

Aprenentatge Automàtic 1

Codi: 104870

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2503852 Estadística Aplicada	OB	3	1

La metodologia docent i l'avaluació proposades a la guia poden experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries.

Professor/a de contacte

Nom: Juan Ramón González Ruíz

Correu electrònic: JuanRamon.Gonzalez@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: català (cat)

Grup íntegre en anglès: No

Grup íntegre en català: Sí

Grup íntegre en espanyol: No

Equip docent

Joan Valls Marsal

Prerequisits

Aquesta assignatura suposa que l'alumne ha assolit els coneixements que s'imparteixen en diferents assignatures sobre els següents temes:

- Càlcul en diverses variables.
- Probabilitat.
- Models lineals.
- Programació en R.

Objectius

Aquesta assignatura pretén familiaritzar l'alumne amb diferents mètodes d'aprenentatge automàtic aplicant el punt de vista utilitzat quan es disposa de grans quantitats de dades.

Competències

- Analitzar dades mitjançant l'aplicació de mètodes i tècniques estadístiques, treballant amb dades de diverses tipologies.
- Aplicar l'esperit crític i el rigor per validar o refutar arguments tant propis com d'altres persones.
- Avaluar de manera crítica i amb criteris de qualitat el treball realitzat.
- Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia.
- Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.

- Que els estudiants sàpiguin aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
- Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes destacats d'indole social, científica o ètica.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.
- Seleccionar els models o tècniques estadístiques per aplicar-los en estudis i problemes reals, així com conèixer-ne les eines de validació.
- Seleccionar i aplicar procediments més apropiats per a la modelització estadística i l'anàlisi de dades complexes.
- Seleccionar les fonts i tècniques d'adquisició i gestió de dades adequades per a fer-ne un tractament estadístic.
- Treballar cooperativament en un context multidisciplinari assumint i respectant el rol dels diferents membres de l'equip.
- Utilitzar correctament un ampli espectre del programari i llenguatges de programació estadístiques, escollint el més apropiat per a cada anàlisi i ser capaç d'adaptar-lo a noves necessitats.
- Utilitzar eficaçment la bibliografia i els recursos electrònics per obtenir informació.

Resultats d'aprenentatge

1. Analitzar dades mitjançant la metodologia d'aprenentatge automàtic.
2. Aplicar l'esperit crític i el rigor per validar o refutar arguments, tant propis com d'altres.
3. Avaluar de manera crítica i amb criteris de qualitat la feina feta.
4. Caracteritzar grups homogenis d'individus mitjançant una anàlisi multivariant.
5. Descobrir comportaments i tipologies d'individus mitjançant tècniques de mineria de dades.
6. Descriure els avantatges i els inconvenients dels mètodes algorítmics vers els mètodes convencionals de la inferència estadística.
7. Fer servir gràfics de resum de dades multivariades o més complexes.
8. Identificar les suposicions estadístiques associades a cada procediment avançat.
9. Identificar, fer servir i interpretar els criteris per avaluar el grau de compliment dels requisits necessaris per aplicar cada procediment avançat.
10. Implementar programes en llenguatges adequats per a la mineria de dades.
11. Obtenir i gestionar bases de dades complexes per analitzar-les posteriorment.
12. Projectar un estudi partint de metodologies multivariants o de mineria de dades per resoldre un problema contextualitzat en la realitat experimental.
13. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia.
14. Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
15. Que els estudiants sàpiguin aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
16. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes destacats d'indole social, científica o ètica.
17. Treballar cooperativament en un context multidisciplinari assumint i respectant el rol dels diferents membres de l'equip.
18. Utilitzar eficaçment bibliografia i recursos electrònics per obtenir informació.
19. Utilitzar mètodes de mineria de dades per validar i comparar possibles models.

Continguts

Aquests són els continguts de l'assignatura*

- Introducció a Tidyverse
- Introducció a l'aprenentatge automàtic

- Regressió lineal i logística
- Passos previs a la creació d'un model predictiu i mesures de validació
- Mètodes d'aprenentatge automàtic
 - Anàlisi mitjançant arbres de classificació
 - K-veïns més propers
 - Random Forest
 - Boosting
- Mètodes d'aprenentatge per a dades $n \ll p$
 - Mètodes de penalització (shrinkage)
 - Mètodes de regularització
 - La llibreria 'caret'
- Mètodes d'aprenentatge per a dades grans (big data)
 - XGBoost
 - Lasso
 - La llibreria 'H2O'

*Llevat que les restriccions imposades per les autoritats sanitàries obliguin a una prioritització o reducció d'aquests continguts.

Metodologia

L'assignatura té programades, a la setmana, dues hores de teoria i dues hores de pràctiques.

- Classe de teoria: es defineixen i s'expliquen els diferents mètodes amb les seves característiques particulars i es mostren exemples concrets.

- Classe de pràctiques: es treballen els mètodes explicats a classe de teoria amb diversos conjunts de dades utilitzant el llenguatge de programació R.

Es considera que, per a cada hora de teoria i pràctiques, l'alumne haurà de dedicar una hora addicional a la preparació i/o finalització de la sessió. Adicionalment realitzarà preguntes d'autoevaluació al Moodle per consolidar els coneixements adquirits a classe.

NOTA: La metodologia docent proposada pot experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Sessions de pràctiques	50	2	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 19
Tipus: Supervisades			
Sessions de teoria	50	2	1, 2, 4, 5, 6, 7, 8, 9, 12, 13
Tipus: Autònomes			
Treball setmanal + autoevaluació	50	2	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

Avaluació

L'avaluació de l'assignatura es durà a terme amb un examen (final), treballs setmanals i preguntes d'autoevaluació. La nota final es calcularà amb la fórmula:

$$NF = 0,3 * NE + 0,5 * NP + 0,2 * NA$$

on NP és la nota promig dels treball semanals, NA es la nota promig de les preguntes d'autoevaluació i NE es la nota del examen que com a mínim ha de ser un 5.

A final de curs hi haurà una prova de recuperació per a aquells alumnes tals que NE és menor que 5 i NF menor que 5. En aquest cas, la nota final es calcularà amb la fórmula:

$$NF = 0,5 * NR + 0,5 * NP$$

on NR és la nota de l'examen de recuperació.

NOTA: L'avaluació proposada pot experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Autoevaluació	20%	0	0	1, 3, 5, 8, 9, 13, 15, 19
Examen final	30%	0	0	2, 4, 6, 7, 8, 9, 13, 15, 18
Pràctiques + autoevaluació	50%	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

Bibliografia

Bibliografia bàsica:

- An Introduction to Statistical Learning with Applications in R - Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

Bibliografia complementària:

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction - Trevor Hastie, Robert Tibshirani and Jerome Friedman

- Data Science from Scratch - Joel Grus

- Computer Age Statistical Inference: Algorithms, Evidence and Data Science - Trevor Hastie and Bradley Efron