

TP 8 : Classification et groupement de données

Exercice :

Générer et dessiner un ensemble de données à 2 dimensions, X_1 , qui consiste de $N = 400$ points. Ces points forment quatre groupes de tailles égaux. Chaque groupe contient des vecteurs générés par des distributions gaussiennes avec les moyennes $m_1 = [0,0]^T$, $m_2 = [10,0]$, $m_3 = [0,9]$, et $m_4 = [9,8]^T$, respectivement, et les matrices de covariance suivantes, respectivement

$$S_1 = I, \quad S_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1.5 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1.1 \end{bmatrix}, \quad S_4 = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}$$

Où I désigne la matrice d'identité de taille 2×2 .

1. Appliquer l'algorithme de k-means sur X_1 pour $m = 4$. En utilisant la fonction MATLAB prédéfinie (`rand`), initialiser les vecteurs de paramètre θ_j . Comparer les estimations finales de θ_j avec les valeurs moyennes des Gaussiens, m_j . Dessiner les vecteurs de paramètre θ_j et les points de X_1 . Utiliser des couleurs différentes pour représenter les données des différents clusters.
2. Répétez l'étape 1 pour $m = 3$.
3. Répétez l'étape 1 pour $m = 5$.
4. Répétez l'étape 1, maintenant avec les paramètres θ_j initialisé comme suit:

$$\theta_1(0) = [-2.0, -2.0]^T, \theta_2(0) = [-2.1, -2.1]^T, \theta_3(0) = [-2.0, 2.2]^T, \theta_4(0) = [-2.1, -2.2]^T.$$

5. Répétez l'étape 1, maintenant avec les paramètres θ_1 , θ_2 et θ_3 initialisés aléatoirement comme précédemment et $\theta_4(0)$ égal à $[20,20]^T$.
6. Commenter les résultats.