

Classification non supervisée

Partitionnement spectral

Charlotte Baey (charlotte.baey@univ-lille.fr)

1 Exercice 1 (R) : données simulées

1. Créer un échantillon de taille $n = 100$ selon le mélange Gaussien suivant :

$$p_1 \mathcal{N}(m_1, \Sigma) + p_2 \mathcal{N}(m_2, \Sigma) + p_3 \mathcal{N}(m_3, \Sigma),$$

$$\text{avec } p_1 = p_2 = p_3 = 1/3, m_1 = \begin{pmatrix} -3 \\ 0 \end{pmatrix}, m_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, m_3 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \text{ et } \Sigma = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix}.$$

2. Tracer le nuage de points obtenu. Les classes sont-elles (visuellement) bien séparées ?
3. Construire la matrice de similarité S à l'aide de la fonction de similarité Gaussienne de paramètre $\sigma = 1$.
4. Calculer la matrice d'adjacence W_{full} du graphe entièrement connecté associé à la matrice S obtenue dans la question précédente. Quel est le degré moyen du graphe ?
5. Tracer ce graphe à l'aide du package `igraph`.

Indication : on utilisera la fonction `graph.adjacency`, puis la fonction `simplify` pour supprimer les boucles. Pour améliorer la lisibilité, on pourra pondérer la largeur des arêtes par leur poids, en modifiant la valeur de `E(p)$width`, si `p` est le nom donné à la sortie de la fonction `graph.adjacency`.

6. En choisissant pour seuil ε le quantile empirique d'ordre 0.75 des indices de similarité, construire la matrice d'adjacence W_ε du graphe du ε -voisinage, et tracer le graphe correspondant. Quel est le degré moyen du graphe ?
7. En prenant $k = 2 \lfloor \log n \rfloor$, construire la matrice d'adjacence W_{knn} des k -plus proches voisins mutuels et tracer le graphe associé. Quel est le degré moyen du graphe ?
8. Construire les matrices Laplaciennes normalisées des graphes obtenus dans les questions 4, 6 et 7, ainsi que leurs valeurs propres.
9. Tracer les 10 premières valeurs propres pour chaque matrice Laplacienne. Que remarque t-on ?
10. Appliquer l'algorithme de partitionnement spectral normalisé pour chaque graphe de similarité, en prenant 3 classes. Comparer les résultats obtenus.

2 Exercice 2 (R) : segmentation d'image

1. Télécharger le fichier `irm_small.jpeg` et l'importer sous R en utilisant la fonction `readJPEG` du package `jpeg`.

2. Afficher l'image à l'aide de la fonction `image`.
3. Appliquer l'algorithme de partitionnement spectral normalisé afin d'identifier différentes zones dans l'image. On utilisera la fonction de similarité Gaussienne et le graphe du ε -voisinage en choisissant un seuil ε égal au quantile d'ordre 75% des indices de similarité. On justifiera le nombre de classes sélectionnées.
4. Afficher les classes obtenues.