

Nom et Prénom : AL MAROUNI YASMINA

Date de soutenance : 02/11/2024

Directeur de Thèse : BENTALEB YOUSSEF

Sujet de thèse :

Optimisation de la régression PLS pour la modélisation latente en Big Data : Les défis de la haute dimension et des données catégoriques à valeurs manquantes

Résumé:

Face aux défis du Big Data dans un monde numérisé, cette thèse se penche sur l'optimisation de la régression des Moindres Carrés Partiels (PLS) et des approches factorielles, intégrant la modélisation latente pour une analyse approfondie de données multidimensionnelles. Elle vise à adresser des problématiques telles que la haute dimensionnalité, la diversité des variables, et les données manquantes. Le travail commence par explorer les défis et opportunités présentés par le Big Data pour l'analyse statistique, avant de se focaliser sur les principes sous-jacents à la régression multidimensionnelle et à la modélisation. Une attention particulière est portée à l'adaptation de la régression PLS pour la modélisation latente de données catégorielles, proposant des méthodes novatrices pour une analyse plus précise et robuste. L'étude détaille ensuite comment ces techniques sont appliquées et optimisées pour le Big Data, mettant en lumière l'Analyse en Composantes Principales (PCA), l'Analyse Factorielle des Correspondances (CFA), et explorant de nouvelles approches pour décrypter les structures complexes typiques des grands ensembles de données. Cette recherche ambitionne de contribuer de manière significative à l'analyse de données multidimensionnelles, offrant des solutions innovantes aux défis du Big Data et marquant une avancée importante dans l'évolution des méthodologies statistiques adaptées à cette nouvelle ère.

Abstract:

Facing the challenges of Big Data in a digitized world, this thesis focuses on optimizing Partial Least Squares (PLS) regression and factorial approaches, incorporating latent modeling for in-depth analysis of multidimensional data. It aims to address issues such as high dimensionality, variable diversity, and missing data. The work begins by exploring the challenges and opportunities presented by Big Data for statistical analysis, before focusing on the underlying principles of multidimensional regression and modeling. Particular attention is paid to adapting PLS regression for latent modeling of categorical data, proposing innovative methods for more precise and robust analysis. The study then details how these techniques are applied and optimized for Big Data, highlighting Principal Component Analysis (PCA), Correspondence Factor Analysis (CFA), and exploring new approaches to decrypt the complex structures typical of large data sets. This research aims to make a significant contribution to the analysis of multidimensional data, offering innovative solutions to the challenges of Big Data and marking an important advancement in the evolution of statistical methodologies adapted to this new era.