



Propuesta de Trabajo Fin de Grado en Matemáticas (curso 2022–2023)

<i>Responsable de tutorización:</i> José Luis Romero Béjar <i>Departamento:</i> Estadística e Investigación Operativa <i>Correo electrónico:</i> jlrbejar@ugr.es
<i>Responsable de cotutorización:</i> José Miguel Angulo Ibáñez <i>Departamento:</i> Estadística e Investigación Operativa <i>Correo electrónico:</i> jmangulo@ugr.es
<i>(Rellenar sólo en caso de que la propuesta esté realizada a través de un estudiante):</i> <i>Estudiante que propone el trabajo:</i> Pablo Andrades Moya

<i>Título del trabajo:</i> Agrupamiento y Clasificación de Datos Multivariantes
<i>Tipología del trabajo (marcar una de las siguientes casillas):</i> <input checked="" type="checkbox"/> <i>Complemento de profundización</i> <input type="checkbox"/> <i>Divulgación de las Matemáticas</i> <input type="checkbox"/> <i>Docencia e innovación</i> <input checked="" type="checkbox"/> <i>Herramientas informáticas</i> <input type="checkbox"/> <i>Iniciación a la investigación</i>
<i>Materias del grado relacionadas con el trabajo:</i> Probabilidad, Estadística Multivariante,...
<i>Descripción y resumen de contenidos:</i> El análisis cluster (AC) es una técnica multivariante cuyo principal objetivo es agrupar objetos formando conglomerados (clusters) cuyos elementos presenten un alto grado de homogeneidad interna y heterogeneidad externa. En otras palabras, el AC es un procedimiento exploratorio que permite encontrar estructuras con similitudes en un conjunto de datos con cierta variabilidad. La motivación de esta técnica va a asociada a la necesidad de diseñar una estrategia que permita definir grupos homogéneos. En este sentido es un método de clasificación. Si bien esta técnica tiene la gran ventaja de su objetividad - no se tiene información acerca de los grupos de clasificación sino que estos se construyen durante el desarrollo del análisis -, presenta desventajas relacionadas con que es meramente descriptiva y no inferencial, junto con que no proporciona una solución única. El AC tiene aplicación en un amplio número de situaciones en distintas áreas de las ciencias: psicología, biología, sociología, economía, ingeniería, investigación de mercados y marketing, etc. En este trabajo se pretende que el alumno, a partir de una discusión general sobre la problemática de la clasificación estadística, con una referencia introductoria a las principales metodologías consolidadas en la literatura relacionada, profundice en el enfoque de AC dando una visión completa y actualizada de sus aspectos fundamentales, su implementación y su aplicación.

Actividades a desarrollar:

- Exploración bibliográfica sobre metodologías consolidadas de agrupamiento y clasificación estadística, con especial referencia a su estado actual, y exposición sintética del conocimiento desde una perspectiva global en el contexto del análisis estadístico de datos multivariantes.
- Profundización en la metodología de análisis cluster a partir de datos multivariantes, con una clara identificación de los elementos conceptuales inherentes, y exposición de sus fundamentos matemáticos y aspectos metodológicos.
- Aplicación en un estudio con datos reales mediante el uso y desarrollo eventual de procedimientos computacionales y gráficos relacionados.

Objetivos matemáticos planteados

Identificación de los aspectos conceptuales inherentes a los enfoques de agrupamiento y clasificación objeto de estudio y su formalización matemática.

Análisis pormenorizado de los fundamentos matemáticos que sustentan el desarrollo de la metodología objeto central del trabajo.

Desarrollo de un estudio aplicado e interpretación de resultados con referencia precisa a la fundamentación matemática de la metodología.

Bibliografía

- [1] Y. FUJIKOSHI, V. V. ULYANOV, R. SHIMIZU, *Multivariate Statistics. High-Dimensional and Large-Sample Approximations*, (John Wiley & Sons, New Jersey, 2010).
- [2] W. K. HARDLE, L. SIMAR, *Applied Multivariate Statistical Analysis*, (Springer, Berlin, 2015).
- [3] A. J. IZENMAN, *Modern Multivariate Statistical Techniques. Regression, Classification, and Manifold Learning*, (Springer, New York, 2008).
- [4] D. J. OLIVE, *Robust Multivariate Analysis*, (Springer, Switzerland, 2017).
- [5] E. PEKALSKA, R. P. W. DUIN, *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications*, (World Scientific, Singapore, 2005).
- [6] A. C. RENCHER, W. F. CHRISTENSEN, *Methods of Multivariate Analysis*, (John Wiley & Sons, New Jersey, 2012).
- [7] N. H. TIMM, *Applied Multivariate Analysis*, (Springer, New York, 2002).
- [8] D. ZELTERMAN, *Applied Multivariate Statistics with R*, (Springer, New York, 2015).

Firma del estudiante

Firma del responsable de tutorización

(sólo para trabajos propuestos por estudiantes) (sólo para trabajos propuestos por estudiantes)

*Firma del responsable de cotutorización
(sólo para trabajos propuestos por estudiantes)*

En Granada, a 11 de mayo de 2022.