

**FITXA IDENTIFICATIVA****Dades de l'Assignatura**

Codi	44276
Nom	Anàlisi exploradora de dades
Cicle	Màster
Crèdits ECTS	3.0
Curs acadèmic	2021 - 2022

Titulació/titulacions

Titulació	Centre	Curs	Període
2199 - M.U. en Enginyeria Electrònica	Escola Tècnica Superior d'Enginyeria	1	Segon quadrimestre
3131 - Enginyeria Electrònica	Escola de Doctorat	0	Primer quadrimestre

Matèries

Titulació	Matèria	Caràcter
2199 - M.U. en Enginyeria Electrònica	1 - Tractament digital de senyals	Obligatòria
3131 - Enginyeria Electrònica	1 - Complementos de Formació	Optativa

Coordinació

Nom	Departament
MARTIN GUERRERO, JOSE DAVID	242 - Enginyeria Electrònica
SERRANO LOPEZ, ANTONIO JOSE	242 - Enginyeria Electrònica

RESUM

Esta assignatura té com a objectiu principal descriure les tècniques més importants d'Anàlisi Exploratòria de Dades que permeten extraure coneixement d'un problema a través de l'estudi estadístic de les dades adquirides.

L'assignatura està formada per cinc temes. En el primer d'ells es presenten i repassen algunes de les eines matemàtiques bàsiques per a l'anàlisi de dades, com la Probabilitat, l'Estadística o l'Àlgebra Lineal. A continuació, en el segon tema, es descriuen alguns conceptes bàsics, que permeten definir les característiques del tipus de problema a resoldre i la tècnica que s'utilitzarà per a això (nombre de patrons, dades atípiques, valors perduts, tipus d'aprenentatge a utilitzar, etc.). Posteriorment, en el tercer tema, es presenten les tècniques per a dur a terme les primeres exploracions sobre les dades per a tindre una descripció succinta del tipus d'informació que emmagatzemen, destacant l'estudi d'algoritmes d'agrupament (*clustering*) així com les tècniques d'extracció i selecció de variables (característiques). Els



dos últims temes descriuen els models de classificació i regressió (lineals i no lineals, respectivament) més àmpliament utilitzats en l'Anàlisi Exploratòria de Dades.

A banda dels continguts purament teòrics l'assignatura proveirà a l'alumne dels coneixements generals necessaris per a la resolució de problemes d'Enginyeria.

Esta és una assignatura de caràcter obligatori, que s'impartix en el primer quadrimestre del Màster en Enginyeria Electrònica. La càrrega lectiva total és de 3 ECTS. La càrrega de treball per a l'alumne és de 75 hores al llarg del quadrimestre, de les quals 30 són presencials i 45 són de treball individual.

Per a les sessions de laboratori, utilitzarem el paquet de codi obert R per a l'anàlisi de dades i la RStudio i el programari Shiny Server per a la representació interactiva i visualització dels resultats

CONEIXEMENTS PREVIS

Relació amb altres assignatures de la mateixa titulació

No heu especificat les restriccions de matrícula amb altres assignatures del pla d'estudis.

Altres tipus de requisits

És recomanable que l'alumne tinga una base matemàtica mínima d'Àlgebra, Càlcul, Estadística i Probabilitat. Si no la posseïx, se li facilitarà una sèrie de tutorials perquè s'adapte al curs sense problemes.

COMPETÈNCIES

2199 - M.U. en Enginyeria Electrònica

- Que els estudiants sàpiguem aplicar els coneixements adquirits i la seua capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seua àrea d'estudi.
- Que els estudiants siguen capaços d'integrar coneixements i afrontar la complexitat de formular judicis a partir d'una informació que, sent incompleta o limitada, incloga reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.
- Que els estudiants sàpiguem comunicar les conclusions (i els coneixements i les raons últimes que les sustenten) a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.
- Que els estudiants posseïsquen les habilitats d'aprenentatge que els permeten continuar estudiant d'una forma que haurà de ser en gran manera autodirigida o autònoma.
- Considerar el context econòmic i social en les solucions d'enginyeria sent conscient de la diversitat i la multiculturalitat, i garantint la sostenibilitat i el respecte dels drets humans i de la igualtat home-dona.



- Dissenyar un sistema, component o procés que complisca unes especificacions des de diferents punts de vista: electrònic, econòmic, social, ètic i mediambiental.
- Demostrar una comprensió sistemàtica d'un camp d'estudi i el domini de les habilitats.
- Realitzar una anàlisi crítica, avaluació i síntesi d'idees noves i complexes.
- Ser capaç de fomentar, en contextos acadèmics i professionals, l'avanç tecnològic, social o cultural dins una societat basada en el coneixement.
- Capacitat per projectar, calcular i dissenyar productes, processos i instal·lacions en tots els àmbits de l'enginyeria electrònica i, en particular, els de tractament del senyal, sistemes digitals i de comunicacions i electrònica industrial.
- Capacitat per a la modelització matemàtica, càlcul i simulació en tots els àmbits relacionats amb l'enginyeria electrònica i camps multidisciplinaris afins. En especial, els de tractament del senyal, sistemes digitals i de comunicacions i electrònica industrial.
- Conèixer les tècniques avançades d'anàlisi de dades.
- Capacitat per analitzar, especificar i dissenyar sistemes de tractament digital de senyals, des de la seua concepció fins a la seua implementació en sistemes maquinari de temps real.

RESULTATS DE L'APRENTATGE

Una vegada s'haja desenvolupat esta assignatura l'alumne haurà de saber afrontar amb garanties un problema d'anàlisi de dades des d'un principi, sent capaç de filtrar la informació útil de la què no ho és (soroll), i tenint capacitat per a trobar els perfils típics de comportament que s'observen en el conjunt de dades.

Així mateix, tindrà els suficients elements de juí i la capacitat d'establir els avantatges i inconvenients de diferents alternatives en el processat de dades. Podrà utilitzar aproximacions lineals o no lineals depenent del problema a resoldre.

DESCRIPCIÓ DE CONTINGUTS

1. Fonaments bàsics per a l'anàlisi exploradora de dades

TEMA 1.

- 1.1 Probabilitat (axiomes; prob. Condicional i conjunta)
- 1.2 Teorema de Bayes. Aplicacions
- 1.3 Variable Aleatòria. Distribucions típiques; moments
- 1.4 Estadística. Contrastos d'hipòtesis
- 1.5 Repàs d'Àlgebra lineal
- 1.6 Optimització de funcions

PRÀCTICA: Aplicació d'estadística i contrastos d'hipòtesis per a l'anàlisi de dades.



2. Introducció a l'anàlisi de dades

TEMA 2.

- 2.1 Variables: tipus. Patrons
- 2.3 Tipus de problemes a resoldre
- 2.4 Aprenentatge. Tipus

3. Anàlisi exploradora descriptiva

TEMA 3.

- 3.1 Adquisició i neteja de les dades
- 3.2 Caracterització de les variables (estadístics i gràfiques)
- 3.3 Transformacions de les dades
- 3.4 Selecció i extracció de característiques
- 3.5 Algorismes d'agrupament (HCM, FCM i SOM)

PRÀCTICA: Preprocessat de dades.

Codificació, normalització, eliminació de dades atípiques, processat de dades perduts, obtenció de perfils típics amb mètodes d'agrupament.

4. Models lineals

TEMA 4.

- 4.1 Descripció de models lineals
- 4.2 Obtenció de paràmetres: equacions normals. Regularització
- 4.3 Validació de models

PRÀCTICA: Aplicació de models lineals per al tractament de dades

Regressió lineal de mínims quadrats per al modelat de funcions. Regularització. Regressió logística aplicada a problemes de classificació. Regressió amb funcions de cost robustes.

5. Models no lineals

TEMA 5.

- 5.1 Arbres (regressió i classificació)
- 5.2 Xarxes Neuronals Artificials
- 5.3 Màquines de Vectors Suport
- 5.4 Extracció de regles

PRÀCTICA: Aplicació de models no lineals al tractament de dades (I)

Aplicació de models no lineals vistos en teoria; s'usaran conjunts de dades semblants a la pràctica anterior per a poder comparar el seu funcionament.

PRÀCTICA: Aplicació de models no lineals al tractament de dades (II)

Atés que es mostren diversos models no lineals en teoria es planteja una segona pràctica per a analitzar



els que no s'hagen vist en a pràctica anterior.

VOLUM DE TREBALL

ACTIVITAT	Hores	% Presencial
Classes de teoria	15,00	100
Pràctiques en laboratori	15,00	100
Estudi i treball autònom	10,00	0
Lectures de material complementari	5,00	0
Preparació d'activitats d'avaluació	5,00	0
Preparació de classes de teoria	5,00	0
Preparació de classes pràctiques i de problemes	10,00	0
Resolució de casos pràctics	10,00	0
TOTAL	75,00	

METODOLOGIA DOCENT

Les metodologies docents a emprar en el desenvolupament de l'assignatura són les següents:

a) Activitats teòriques.

Desenvolupament expositiu de la matèria amb la participació de l'estudiant en la resolució de qüestions puntuals.

b) Activitats pràctiques.

Resolució de casos pràctics.

c) Treball personal de l'estudiant.

Descripció: Realització fora de l'aula de qüestions i problemes, així com la preparació de classes i exàmens (estudi). Aquesta tasca es realitzarà de manera individual i intenta potenciar el treball autònom.

S'utilitzaran les plataformes d'e-learning (Aula Virtual) com a suport de comunicació amb els estudiants. A través d'ella es tindrà accés al material didàctic utilitzat en classe, així com els problemes i exercicis a resoldre.

AVALUACIÓ

L'avaluació de l'assignatura es durà a terme mitjançant la realització d'una prova de coneixement que es realitzarà en forma d'examen individual o de treball per grups sobre els continguts de l'assignatura.



REFERÈNCIES

Bàsiques

- Advances in knowledge discovery and data mining. Edited by Usama M. Fayyad [et al.]. MIT Press, 1996.
- Data mining for scientific and engineering applications. Edited by Robert L. Grossman [et al.]. Kluwer, 2001.
- Análisis de datos experimentales. Emilio Soria, José D. Martín, Antonio J. Serrano, Daniel Aguado. Universidad Politécnica de Valencia, 2007.
- Machine Learning. Ethem Alpaydin, MIT Press, 2009.

Complementàries

- Neural Networks and Learning Machines. Simon Haykin. Pearson Education, 2009.
- Time Series Prediction: Forecasting the Future and Understanding the Past: Proceedings of the NATO by Andreas Weigend y Neil Gershenfeld (Editores). Addison-Wesley, 1993.
- Pattern Classification (2nd Edition). Richard O. Duda, Peter E. d G. Stork. Wiley-Interscience, 2n edition, November 2000.
- Kernel Methods for Pattern Analysis. John Shawe-Taylor and Nello Cristianini. Cambridge University Press, New York, NY, USA, 2004.
- The elements of statistical learning: data mining, inference, and prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer, 2001.

ADDENDA COVID-19

Aquesta addenda només s'activarà si la situació sanitària ho requereix i previ acord del Consell de Govern