

Projet « Data Analytics & Machine Learning »

Les critères suivants seront appréciés :

- Compte-rendu : clarté et pertinence des explications, qualité de la présentation (en particulier les équations), analyse et interprétation des résultats ;
- Travail fourni : conformité aux objectifs précisés dans le texte, quantité et qualité des résultats et conclusions ;
- Initiative et originalité.

De la théorie ...

Présenter (avec synthèse) l'algorithme *k-means* et les algorithmes EM et CEM appliqués au modèle de mélange gaussien.

... à la pratique

1. Simulation d'un modèle de mélange :
 - Écrire une fonction qui simule n points du plan suivant un modèle de mélange de deux lois gaussiennes définies par leurs proportions, centres, volumes, formes et orientations.
 - Donner quelques exemples de simulation et donner les représentations graphiques correspondantes.
2. Programmer l'algorithme CEM pour les modèles parcimonieux $[\pi, \lambda \mathbf{I}]$, $[\pi_k, \lambda \mathbf{I}]$, $[\pi, \lambda_k \mathbf{I}]$ et $[\pi_k, \lambda_k \mathbf{I}]$.
3. Comparer votre algorithme et l'algorithme des **kmeans** sur des jeux de données simulées et réelles. Pour les données simulées, on comparera tout d'abord les deux algorithmes lorsque les classes sont très séparées ($err \approx 0.05$ c.à.d. chevauchement des classes de 5%) ; pour cela, on étudiera des situations correspondant à des proportions égales, différentes ou très différentes et des situations correspondant à des volumes égaux, différents ou très différents. On fera les mêmes comparaisons lorsque les classes sont moyennement séparées ($err \approx 0.15$) et très séparées ($err \approx 0.30$). Pour les données réelles, vous pouvez choisir (en motivant ce choix) une application et un jeu de données parmi ceux proposés sur le site web : <http://archive.ics.uci.edu/ml/datasets.php>
4. Comment peut-on utiliser les modèles de mélange pour faire de l'apprentissage supervisé (classement) ? Illustrer votre réponse par une application de votre choix. Proposer ainsi une autre méthode d'apprentissage supervisé, la Présenter (avec synthèse) et la comparer avec les modèles de mélange.