Universitat Autònoma de Barcelona

2019/2020

Aprendizaje No Supervisado

Código: 104869
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503852 Estadística Aplicada	OB	2	2

Contacto

Nombre: Mercè Farré Cervelló

Correo electrónico: Merce.Farre@uab.cat

Uso de idiomas

Lengua vehicular mayoritaria: catalán (cat)

Algún grupo íntegramente en inglés: No

Algún grupo íntegramente en catalán: Sí

Algún grupo íntegramente en español: No

Prerequisitos

Es imprescindible un curso previo de Álgebra lineal, así como cursos básicos de Probabilidades y de Inferencia estadística. También se supone que sabe utilizar con agilidad el lenguaje R.

Objetivos y contextualización

La mayoría de datos recogidos son multivariantes, es decir, para una misma unidad experimental (que puede ser un objeto de naturaleza compleja) observamos de manera simultánea los valores de varias variables. El Análisis Multivariante trata los métodos que son más apropiados para describir, explorar y modelizar estos datos vectoriales, así como para aplicar la inferencia estadística. La necesidad de procesar gran cantidad de datos con muchas variables de naturaleza diversa, con la voluntad de reducir la información que no sea relevante y de descubrir patrones de asociación entre variables, así como el interés en la clasificación supervisada o autónoma de casos, han hecho desarrolló recientemente un gran número de procedimientos que se sitúan en el escenario multivariante. 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, al tiempo las limitaciones, los métodos, algunos de los cuales se basan en ideas heurísticas bastante intuitivas. El enfoque de la asignatura es eminentemente aplicado con respecto 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 los resultados.

Competencias

- Analizar datos mediante la aplicación de métodos y técnicas estadísticas, trabajando con datos de diversas tipologías.
- Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otras personas.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- Resumir y descubrir patrones de comportamiento en la exploración de los datos.
- Seleccionar las fuentes y técnicas de adquisición y gestión de datos adecuadas para su tratamiento estadístico.
- Seleccionar y aplicar procedimientos más apropiados para la modelización estadística y el análisis de datos complejos.
- Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
- Utilizar eficazmente la bibliografía y los recursos electrónicos para obtener información.

Resultados de aprendizaje

1. Analizar datos mediante metodología de aprendizaje automático.
2. Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otros.
3. Caracterizar grupos homogéneos de individuos mediante análisis multivariante.
4. Describir las ventajas e inconvenientes de los métodos algorítmicos frente a los métodos convencionales de la inferencia estadística.
5. Emplear gráficos de resumen de datos multivariados o más complejos.
6. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
7. Identificar las suposiciones estadísticas asociadas a cada procedimiento avanzado.
8. Identificar, emplear e interpretar los criterios para evaluar el grado de cumplimiento de los requisitos necesarios para aplicar cada procedimiento avanzado.
9. Obtener y gestionar bases de datos complejas para su posterior análisis.
10. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
11. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
12. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
13. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
14. Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
15. Utilizar eficazmente bibliografía y recursos electrónicos para obtener información.

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 correlaciónes. 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.

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 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 procedimientos de conglomeración.
- 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) y otros métodos supervisados

- Objetivos and criterios de AD.
- Análisis discriminante en modelos gaussianos.
- Discriminante lineal de Fisher.
- Regresión por mínimos cuadrados parciales (PLS). Otros métodos.

Metodología

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.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases de prácticas con ordenador	26	1,04	1, 4, 5, 7, 8, 9, 12, 13, 14, 15
Clases de teoría	26	1,04	1, 2, 6, 3, 4, 5, 7, 8, 9
Tipo: Autónomas			
Estudio y consultas	42	1,68	2, 4, 7, 8, 10, 15
Resolución y entrega de ejercicios	44	1,76	1, 3, 4, 5, 9, 10, 11, 12, 13, 14, 15

Evaluación

La nota de curso (NC) se calculará a partir de las entregas de exercicios teóricos y de prácticas (10%) y de las notas de dos exámenes parciales (P1 y P2), tanto de teoría (45%) como de prácticas (45%). Detalle de la nota de curso:

NC = 0,2•P1t + 0,2•P2p + 0,25•P2t + 0,25•P2p + 0,10•Lli

donde P1t y P1p son las notas de teoría y de prácticas del primer parcial, P2t y P2p son análogas para el segundo parcial, y Lli es la nota de las entregas que se propongan y que no se podrán recuparar.

Si NC es igual o superior a 5, el alumno superará la materia sin necesidad de acudir a la recuperación. El examen final de recuperación (F) será una prueba de síntesis de todo el semestre con una parte teórica y una de computacional que promediarán. Los alumnes 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=Max(NC , 0,90•F + 0,10Lli)

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

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de tareas	0,1	2	0,08	2, 6, 9, 10, 11, 12, 13, 14, 15
Primer parcial	0,4	5	0,2	1, 5, 7, 8, 10, 11
Segundo parcial	0,5	5	0,2	1, 3, 4, 7, 8, 9, 10, 11, 12, 15

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.

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 pràctica 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.