

Titulació	Tipus	Curs
4313136 Modelització per a la Ciència i l'Enginyeria / Modelling for Science and Engineering	OT	0

Professor/a de contacte

Nom: Amanda Fernandez Fontelo

Correu electrònic: amanda.fernandez@uab.cat

Equip docent

Sundus Zafar

Carles Barril Basil

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

Els estudiants haurien de tenir coneixements bàsics d'àlgebra lineal, probabilitat, inferència estadística i models lineals.

L'experiència prèvia amb R i Python és recomanable.

Objectius

Avui dia, quantitats enormes de dades estan sent generades a molts camps, i el propòsit d'aquest curs és per aprendre com extreure informació a partir d'aquestes dades. Aquest curs es centra doncs en l'aprenentatge i l'aplicació de diversos mètodes matemàtics i estadístics per a la descoberta de patrons a conjunts de dades.

Competències

- "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."
- Analitzar, sintetitzar, organitzar i planificar projectes del seu camp d'estudi.
- Aplicar les tècniques de resolució dels models matemàtics i els seus problemes reals d'implementació.
- Concebre i dissenyar solucions eficients, aplicant tècniques computacionals, que permetin resoldre models matemàtics de sistemes complexos.
- Extreure d'un problema complex la dificultat principal, separada d'altres qüestions d'índole menor.

- Formular, analitzar i validar models matemàtics de problemes pràctics de diferents camps.
- Resoldre problemes complexos aplicant els coneixements adquirits a àmbits diferents dels originals

Resultats d'aprenentatge

1. "Aplicar el pensament lògic/matemàtic: el procés analític a partir de principis generals per arribar a casos particulars; i el sintètic, para a partir de diversos exemples extreure una regla general."
2. Analitzar, sintetitzar, organitzar i planificar projectes del seu camp d'estudi.
3. Aplicar tècniques d'Estadística Bayesiana per predir el comportament futur de certs fenòmens.
4. Extreure d'un problema complex la dificultat principal, separada d'altres qüestions d'indole menor.
5. Identificar fenòmens reals com a models de processos estocàstics i saber extreure d'aquí informació nova per interpretar la realitat
6. Resoldre problemes complexos aplicant els coneixements adquirits a àmbits diferents dels originals
7. Resoldre problemes reals d'anàlisis de dades identificant-los adequadament des de l'òptica de l'Estadística *Bayesiana.
8. Usar paquets estadístics i mètodes bayesians apropiats per solucionar problemes concrets.

Continguts

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.

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Clases de Teoria	38	1,52	2, 5
Exercicis (problemes i ordinador)	36	1,44	1, 8
Tipus: Autònomes			
Estudi autònom	20	0,8	5
Homework	44	1,76	2, 1, 5, 4, 6, 8

Veure la versió de la guia en anglès.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Deep Learning	0,25	3	0,12	2, 1, 5, 4, 6, 8
First Homework Statistics	0,25	3	0,12	2, 3, 1, 5, 4, 6, 7
Homework Text Mining	0,25	3	0,12	2, 1, 4, 6, 8
Second Homework Statistics	0,25	3	0,12	2, 1, 5, 4, 6, 7, 8

Homework: Presentació escrita dels exercicis proposats.

Projecte final: Els estudiants hauran de desenvolupar un projecte segons les indicacions publicades al Campus Virtual.

Les dates previstes seran anunciades durant el curs i seran estrictes.

Bibliografia

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. Mémoli, 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).

Programari

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

Llista d'idiomes

Nom	Grup	Idioma	Semestre	Torn
(TEm) Teoria (màster)	1	Anglès	segon quadrimestre	tarda