UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

Propuesta TFG. Curso 2024/2025

GRADO: Grado en Matemáticas

CÓDIGO DEL TFG: 270-073-2024/2025

1. DATOS BÁSICOS DEL TFG:

Título: Profundidad de datos en un contexto multivariante. Aplicación a problemas de clustering

Descripción general (resumen y metodología):

El concepto de 'profundidad de datos' (data depth) fue introducido con el propósito de establecer ordenaciones totales de conjuntos de datos en un contexto mutivariante basadas en criterios de centralidad, siendo tales ordenaciones útiles en relación con distintos procedimientos de inferencia relativos a rangos, y particularmente en la detección de observaciones anómalas (outliers) y el diseño de métodos robustos. Desde su aparición se han propuesto diversas funciones de profundidad, sobre la base de determinados principios axiomáticos formales. La teoría y metodologías desarrolladas en torno a este concepto han tenido y tienen importantes aplicaciones en múltiples campos.

En este Trabajo de Fin de Grado se trata, en primer lugar, de introducir los aspectos matemáticos fundamentales, tanto conceptuales como formales, en relación con la profundidad de datos. A partir de una revisión acerca de distintos problemas y enfoques metodológicos basados en el concepto de profundidad de datos surgidos en la literatura, el trabajo se centrará de forma específica en métodos de agrupamiento (clustering) desarrollados bajo criterios de profundidad de datos. Finalmente, se analizarán y discutirán aspectos computacionales relativos a la implementación de distintos procedimientos, así como la interpretación de los resultados derivados de éstos en las aplicaciones.

La metodología de trabajo contempla las siguientes fases:

- Revisión de literatura base y de carácter específico en relación con el concepto de profundidad de datos.
- Síntesis sobre aspectos teóricos fundamentales, así como enfoques metodológicos en relación con distintos problemas estadísticos en el contexto multivariante.
- Estudio específico en profundidad de algunos enfoques y procedimientos clustering basados en criterios de profundidad de datos.
- Revisión sobe aspectos computacionales y gráficos, así como aplicaciones, mediante software disponible en el entorno R.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

- Identificación de los aspectos conceptuales y formales que constituyen el fundamento teórico del análisis basado en profundidad de datos.
- Conocimiento de resultados fundamentales, enfoques metodológicos, así como su implementación en procedimientos computacionales, en este contexto.
- Análisis pormenorizado de los fundamentos matemáticos que sustentan el desarrollo de algunas metodologías de clustering basadas en criterios de profundidad de datos.
- Desarrollo de un estudio aplicado e interpretación y discusión de resultados con referencia precisa a la fundamentación matemática de la metodología utilizada.

Bibliografía básica:

Cascos, I. (2009) Data depth: multivariate statistics and geometry. En: New Perspectives in Stochastic Geometry, W.S. Kendall, I. Molchanov (eds.), pp. 398-424. Oxford University Press.

Hoberg, R. (2000) Cluster analysis based on data depth. En: Data Analysis, Clasification, and Related Methods, H.A.L. Kiers, J.-P. Rasson, P.J.F. Groenen, M. Schader (eds.), pp. 17-22. Springer.

Liu, R.Y., Parelius, J.M., Singh, K. (1997) Multivariate analysis by data depth: descriptive statistics, graphics and inference. The Annals of Statistics 27, 783-858.

Pokotylo, O., Mozharovski, P., Dycherhoff, R. (2019) Depth and depth-based classification with R package ddalpha. Journal of Statistical Software 91, doi:10.18637/jss.v091.i05

Mosler, K., Mozharovskyi, P. (2022) Choosing among notions of multivariate depth statistics. Statistical Science 37, 348-368.

Zuo, Y., Serfling, R. (2000) General notions of statistical depth function. The Annals of Statistics 28, 461-482.

Recomendaciones y orientaciones para el estudiante:

Se requieren conocimientos previos sobre Estadística Multivariante, Inferencia Estadística y Estadística Computacional, al nivel de las asignaturas correspondientes a estas materias que se imparten en el Grado en Matemáticas.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: JOSÉ MIGUEL ANGULO IBÁÑEZ

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: jmangulo@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: SOFIA SALCEDO ARMIJO

Correo electrónico: sofiasalcedo@correo.ugr.es