

Nom et Prénom : SARA JEDDIN

Date de soutenance : 01/11/2024

Directeur de Thèse : BENTALEB YOUSSEF

Sujet de thèse :

Exploration approfondie des méthodes de classification : innovations et applications

Résumé:

Cette thèse vise à approfondir la compréhension des méthodes de classification, qu'elles soient supervisées ou non, en mettant en lumière leurs fondements théoriques, leurs applications pratiques, et les défis inhérents. Elle se divise en plusieurs volets, explorant d'abord les méthodes de classification supervisée telles que la régression linéaire, les arbres de décision, et les Machines à Vecteurs de Support (SVM), en soulignant les avantages spécifiques et les limites de chacune. Le deuxième volet se concentre sur la classification non supervisée, explorant l'Analyse en Composantes Principales (ACP) et de l'Analyse Factorielle des Correspondances (AFC) pour la réduction de la dimensionalité, ainsi que les techniques de clustering telles que le k-means et la Classification Ascendante Hiérarchique (CAH). La thèse propose des contributions novatrices, notamment l'introduction du concept de Barycentre pondéré pour améliorer la Classification Ascendante Hiérarchique (CAH). Une autre innovation réside dans l'amélioration de l'algorithme K-means par l'introduction d'une approche de barycentre pondéré pour déterminer le nombre optimal de clusters. La recherche s'articule autour d'une démarche empirique approfondie, développant de nouvelles méthodes pour pallier les limitations des techniques existantes. Ces avancées ont été consignées dans des publications de revues indexées, soulignant leur pertinence dans le paysage académique. En résumé, cette thèse aspire à être une ressource complète pour la communauté académique et professionnelle, renforçant la compréhension des utilisateurs potentiels et suscitant des réflexions critiques sur l'application de ces méthodes d'analyse de données. Mots clés : Classification, clusters, distance , inertie , pondération, barycentre pondéré

Abstract:

This thesis aims to deepen the understanding of classification methods, whether supervised or unsupervised, by shedding light on their theoretical foundations, practical applications, and inherent challenges. It is divided into several sections, initially exploring supervised classification methods such as linear regression, decision trees, and Support Vector Machines (SVM), highlighting the specific advantages and limitations of each. The second section focuses on unsupervised classification, exploring Principal Component Analysis (PCA) and Correspondence Factor Analysis (CFA) for dimensionality reduction, as well as clustering techniques such as k-means and Hierarchical Cluster Analysis (HCA). The thesis proposes innovative contributions, notably the introduction of the concept of weighted centroid to improve Hierarchical Cluster Analysis (HCA). Another innovation lies in the enhancement of the k-means algorithm by introducing a weighted centroid approach to determine the optimal number of clusters. The research is based on a comprehensive empirical approach, developing new methods to overcome the limitations of existing techniques. These advancements have been documented in indexed journal publications, highlighting their relevance in the academic landscape. In summary, this thesis aspires to be a comprehensive resource for the academic and professional community, enhancing the understanding of potential users and stimulating critical reflection on the application of these data analysis methods. Keywords : Classification, clusters, distance, inertia, weighting, weighted Barycenter