

Mean Shift Clustering

N.EL MENANI

K.MARZAQ

F.TOUBA

March 28, 2023

Plan

- 1 Introduction
- 2 Algorithme
- 3 Avantages et inconvénients
- 4 Conclusion

Introduction

Mean Shift Clustering

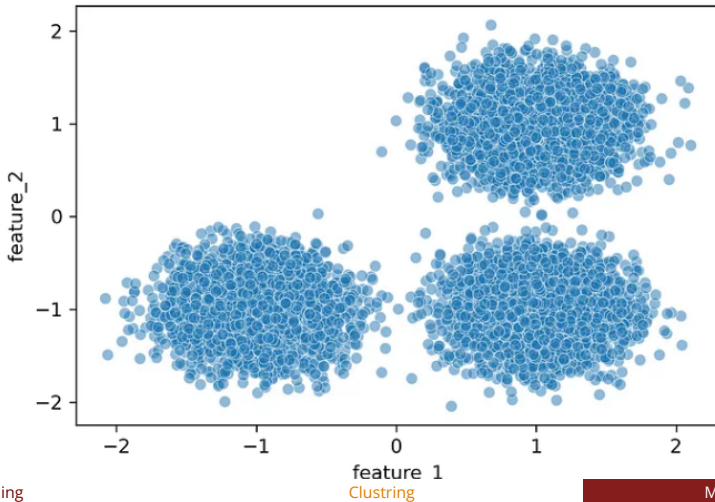
L'algorithme Mean Shift Cluster est une méthode de clustering non-paramétrique qui peut être utilisée pour identifier les groupes naturels de données dans un ensemble de données sans connaître à l'avance le nombre de clusters.

Introduction

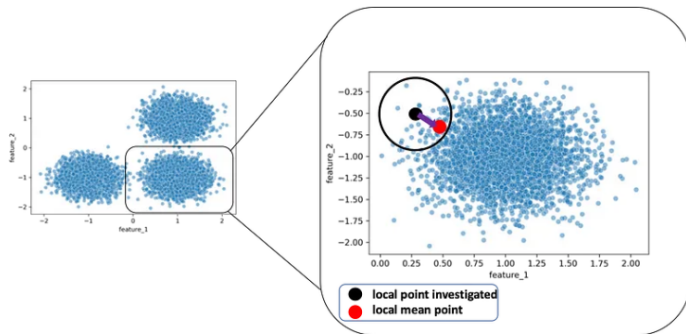
Principe :

- 1 Initialisation
- 2 Calcul de la densité locale
- 3 Calcul du vecteur de déplacement moyen
- 4 Mise à jour de la position
- 5 Répéter les étapes 2 à 4 pour tous les points de données
- 6 Fusion des clusters
- 7 Répéter les étapes 1 à 6

Initialisation



Initialisation



step 1 for a point in Mean Shift (image by author)

Calcul de la densité locale

Gauss :

$$K(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}}$$

Densité :

$$e_i = \sum_{j=1}^n K\left(\frac{\|x_i - x_j\|}{h}\right)$$

h : bande passante

Calcul du vecteur de déplacement moyen

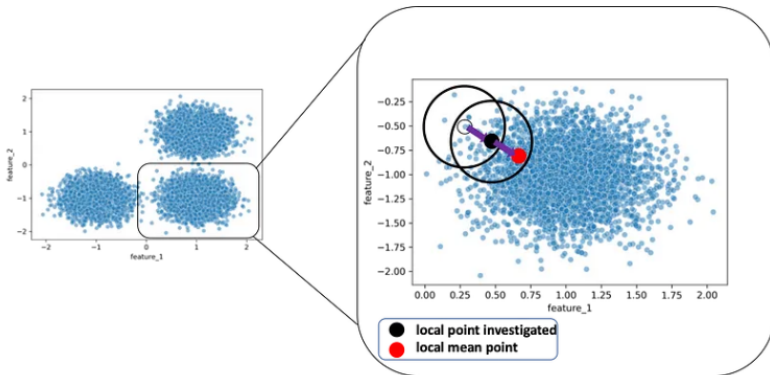
Poids :

$$w_i = K\left(\frac{\|x_i - x_c\|}{h}\right)$$

Vecteur de déplacement moyen :

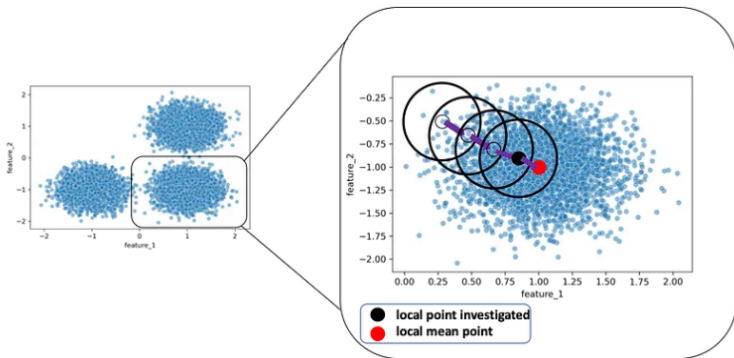
$$m(x) = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} - x$$

Mise à jour de la position



step 2 for a point in Mean Shift (image by author)

Répétition



step N for a point in Mean Shift (image by author)

Etape final

Fusion des clusters

les points de données qui convergent vers le même maximum local sont fusionnés en un cluster unique. Les points de données sont assignés à un cluster en fonction de leur proximité avec les maxima locaux de densité.

Les étapes 1 à 6 sont répétées plusieurs fois jusqu'à ce que les clusters finaux convergent et ne changent plus.

Avantages et inconvénients

Avantages

- Il n'a pas besoin de connaître à l'avance le nombre de clusters.
- Il est capable de trouver des clusters de formes et de tailles arbitraires.
- Il est robuste aux valeurs aberrantes.
- Il est relativement simple à mettre en œuvre

inconvénients

- Il est sensible à la taille de la fenêtre de recherche.
- Il est coûteux en termes de calcul.
- Il peut être sensible aux paramètres.
- Il peut souffrir de problèmes de convergence.

Conclusion

Conclusion

L'algorithme Mean Shift Clustering est itératif et adaptatif, car il peut trouver automatiquement la structure sous-jacente des données sans nécessiter de spécifier le nombre de clusters à l'avance.