

Titulación	Tipo	Curso
2503852 Estadística Aplicada	OB	3

Contacto

Nombre: María Merce Farre Cervello

Correo electrónico: merce.farre@uab.cat

Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

Prerrequisitos

Es imprescindible un curso previo de álgebra lineal, así como los cursos de probabilidades, dustrubuciones multidimensionales e inferencia estadística. También se supone que sabe utilizar con agilidad el lenguaje R.

Objetivos y contextualización

La necesidad de procesar gran cantidad de datos con muchas variables de naturaleza diversa, aunque reduciendo la información que no sea relevante y descubriendo patrones de asociación entre variables y/o casos, han ocasionado el desarrollo de un gran número de procedimientos que se sitúan en el escenario multivariante. En Aprendizaje No Supervisado se tratan los métodos más apropiados para describir, explorar y modelizar datos vectoriales. La asignatura se plantea como un primer contacto del estudiante con el mundo del llamado "aprendizaje estadístico" ("statistical learning"), para que comprenda la potencia y aplicabilidad, y al mismo tiempo las limitaciones, de los métodos, algunos de los que se basan en ideas heurísticas bastante intuitivas. La mayoría de los métodos trabajados en el curso son de tipo no supervisado, es decir, no se dispone de un conjunto de casos con respuestas conocidas que permitan evaluar el método. El enfoque de la asignatura es eminentemente aplicado en lo que se refiere al trabajo con datos usando el potencial del software libre R, acompañado del rigor y generalidad adecuados en la definición de los modelos teóricos y los correspondientes métodos de análisis y validación de resultados.

Resultados de aprendizaje

1. CM11 (Competencia) Crear nuevos modelos de aprendizaje automático, ejecutando experimentos para demostrar su viabilidad y mejora de rendimiento con respecto al estado del arte.
2. CM12 (Competencia) Valorar la existencia de desigualdades por razón de género en las bases de datos, para evitar los sesgos en la toma de decisiones automática (algorítmica).
3. KM16 (Conocimiento) Reconocer modelos de aprendizaje automático, supervisado y no supervisado, profundo y generativo, fomentando la innovación en el ámbito de la estadística.
4. KM16 (Conocimiento) Reconocer modelos de aprendizaje automático, supervisado y no supervisado, profundo y generativo, fomentando la innovación en el ámbito de la estadística.

Contenido

Aprendizaje estadístico (statistical learning) y reducción de la dimensión

- Aprendizaje supervisado y no-supervisado. Clasificación de los métodos. Ejemplos.
- Vectores aleatorios. El vector de esperanzas y las matrices de covarianzas y correlaciones. Propiedades.
- Datos multivariantes. El vector de medias y las matrices de covarianza y correlación muestrales.
- Estimadores de máxima verosimilitud en el caso Gaussiano.
- Las descomposiciones espectral (SD) y en valores singulares (SVD).
- Teorema fundamental de maximización de formas cuadráticas con restricciones.

Métodos factoriales I: Análisis de componentes principales (ACP)

- Introducción al ACP. Definición de componentes. El resultado fundamental.
- Criterios para decidir el número de componentes: las componentes principales.
- Gráficas de variables y de individuos. Estandarizaciones.
- Análisis de la matriz de vectores propios, y de matrices derivadas, por filas y por columnas.
- Enfoque geométrico del ACP.

Métodos factoriales II: Análisis factorial (AF)

- El modelo factorial. Comunalidades y especificidades.
- Teorema de descomposición de la matriz de covarianza.
- Existencia y unicidad del modelo factorial? Rotaciones.
- Métodos de estimación de los parámetros del modelo factorial. Las puntuaciones factoriales (scores): estimación o predicción.
- Interpretación de los resultados. Comparación del ACP y el AF.

Métodos factoriales III: Escalamiento multidimensional (EMD) y análisis de correspondencias (AC)

- Objetivos y métodos.
- EMD: clásico y métrico.
- EMD no métrico.
- Distancias, proximidades y disimilaridades.
- Datos categóricos: distancia ji -cuadrado y otras.
- Análisis de correspondencias (AC) como caso particular de EMD
- Perfiles e inercia. Descomposición de la inercia.
- Representación gráfica e interpretación de los resultados del AC.

Análisis de clústers (ACL)

- Comparación de diferentes enfoques. Ejemplos.
- Análisis y validación de los resultados de los clústers.
- Conglomerados jerárquicos: Funciones de enlace.
- Métodos basados en centroides: El método de las k -medias.
- Métodos basados en modelos probabilísticos: Esperanza y maximización (EM).

Básicos en inferencia multivariante

- Test de la razón de verosimilitudes.
- Tests para vectores de medias.
- Tests para matrices de covarianzas. ANOVA y MANOVA.

Análisis discriminante (AD)

- Objetivos y criterios de AD.
- Análisis discriminante en modelos gaussianos.
- Discriminante lineal de Fisher.

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases de prácticas con ordenador	26	1,04	
Clases de teoría	26	1,04	
Tipo: Autónomas			
Estudio y consultas	42	1,68	
Resolución y entrega de ejercicios	44	1,76	

Clases de teoría, donde se expondrán detalladamente y se discutirán los contenidos en base a ejemplos de aplicación convenientemente interpretados. Los resúmenes de los apuntes se colgarán en el Campus virtual. Se estimulará la revisión y ampliación de contenidos con la bibliografía básica y complementaria.

Clases prácticas pensadas para implementar con el lenguaje estadístico R. Los guiones de las prácticas, los enunciados de problemas y otro material auxiliar se pondrán a disposición del alumnado en el Campus Virtual. En las clases prácticas se propondrán ejercicios de ampliación para resolver autónomamente.

Se procurará la colaboración y participación de todos los estudiantes, sin ningún tipo de discriminación por razón de sexo u otra causa.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de tareas	0,2	4	0,16	CM11, CM12, KM16
Primer parcial (teoría i prácticas)	0,35	4	0,16	CM11, KM16
Segundo parcial (teoría y prácticas)	0,45	4	0,16	CM11, KM16

La nota de curso (NC) se calculará a partir de las entregas de ejercicios teóricos y de prácticas y de las notas de dos exámenes parciales (P1 y P2), tanto de teoría como de prácticas. Detalle de la nota de curso:

$$NC = 0,35 \cdot P1 + 0,45 \cdot P2 + 0,20 \cdot Lli$$

donde P1 y P2 son las notas del primer parcial y del segundo parcial y Lli es la nota de las entregas que se propongan y que no se podrán recuperar.

El aprobado por curso requiere que $NC \geq 5$ y que P1 y P2 sean ambas mayores que 3.5 (sobre 10). El examen final de recuperación (F) será una prueba de síntesis de todo el semestre con una parte teórica y una de computacional. Los alumnos pueden presentarse a la recuperación para mejorar su NC. No obstante, el profesor se reserva el derecho a decidir las matrículas d'honor, si las hubiere, antes del examen de recuperación. La nota final se calculará:

$$NF = \text{Max}(NC, 0,80 \cdot F + 0,20Lli)$$

Observación: Solo se podrán presentar a examen final aquellos estudiantes que hayan participado en 2/3 partes de las actividades de evaluación continuada.

Evaluación única

La evaluación única será una prueba de síntesis de las competencias de ambos parciales, en base a: (1) Un examen con cuestiones de teoría y de prácticas (peso: 50%). (2) Una prueba de prácticas delante del ordenador (peso: 40%). (3) La entrega de las tareas programadas que se indiquen, con la posibilidad de que el profesorado pida que el estudiante explique detalles de estas entregas (peso: 10%).

Bibliografía

Everitt, B., Hothorn, T. ; An introduction to Applied Multivariate Analysis with R. Springer, 2011.

Härdle, W., Simar, L.; Applied Multivariate Statistical Analysis. Springer,2007.

Peña, D.; Análisis de datos multivariantes. McGraw Hill, 2002.

Rencher, A., Christensen, W.; Methods of Multivariate Analysis. Wiley Series in Probability and Mathematical Statistics, 2012.

Wehrens, R. (2020). Chemometrics with R: Multivariate data analysis in the natural sciences and life sciences. Heidelberg: Springer. <https://link-springer-com.are.uab.cat/book/10.1007/978-3-662-62027-4>

Bibliografia complementària

Coghlan, A.; Little book of R for Multivariate Analysis.

<https://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/>

Cuadras, C.; Nuevos Métodos de Análisis Multivariante (web), 2014.

Greenacre, M.; La práctica del análisis de correspondencias. Fundacion BBA, 2003.

James, G., Witten, D., Hastie, T., Tibshirani, R.; An Introduction to Statistical Learning. Springer, 2014.

Mardia, K.V, Kent, J.T., Bibby, J.M.; Multivariate Analysis. Academic Press, 2003.

Rencher, A.; Multivariate Statistical Inference and Applications. John Wiley & Sons, 1998.

Software

R y RStudio

Lista de idiomas

La información sobre los idiomas de impartición de la docencia se puede consultar en el apartado de CONTENIDOS de la guía.

PROVISIONAL