INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Departamento Académico de Estadística
Río Hondo # 1, Col. Progreso Tizapán,
Ciudad de México, C.P. 01080, México

Seminario Aleatorio

Sesión 432

Divisive Hierarchical Clustering of Variables Identified by Singular Vectors

Jan Bauer

Vrije Universiteit Amsterdam

Abstract

In this work, we present a novel method for divisive hierarchical variable clustering. A cluster is a group of elements that exhibit higher similarity among themselves than to elements outside this cluster. The correlation coefficient serves as a natural measure to assess the similarity of variables. This means that in a correlation matrix, a cluster is represented by a block of variables with greater internal than external correlation. Our approach provides a solution to identify such block structures in the correlation matrix using its singular vectors (eigenvectors). When divisively clustering p variables, there are 2^{p-1} *possible splits. Using the eigenvectors for cluster identification, we can effectively reduce these number to at most p(p-1), thereby making it computationally feasible. We elaborate on the methodology and outline the incorporation of linkage functions to assess distances between clusters. Additionally, we demonstrate that these distances are ultrametric, ensuring that the resulting hierarchical cluster structure can be uniquely represented by a dendrogram, with the heights of the dendrogram being interpretable. We also discuss the advantages of our method over agglomerative clustering, especially over their sensitivity to misspecification due to their greedy nature. To validate the efficiency of our approach, we perform simulation studies and analyze real world data from psychology and finance.*

Viernes 22 de noviembre de 2024,
13:00 horas de CDMX
Salón 301