1. **Simulation d’un modèle de mélange gaussien diagonal**

n = 1000

proportions = [0.4, 0.6]

means = [[0, 2], [2, 5]]

sds = [[1, 5], [1, 2]]

n = 1000

proportions = [0.5, 0.5]

means = [[0, 0], [5, 5]]

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquementstds = [[1, 3], [3, 1]]

Une image contenant capture d’écran, Caractère coloré, Tracé, diagramme

Description générée automatiquement

1. **Algorithme des centres-mobiles**

**Une image contenant texte, capture d’écran, logiciel, Système d’exploitation

Description générée automatiquement**

1. **Génération des jeux de données**

Degré d’erreur de 6% :

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Degré d’erreur de 16%

Une image contenant capture d’écran, diagramme, texte, ligne

Description générée automatiquement

Une image contenant capture d’écran, diagramme, texte, ligne

Description générée automatiquementDegré d’erreur de 26%

1. **Résultats sur le premier dataset**

**Une image contenant capture d’écran, Caractère coloré

Description générée automatiquementUne image contenant capture d’écran, Caractère coloré

Description générée automatiquement**

**Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement**

On peut voir dans les résultats pour 2 applications de notre algorithme qu’il n’y a pas besoin d’améliorer l’initialisation puisque le résultat est correct et est toujours le même. On pourrait améliorer l’initialisation des centroïdes dans le cas K =2 en initialisant avec deux points éloignés, pour améliorer la vitesse de calcul on pourrait simplement prendre les 2 points extrêmes dans une coordonnée (aléatoirement x ou y). Faire cela permettrait d’accélérer la convergence des centroïdes.

1. **Résultats sur les 2 autres datasets**

Résultats dataset 2 :

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Résultat dataset 3 :

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Le pourcentage de points mal classés augmente avec le degré de mélange, ce à quoi on s’attendait.

1. **Algorithme CEM adapté aux trois jeux de données**

On choisit un modèle gaussien parcimonieux diagonal (car chaque cluster a une matrice de covariance diagonale).

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquementPremier dataset :

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Second dataset :

Une image contenant capture d’écran, Caractère coloré

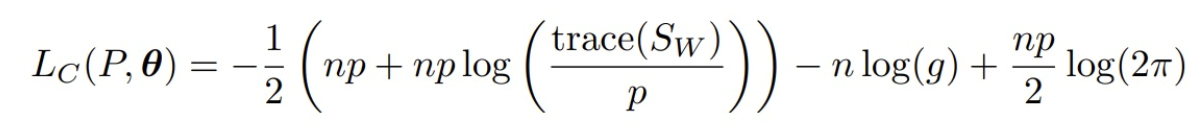
Description générée automatiquementUne image contenant capture d’écran, Caractère coloré

Description générée automatiquement

Une image contenant capture d’écran, Caractère coloré

Description générée automatiquementUne image contenant capture d’écran, Caractère coloré

Description générée automatiquementTroisième dataset :

La valeur de la vraisemblance maximisée est : 

Le modèle CEM ne parvient pas à de meilleurs résultats que l’algorithme des K-means dans notre cas car ici, comme présenté dans le cours, l’algorithme fonctionne comme les K-means.