



Propuesta de Trabajo Fin de Grado en Matemáticas (curso 2021-2022)

Responsable de tutorización: Nuria Rico Castro
Departamento: Estadística e Investigación Operativa
Correo electrónico: nrico@ugr.es

Responsable de cotutorización: Desirée Romero Molina
Departamento: Estadística e Investigación Operativa
Correo electrónico: deromero@ugre.s

(Rellenar sólo en caso de que la propuesta esté realizada a través de un estudiante)
Estudiante que propone el trabajo: Laura Gavira Morillo

Título del trabajo: Estudio del Análisis Cluster. Aplicación a datos Reales.

Tipología del trabajo (marcar una o varias de las siguientes casillas):

- Complementario de profundización
- Divulgación de las Matemáticas
- Docencia e innovación
- Herramientas informáticas
- Iniciación a la investigación

Materias del grado relacionadas con el trabajo:
Estadística Multivariante, Estadística Computacional.

Descripción y resumen de contenidos:

El Análisis Cluster (AC) es una técnica estadística que se engloba dentro de las denominadas técnicas multivariantes, por lo tanto es un método estadístico cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado.

Como en cualquier técnica estadística multivariante su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes son incapaces de conseguir.

El AC se engloba, a su vez, dentro de los denominados métodos de interdependencia ya que no distingue entre variables dependientes e independientes y su objetivo es clasificar una muestra de entidades (individuos o variables) en un número pequeño de grupos de forma que las observaciones pertenecientes a un grupo sean homogéneas entre sí y a su vez heterogéneas con respecto al resto. A diferencia de otras técnicas multivariantes que realizan un análisis similar, en el AC se desconoce el número y la composición de los grupos que se buscan.

Existen dos cuestiones fundamentales en la construcción de un método de AC; la primera de ellas es cómo establecer una medida de homogeneidad entre individuos o entre variables. En este aspecto juega un papel fundamental la teoría de la medida. La segunda cuestión es la de establecer cómo llevar a cabo la agrupación de elementos similares. Se diferencian en este sentido los métodos jerárquicos, donde la agrupación (o desagrupación) se realiza por etapas y los métodos no jerárquicos, los cuales pueden ser recursivos pero la agrupación en todo caso se hace de forma simultánea para todos los individuos o variables.

La asignatura Estadística Multivariante puede ser una buena base para el alumno ya que en ella se estudian algunas técnicas estadísticas como ejemplo de aplicación de la teoría multivariante que se

desarrolla en la materia. Por otro lado, la asignatura Estadística Computacional será fundamental para el estudiante ya que le puede proporcionar los conocimientos necesarios para usar el paquete R con el fin de aplicar la técnica o incluso para poder hacer una implementación propia.

Actividades a desarrollar:

El estudiante deberá comenzar recopilando bibliografía sobre el AC para su posterior estudio, buscando las bases de esta técnica y todos los desarrollos que se han hecho de la misma.

A continuación el alumno/a estudiará la base matemática en que se basa esta técnica, donde tendrá que usar conocimientos de otras áreas de las matemáticas.

Finalmente, el estudiante deberá aplicar la técnica, utilizando algún paquete estadístico como R, a un conjunto de datos reales, enfrentándose a un gran conjunto de datos y no a meros ejemplos básicos de libros.

Si el estudiante así lo desea puede ampliar el trabajo incluyendo una implementación propia de la técnica.

Objetivos matemáticos planteados

Recopilación bibliográfica

Estudio matemático de la técnica

Aplicación a datos reales

Implementación de la técnica

Bibliografía para el desarrollo matemático de la propuesta:

- Anderson, T.W. (2003, 3º ed.) An introduction to Multivariate Statistical Analysis. John Wiley & Sons.
- Cuadras, C. M. (2018, 2º ed.) Nuevos métodos de Análisis Multivariante. CMC Editions.
- Peña Sánchez de Rivera, D. (2002) Análisis de Datos Multivariantes. McGraw-Hill.
- Timm, N. H. (2002) Applied Multivariate Analysis, Springer.

Otras referencias (si procede):

- <https://cran.r-project.org/web/packages/cluster/cluster.pdf>



Firma del estudiante

(solo para trabajos propuestos por estudiantes)



Núria Pardo Castro

Firma del responsable de tutorización

(solo para trabajos propuestos por estudiantes)



Firma del responsable de cotutorización

(solo para trabajos propuestos por estudiantes)

En, Granada, a 26 de abril de 2021