

Titulación	Tipo	Curso
4313136 Modelización para la Ciencia y la Ingeniería / Modelling for Science and Engineering	OT	0

Contacto

Nombre: Amanda Fernandez Fontelo

Correo electrónico: amanda.fernandez@uab.cat

Equipo docente

Sundus Zafar

Carles Barril Basil

Idiomas de los grupos

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

Prerrequisitos

Los estudiantes deberían tener conocimientos básicos de álgebra lineal, inferencia estadística, y modelos lineales.

La experiencia previa con R y Python es recomendable.

Objetivos y contextualización

Hoy en día se están generando enormes cantidades de datos en muchos campos, y el propósito de este curso es aprender a extraer información a partir de estos datos. El objetivo de este curso es para aprender y aplicar varios métodos matemáticos y estadísticos para el descubrimiento de patrones en conjuntos de datos. Cuando se trabaja con datasets grandes, los procedimientos matemáticos tienen que ser escalables, así que nos ocuparemos de métodos que puedan ser escalados y/o paralelizados.

Competencias

- "Aplicar el pensamiento lógico/matemático: el proceso analítico a partir de principios generales para llegar a casos particulares; y el sintético, para a partir de diversos ejemplos extraer una regla general."
- Analizar, sintetizar, organizar y planificar proyectos de su campo de estudio.

- Aplicar las técnicas de resolución de los modelos matemáticos y sus problemas reales de implementación.
- Concebir y diseñar soluciones eficientes, aplicando técnicas computacionales, que permitan resolver modelos matemáticos de sistemas complejos.
- Extraer de un problema complejo la dificultad principal, separada de otras cuestiones de índole menor.
- Formular, analizar y validar modelos matemáticos de problemas prácticos de distintos campos.
- Resolver problemas complejos aplicando los conocimientos adquiridos a ámbitos distintos de los originales

Resultados de aprendizaje

1. "Aplicar el pensamiento lógico/matemático: el proceso analítico a partir de principios generales para llegar a casos particulares; y el sintético, para a partir de diversos ejemplos extraer una regla general."
2. Analizar, sintetizar, organizar y planificar proyectos de su campo de estudio.
3. Aplicar técnicas de Estadística Bayesiana para predecir el comportamiento futuro de ciertos fenómenos.
4. Extraer de un problema complejo la dificultad principal, separada de otras cuestiones de índole menor.
5. Identificar fenómenos reales como modelos de procesos estocásticos y saber extraer de aquí información nueva para interpretar la realidad
6. Resolver problemas complejos aplicando los conocimientos adquiridos a ámbitos distintos de los originales
7. Resolver problemas reales de análisis de datos identificándolos adecuadamente desde la óptica de la Estadística Bayesiana.
8. Usar paquetes estadísticos y métodos bayesianos apropiados para solucionar problemas concretos.

Contenido

Text Mining

- Fundamentals of Text Mining - From text to numbers
- Data cleaning
- Tokenization
- Stemming
- Lemmatization
- POS, NER
- Data chunking

Statistics

- Summarising the information from large data sets:
 - The principle of sufficiency and sufficient statistics.
 - Applications to classical and generalised linear models.
 - The Biglm package.
- Problems of likelihood estimation problems for large data sets:
 - The method of "Divide and Recombine" and generalisations.
 - The idea of segmentation, analysis of chunks of data, and methods based on meta-analysis.
 - Applications to linear and generalised linear models.
- The problem of multiple testing and false discovery rate:
 - The idea of knockoff variables.
- Functional Data Analysis:
 - Observed functional data and its computational representation.
 - Descriptive statistics and dimensionality reduction.
 - Depth measures for functional data.
 - Functional linear models and classification techniques.

Deep Learning

- Fully Connected Neural Networks.
- Convolutional Neural Networks.
- Recurrent Neural Networks
- Keras and Tensorflow.

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases teóricas	38	1,52	2, 5
Ejercicios (problemas y programación)	36	1,44	1, 8
Tipo: Autónomas			
Estudio autónomo	20	0,8	5
Homework	44	1,76	2, 1, 5, 4, 6, 8

Ver la versión de la guía en inglés.

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
Deep Learning	0.25	3	0,12	2, 1, 5, 4, 6, 8
First Homework Statistics	0.25	3	0,12	2, 1, 5, 4, 6, 7, 8
Homework Text Mining	0.25	3	0,12	2, 1, 4, 6, 8
Second Homework Statistics	0.25	3	0,12	2, 3, 1, 5, 4, 6, 7

Ver la versión de la guía en inglés.

Bibliografía

Referències bàsiques

- B. Efron, T. Hastie, *Computer Age Statistical Inference*, Cambridge University Press (2016) (5th Ed 2017) <https://web.stanford.edu/~hastie/CASI/index.html>
- G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning (with applications in R)*. Springer, 2013.
- D. Skillicorn, "Understanding Complex Data. Data Mining with Matrix Decomposition". Chapman&Hall, 2007.

Referències Complementàries

- B. Everitt and T. Hothorn, "An introduction to Applied Multivariate Analysis with R". Springer, 2011.
- B. Everitt, "An R and S+ Companion to Multivariate Analysis", Springer, 2005.
- J. Faraway, "Extending de Linear Model with R", Chapman & Hall, Miami, 2006.
- J. Faraway, "Linear Models with R", Chapman & Hall, Boca Raton, 2005.
- W. Härdle and L. Simar, "Applied Multivariate Statistical Analysis". Springer. 2007.
- B. Ripley, "Pattern Recognition and Neural Networks". Cambridge University Press, 2002.
- L. Torgo. "Data Mining with R. Learning with Case Studies". Chapman & Hall, Miami. 2010
- W Venables, B Ripley, "Modern Applied Statistics with S-PLUS", Springer, New York.
- Collins FS and Varmus H, "A new initiative on precision medicine". N Engl J Med. 2015 Feb 26;372(9):793-5 .
- Jensen A.B. et al, "Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients". Nat Commun 2014 Jun 24; 5:4022.
- J.D. Jobson, "Applied Multivariate Analysis". Vol I i II. Springer, 1992.
- R. Johnson and D.W. Wichern, "Applied Multivariate Statistical Analysis". Pearson Education International, 2007.
- P.Y.Lum et al., "Extracting insights from the shape of complex data using topology". Sci. Rep. 3, 1236; DOI:10.1038/srep01236 (2013).
- A. Rencher, "Methods of Multivariate Analysis". Wiley Series in Probability and Mathematical Statistics, 2002.
- G. Singh, F. Mévoli, G. Carlsson, "Topological methods for the analysis of High dimensional data sets and 3D object recognition". Eurographic Symp. on Point-Based Graphics, 2007
- P. Kokoszka, M. Reimherr, *Introduction to Functional Data Analysis*. CRC Press.(2017).
- Ramsay, J. , B. W. Silverman, *Functional Data Analysis Springer* (2nd Ed. 2005).

Software

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Python

Lista de idiomas

Nombre	Grupo	Idioma	Semestre	Turno
(TEm) Teoría (máster)	1	Inglés	segundo cuatrimestre	tarde