

Guia docent

270962 - AMA - Anàlisi Multivariant Avançat

Última modificació: 23/11/2023

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.

Titulació: MÀSTER UNIVERSITARI EN CIÈNCIA DE DADES (Pla 2021). (Assignatura optativa).

Curs: 2023 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: PEDRO FRANCISCO DELICADO USEROS

Altres: Primer quadrimestre:
PEDRO FRANCISCO DELICADO USEROS - 10

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CE13. Identificar les principals amenaces en l'àmbit de l'ètica i la privacitat de dades en un projecte de ciència de dades (tant en l'aspecte de gestió com d'anàlisi de dades) i desenvolupar i implantar mesures adequades per esmorteir aquestes amenaces.

CE3. Aplicar mètodes d'integració de dades per donar solució a problemes de ciència de dades en entorns heterogenis

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE8. Extreure informació de dades estructurades i no estructurades, tenint en compte la naturalesa multivariant de les mateixes.

Genèriques:

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

Transversals:

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

Bàsiques:

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

METODOLOGIES DOCENTS

Hi ha dues sessions setmanals de 2 hores.

Les primeres tres hores es dediquen a l'exposició dels temes teòrics per part del professor.

L'última hora es dedica a implementar aquests continguts: Cada alumne té el seu portàtil a classe i realitza les tasques proposades pel professor.

Cada setmana finalitza amb una tasca per als alumnes que s'ha de lliurar en 7 dies. El programari utilitzat serà principalment R.



OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Conèixer l'estructura dels principals problemes d'aprenentatge no supervisat.
2. Apreneu diferents mètodes per a la reducció de la dimensionalitat quan no es compleixen els supòsits estàndard de l'anàlisi multivariant clàssic
3. Apreneu a combinar tècniques de reducció de dimensionalitat amb algorismes de predicció
4. Al final de l'assignatura l'estudiant serà capaç de proposar, estimar, interpretar i validar versions no paramètriques de models de regressió lineal i models lineals generalitzats.
5. Al final del curs l'estudiant sabrà triar correctament els paràmetres de suavització que en els models de regressió no paramètrics controlen el compromís entre un bon ajust a la mostra observada i una bona generalització.
6. Al final del curs, l'estudiant serà conscient de la necessitat de dotar d'interpretabilitat als algorismes d'aprenentatge automàtic, coneixerà les tècniques d'interpretabilitat més habituals, sabrà classificar-les i quines relacions hi ha entre ells, i sabrà com utilitzar-los en R i/o Python.

HORES TOTALS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup gran	54,0	36.00
Hores aprenentatge autònom	96,0	64.00

Dedicació total: 150 h

CONTINGUTS

Aprenentatge no supervisat mitjançant anàlisi multivariada avançada

Models de regressió no paramètrics

Aprenentatge automàtic interpretable

ACTIVITATS

Aprentatge no supervisat mitjançant l'anàlisi multivariant avançada

Objectius específics:

1, 2, 3

Competències relacionades:

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents
CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE8. Extreure informació de dades estructurades i no estructurades, tenint en compte la naturalesa multivariant de les mateixes.

CE3. Aplicar mètodes d'integració de dades per donar solució a problemes de ciència de dades en entorns heterogenis

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

Dedicació: 52h 18m

Grup gran/Teoria: 18h

Aprentatge autònom: 34h 18m

Models de regressió no paramètrica

Objectius específics:

4, 5

Competències relacionades:

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents
CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE8. Extreure informació de dades estructurades i no estructurades, tenint en compte la naturalesa multivariant de les mateixes.

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

Dedicació: 54h 18m

Grup gran/Teoria: 20h

Aprentatge autònom: 34h 18m



Interpretabilitat en aprenentatge automàtic

Objectius específics:

6

Competències relacionades:

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents
CE13. Identificar les principals amenaces en l'àmbit de l'ètica i la privacitat de dades en un projecte de ciència de dades (tant en l'aspecte de gestió com d'anàlisi de dades) i desenvolupar i implantar mesures adequades per esmorteir aquestes amenaces.

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE8. Extreure informació de dades estructurades i no estructurades, tenint en compte la naturalesa multivariant de les mateixes.

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

Dedicació: 43h 18m

Grup gran/Teoria: 16h

Aprenentatge autònom: 27h 18m

SISTEMA DE QUALIFICACIÓ

Al llarg del curs s'assignaran tasques com a deures. Les notes dels deures valdran el 40% de la nota del curs.

Hi haurà un examen al final del quadrimestre i s'avaluarà l'assimilació dels conceptes bàsics de tota l'assignatura. L'examen final tindrà una primera part teòrica breu (llibres tancats) i una segona part pràctica més llarga (llibres oberts, a realitzar pels alumnes amb el seu propi ordinador portàtil, amb una estructura semblant als deures).

Nota del curs = 0.4 * Nota Hwk + 0.6 * Nota de l'examen

BIBLIOGRAFIA

Bàsica:

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York [etc.]: Springer, cop. 2009. ISBN 9780387952840.
- Wasserman, Larry. All of nonparametric statistics. Pittsburgh: Springer, cop. 2010. ISBN 9781441920447.
- Wood, Simon N. Generalized additive models : an introduction with R. 2nd ed. Boca Raton: CRC Press/Taylor & Francis Group, [2017]. ISBN 9781498728331.
- Biecek, P.; Burzykowski, T. Explanatory model analysis: explore, explain and examine predictive models. Oxford: Oxford University Press, 2018. ISBN 9780367135591.
- Fan, Jianqing; Li, Runze; Zhang, Cun-hui; Zou, Hui. Statistical foundations of data science. Boca Raton: Oxon : CRC Press, 2020. ISBN 9781466510845.