

Línea de Trabajo fin de Máster

(Fecha última actualización: 12/10/2020)

DOBLE MÁSTER DE FORMACIÓN DEL PROFESORADO DE SECUNDARIA Y BACHILLERATO, FORMACIÓN PROFESIONAL Y ENSEÑANZA DE IDIOMAS - ESTADÍSTICA APLICADA. CURSO ACADÉMICO 2020-21	
Título	Análisis de datos ("Big-data"). Técnicas de MDS aplicadas a datos de disimilaridad
Profesor(es)	José Fernando Vera
Descripción	<p>Uno de los problemas más interesantes en muchas disciplinas se plantea cuando necesitamos medir y entender las relaciones entre objetos o variables, siendo desconocidas las dimensiones subyacentes de los mismos. El problema resulta de especial interés cuando la información disponible está dada en términos de disimilaridades o pseudodistancias entre los elementos que pretendemos estudiar, y su tamaño además es muy grande, lo que hoy es conocido como "big-data". El análisis multidimensional de estructuras de proximidad o "Multidimensional Scaling" (MDS), en combinación con otras técnicas de reducción como el análisis clúster, permiten abordar este problema y facilitar la interpretación de los resultados.</p> <p>En este contexto, dada la gran flexibilidad de la técnica y puesto que no existen soluciones exactas para la estimación de los parámetros en MDS, el desarrollo de técnicas específicas para cada problema en cuestión y su tratamiento computacional constituye una área de investigación muy atractivo y de gran actualidad, en la que diversas técnicas estadísticas y de optimización heurística juegan actualmente un papel importante. Puesto que en MDS la información proviene de datos de proximidad (pseudodistancias), ofrece una alternativa a los problemas actuales para el análisis de grandes conjuntos de datos aplicable tanto a problemas clásicos de análisis multivariante como al análisis funcional.</p>
Objetivos particulares	Estudio y aplicación de técnicas estadísticas para el análisis y representación multidimensional de grandes conjuntos de datos de disimilaridad.
Prerrequisitos y recomendaciones	Haber cursado o estar cursando la asignatura: Análisis de datos. Técnicas aplicadas a datos de proximidad.
Plan de trabajo	<ul style="list-style-type: none"> -Estudio de las técnicas adecuadas para la resolución de problemas y el análisis de datos de proximidad mediante MDS. -Determinación de un modelo aplicado para el análisis real de datos de proximidad. -Implementación del modelo mediante R o MatLab. -Elaboración de la memoria y presentación de resultados.
Competencias generales y específicas	<p>G01. Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.</p> <p>G02. Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.</p> <p>G03. Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.</p> <p>G04. Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u</p>



oral, a un público tanto especializado como no especializado.

G05. Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

G06. Saber utilizar herramientas de búsqueda de recursos bibliográficos.

G08. Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.

Competencias específicas:

E01. Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.

E02. Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.

E03. Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.

E04. Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.

E07. Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.

E08. Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.

Bibliografía

1. BORG, I. & GROENEN, P.J.F. (2005).- Modern Multidimensional Scaling. Theory and Applications. Second Edition. Springer Series in Statistics. Springer.
2. COX, T.F. & COX, M.A.A. (2001).-Multidimensional Scaling. Second Edition. Monographs on statistics and applied probability, 59. London: Chapman Hall.
3. HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2010). The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second Edition. Springer Series in Statistics. Springer.
4. MARDIA K.V. & BIBBY, K.J.M. (1997). Multivariate Analysis. London: Academic Press.
5. VERA, J.F., MACÍAS, R., and ANGULO, J.M., (2008), "Non-Stationary Spatial Covariance Structure Estimation in Oversampled Domains by Cluster Differences Scaling with Spatial Constraints", Stochastic Environmental Research and Risk Assessment, 22, 95–106.
6. VERA, J.F., MACÍAS, R., and HEISER, W.J., (2009a), "A Latent Class Multidimensional Scaling Model for Two-Way One-Mode Continuous Rating Dissimilarity Data", Psychometrika, 74(2), 297–315.
7. VERA, J.F., MACÍAS, R., HEISER, W.J. 2013. Cluster Differences Unfolding for Two-Way Two-Mode Preference Rating Data.. Journal of Classification. 30: 370-396
8. VERA, J.F., ROIJ, M., HEISER, W.J. 2014. A latent class distance association model for cross-classified data with a categorical response variable. British Journal of Mathematical and Statistical Psychology. 67: 514-540
9. VERA, J.F., RIVERA, C.D. 2014. A Structural Equation Multidimensional Scaling Model for One-Mode Asymmetric Dissimilarity Data. Structural Equation Modeling. 21: 54-62
10. VERA, J.F., MACÍAS, R., 2017. Variance-based cluster selection criteria in a k-means framework for one-mode dissimilarity data. Psychometrika. 82: 275-294
11. VERA, J.F. 2017. Distance stability analysis in multidimensional scaling using the jackknife method. British Journal of Mathematical and Statistical Psychology. 70: 25-41
12. VERA, J.F., MAIR, P. 2019. SEMDS: An R Package for Structural Equation Multidimensional Scaling. Structural Equation Modeling: A Multidisciplinary Journal, 26:5, 803-818
13. VERA, J.F., ROIJ, M. 2020. A Latent Block Distance-Association Model for Profile by Profile Cross-Classified Categorical Data, Multivariate Behavioral Research, 55:3, 329-343