

Guía docente de la asignatura

Fecha de aprobación por la Comisión Académica: 13/07/2022

Análisis de Datos. Técnicas Aplicadas a Datos de Proximidad (M42/56/2/31)**Máster**

Máster Universitario en Estadística Aplicada

MÓDULO

Módulo I: Aplicaciones de la Estadística

RAMA

Ciencias

CENTRO RESPONSABLE DEL TÍTULO

Escuela Internacional de Posgrado

Semestre

Primero

Créditos

4

Tipo

Optativa

Tipo de enseñanza

Enseñanza Virtual

PRERREQUISITOS Y/O RECOMENDACIONES

Es recomendable que el alumno tenga conocimientos del lenguaje R y/o MatLab, así como un nivel medio alto en matemáticas acorde con los requisitos del Máster.

BREVE DESCRIPCIÓN DE CONTENIDOS (Según memoria de verificación del Máster)

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 Multidimensional Scaling (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.

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.

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.

Objetivos particulares

- 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 con 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 conl 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.

COMPETENCIAS

COMPETENCIAS BÁSICAS

- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
- CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

RESULTADOS DE APRENDIZAJE (Objetivos)



El alumno sabrá/comprenderá:

- 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.
- Elegir la metodología computacional en relación al software estadístico más adecuado para cada análisis de datos de proximidad.
- Buscar información individual y en equipo, así como el estudio y aplicación de nuevas técnicas estadísticas de MDS.

El alumno será capaz:

- Aplicar estas técnicas a casos concretos.

PROGRAMA DE CONTENIDOS TEÓRICOS Y PRÁCTICOS

TEÓRICO

TEMARIO TEÓRICO:

- Tema 1. Análisis de datos de proximidad. Multidimensional Scaling.
- Tema 2. MDS clásico. Análisis de coordenadas principales.
- Tema 3. MDS métrico y no métrico.
- Tema 4. Diferencias individuales en MDS.
- Tema 5. Análisis de datos de proximidad con MDS. Unfolding. Procruster. Otros modelos de MDS. Clasificación y MDS

PRÁCTICO

TEMARIO PRÁCTICO:

Seminarios/Talleres

- MDS con R
- MDS con SPSS

BIBLIOGRAFÍA

BIBLIOGRAFÍA FUNDAMENTAL

- BORG, I. & GROENEN, P.J.F. (2005).- Modern Multidimensional Scaling. Theory and Applications. Second Edition. Springer Series in Statistics. Springer.



- COX, T.F. & COX, M.A.A. (2001).-Multidimensional Scaling. Second Edition. Monographs on statistics and applied probability, 59. London: Chapman Hall.
- KRUSKAL, B. & WISH, M (1981). Multidimensional Scaling. Sage.
- MARDIA K.V. & BIBBY, K.J.M. (1997). Multivariate Analysis. London: Academic Press.
- ARABIE, P. & CARROLL, J.D. & DeSARBO, W. (1987). Three-Way scaling and clustering. Sage.

BIBLIOGRAFÍA COMPLEMENTARIA

- HOAGLIN, D., MOSTELLER, F. & TUKEY, J.W.(1983).-Understanding robust and Exploratory Data Analysis. New York. John Wiley & Sons.
- SEBER, G.A.F. (1984). Multivariate Observations. New York. John Willey.

ENLACES RECOMENDADOS

<https://www.ibm.com/products/spss-statistics>

EVALUACIÓN (instrumentos de evaluación, criterios de evaluación y porcentaje sobre la calificación final)

EVALUACIÓN ORDINARIA

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:

- 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%)
- Trabajo final (25%)

EVALUACIÓN EXTRAORDINARIA

- 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 (60%)
- Trabajo final (25%)

EVALUACIÓN ÚNICA FINAL

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.

