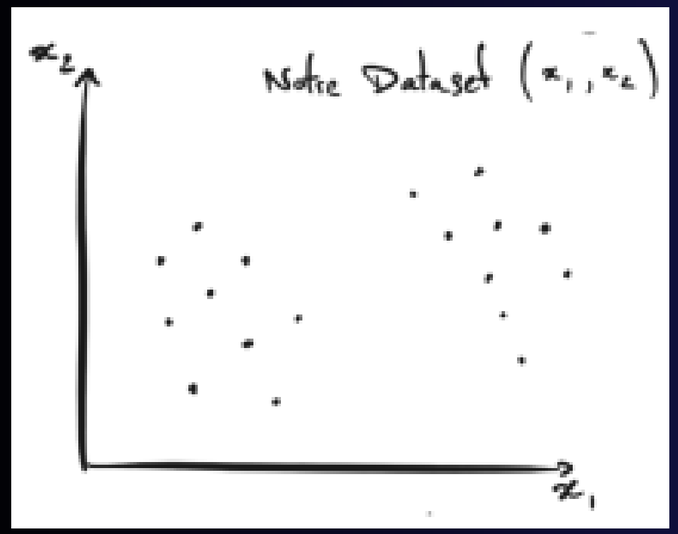
**Courbe d'Elbow :**

La méthode du coude permet de tracer la somme des carrés intra-cluster (WCSS) en fonction de l'augmentation des valeurs de k et de rechercher un point où l'amélioration ralentit  ce point est appelé « coude ».

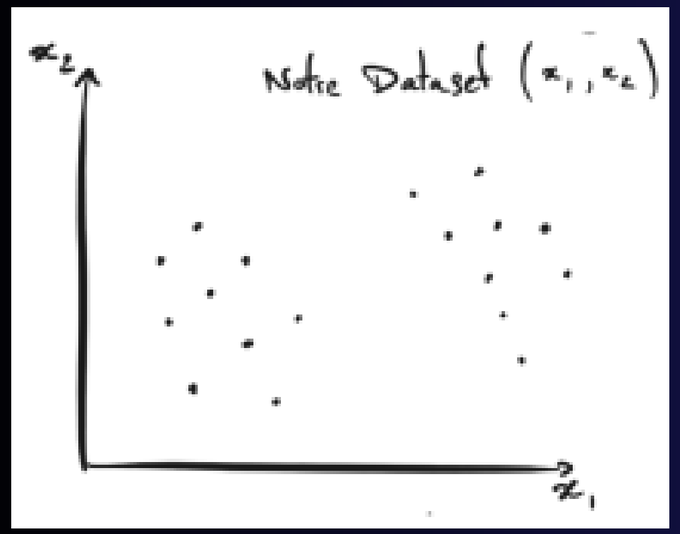
$$\text{WCSS} = \sum_{i=1}^{K} \sum_{j=1}^{|P_i|} \text{distance}(P_{ij}, C_i)^2$$



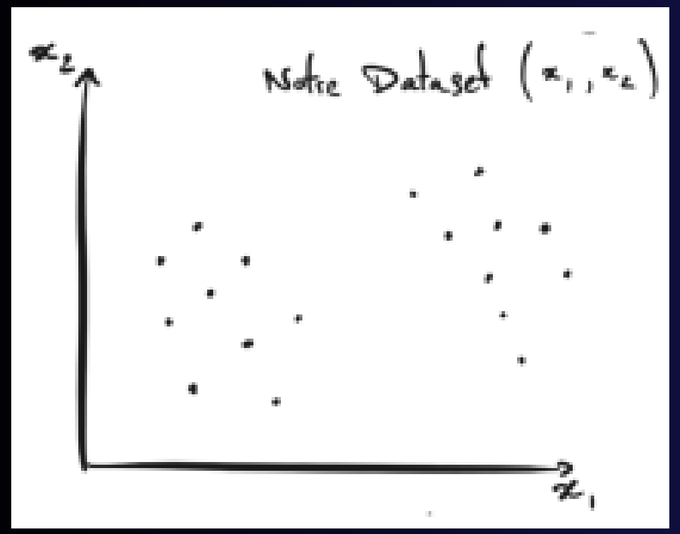
# Déroulement de l'algorithme K-means



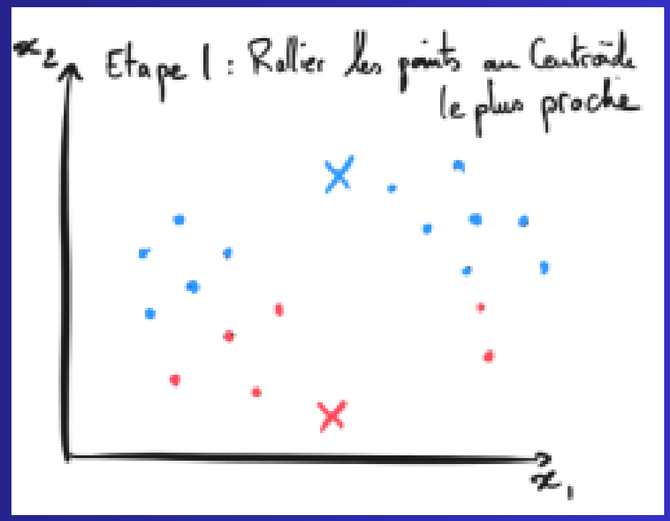
# Déroulement de l'algorithme K-means



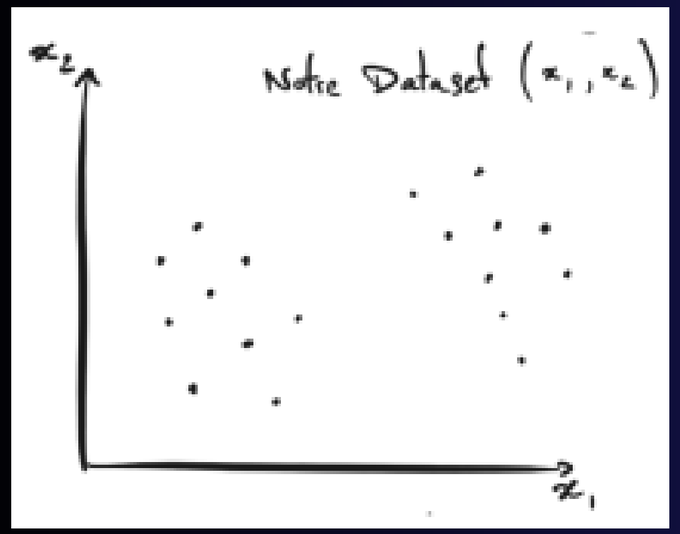

# Déroulement de l'algorithme K-means



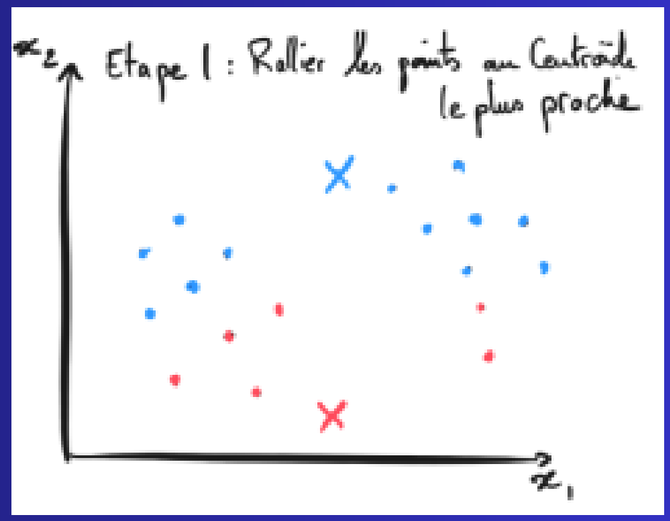


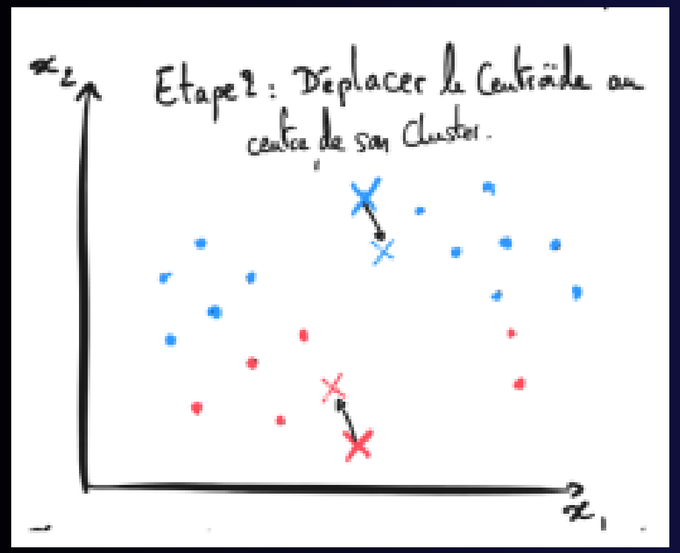


# Déroulement de l'algorithme K-means

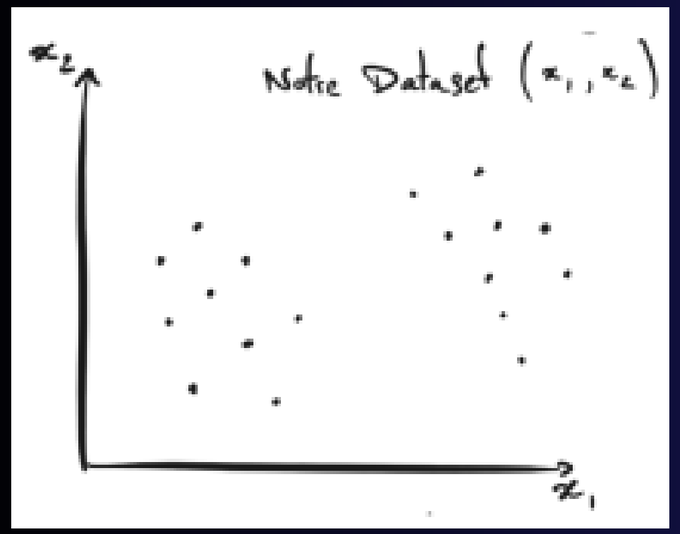




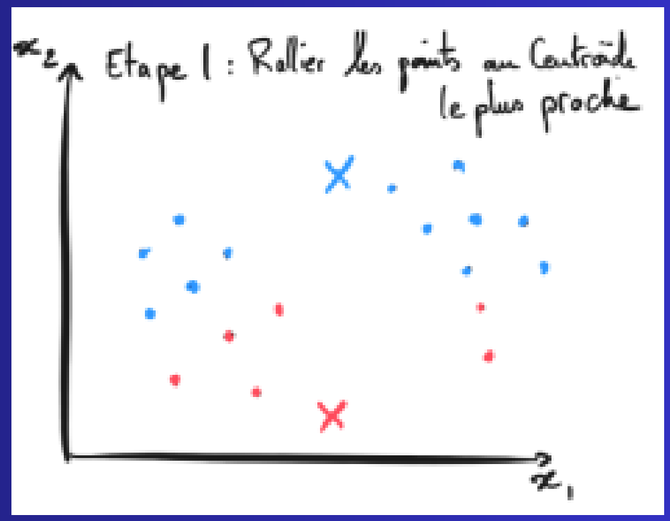


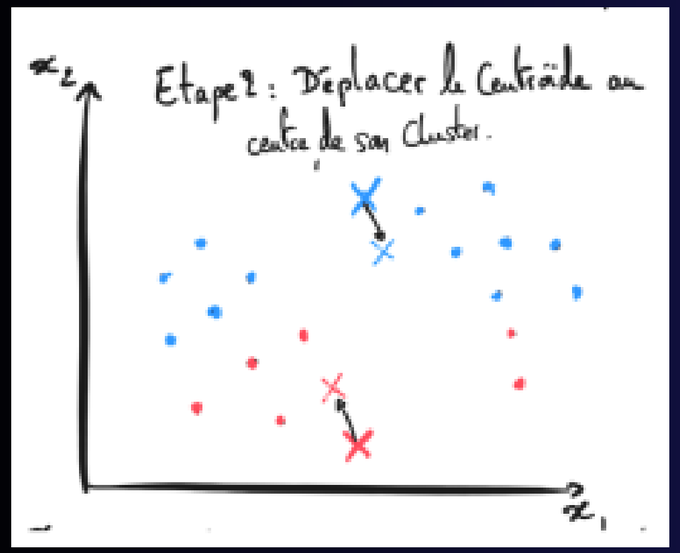



# Déroulement de l'algorithme K-means








# Déroulement de l'algorithme K-means