PhD program in Statistics

DSS Statistics Seminar

January 26, 2026, 12:00

In person Room 24 (CU002)

Webinar https://uniroma1.zoom.us/j/83625004899?pwd=bXCtz0mp759PUh2IkqT0BUoVa0Uegg.1

Passcode: 123456

Robust clustering based on trimming

Luis Angel García Escudero

Departamento de Estadística e I.O. and IMUVA, Universidad de Valladolid

Clustering is one of the most widely used unsupervised learning techniques. However, it is well known that outliers can severely affect the performance of standard clustering methods. Clustered outliers, in particular, can be especially detrimental even for robust statistical procedures. This motivates the combination of concepts from Robust Statistics and Cluster Analysis to handle both clusters and outliers simultaneously through robust clustering approaches. Outliers may arise from errors in data collection, but they can also reflect genuine anomalies of significant scientific or commercial interest. Therefore, anomaly detection becomes a goal in itself, and the use of robust clustering techniques for their automated identification in heterogeneous datasets can be a very useful tool. Among existing robust clustering methods, we focus on those based on trimming. Trimming provides an intuitive framework: a small fraction of the most extreme observations is removed, and standard clustering methods are then applied to the remaining data. When combined with suitable constraints on cluster dispersion parameters, trimming-based approaches have demonstrated strong performance and can be efficiently implemented using available algorithms.

This seminar will start from the fundamental principles of robustness and provide a guided overview of various robust model-based clustering techniques across different scenarios and frameworks. It will cover the necessary algorithms for implementation and discuss solutions to key challenges, including those arising from increasing dimensionality.

SAPIENZA UNIVERSITÀ DI ROMA