

GUIA DOCENTE DEL CURSO	
Nombre del curso	Análisis de datos. Técnicas aplicadas a datos de proximidad.
Profesor(es)	José Fernando Vera. (http://www.ugr.es/local/jfvera) Pedro A. García. (http://www.ugr.es/local/pagarcia)
Descripción	<p>Uno de los problemas más interesantes en muchas disciplinas se plantea cuando necesitamos medir y entender las relaciones entre objetos, siendo desconocidas las dimensiones subyacentes de los mismos, especialmente en aquellas situaciones en las que la información disponible se refiere exclusivamente a la semejanza o desemejanza entre los elementos que son motivo de estudio. El análisis multidimensional de estructuras mediante proximidades o <i>Multidimensional Scaling</i> (MDS) puede definirse como un conjunto de técnicas para el análisis de datos de similaridad o de disimilaridad sobre un conjunto de objetos o variables. En general, las medidas de proximidad son modelizadas mediante distancias en un espacio métrico de dimensión baja para obtener una representación visual de la estructura de los datos, lo que además de resultar más fácil de entender que una tabla de datos, permite mostrar la información esencial minimizando las perturbaciones debidas a errores.</p> <p>Existen diferentes tipos de MDS en función de la geometría utilizada para la representación de los datos, el tratamiento estadístico de los errores, o los procedimientos empleados para encontrar una representación óptima de los mismos, entre otras características. La gran flexibilidad del MDS ha hecho que además, algunos modelos hayan sido combinados teóricamente con otras técnicas tradicionales del Análisis Multivariante como el análisis clúster para facilitar la interpretación de los datos. Desde otra perspectiva, el MDS también ha sido empleado como herramienta fundamental en campos de la estadística tan diferentes como los procesos espacio-temporales.</p> <p>Dada la gran flexibilidad de la técnica y puesto que no existen soluciones exactas para la estimación de la configuración en MDS, el desarrollo de las técnicas específicas de cada problema en cuestión y su tratamiento computacional constituye una área de investigación muy atractivo y de gran actualidad, en la que las técnicas de optimización heurística juegan un papel importante.</p>
Objetivos particulares	<ul style="list-style-type: none"> • Introducir al alumno en el análisis de datos de proximidad y en la geometría de la técnica de Multidimensional Scaling (MDS). • Adquirir los conocimientos y habilidades necesarios en relación a la minería de datos de proximidad y a la metodología del análisis estadístico computacional asociado. • Describir los principales modelos de MDS y su aplicación mediante el manejo de software estadístico GNU de alto nivel. • Enseñar a elegir la metodología computacional en relación al software estadístico más adecuado para cada análisis de datos de proximidad. • Motivar la búsqueda de información individual y en equipo, así como el estudio y aplicación de nuevas técnicas estadísticas de MDS.
Prerrequisitos y recomendaciones	Para el desarrollo de la investigación en este campo se requiere por parte del alumno un conocimiento medio-avanzado en Matemáticas, Estadística y Cálculo de Probabilidades. En particular, es recomendable tener conocimientos medios en Análisis Multivariante y Técnicas de Regresión. Por otra parte, el conocimiento de lenguajes de programación avanzada como R o MatLab, así como paquetes estadísticos avanzados como SPSS es a su vez muy recomendable. Es por tanto aconsejable, aunque no imprescindible, que el alumno haya realizado alguna asignatura relacionada con la Estadística Computacional y el Análisis Exploratorio de Datos.
Contenidos	<ol style="list-style-type: none"> 1. Análisis de datos de proximidad. Multidimensional Scaling 2. MDS clásico. Análisis de coordenadas principales 3. MDS métrico y no métrico 4. Diferencias individuales en MDS

	<ol style="list-style-type: none"> 5. Análisis de datos de proximidad con MDS 6. Unfolding. 7. Procruster. 8. Otros modelos de MDS 9. Clasificación y MDS
Metodología	<p>El curso se desarrollará en el primer cuatrimestre, aunque los bloques de contenidos concluirán una semana antes para destinar los últimos días a resolución de dudas y orientación en la realización de las actividades propuestas.</p> <p>La temporización fijada para el curso plantea el estudio de entre 1 y 2 horas diarias durante el primer cuatrimestre. No obstante, esto es una recomendación o sugerencia. Evidentemente, cada alumno podrá fijar su propio ritmo de estudio, decidir cuándo se conecta a la plataforma, cuándo realizar una actividad, cuándo remitir una aportación al foro, etcétera.</p> <p>Como se puede apreciar, en general los temas son independientes entre sí, aunque complementarios, y no se solapan en ningún momento: salvo que el tutor lo advierta de forma puntual, no se pasará de un bloque a otro hasta que se haya dado por concluido y se hayan entregado las actividades correspondientes.</p>
Bibliografía	<ol style="list-style-type: none"> 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. KRUSKAL, B. & WISH, M (1981). Multidimensional Scaling. Sage. 4. MARDIA K.V. & BIBBY, K.J.M. (1997). Multivariate Analysis. London: Academic Press. 5. HOAGLIN, D., MOSTELLER, F. & TUKEY, J.W.(1983).-Understanding robust and Exploratory Data Analysis. New York. John Wiley & Sons. 6. SEBER, G.A.F. (1984). Multivariate Observations. New York. John Willey. 7. ARABIE, P. & CARROLL, J.D. & DeSARBO, W. (1987). Three-Way scaling and clustering. Sage.
Criterios de evaluación	<p>Se considerará aprobado el curso y se expedirá el correspondiente certificado de aprovechamiento con la superación del 85% del contenido propuesto, tanto de la parte teórica como práctica. Para la evaluación se tendrá en cuenta:</p> <ul style="list-style-type: none"> • Actitud participativa del alumno (10%) • Acceso a la plataforma (módulo de contenidos, glosario, utilización de enlaces web y bibliografía complementaria, etc.) (5%) • Participación en foros y chats (10%) • Entrega de actividades (40%) • Realización de consultas y utilización del resto de herramientas de comunicación (10%) • Prácticas (25%)