

Modelització de Dades Complexes

Codi: 104864
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2503852 Estadística Aplicada	OB	3	2

Professor/a de contacte

Nom: Rosario Delgado de la Torre

Correu electrònic: rosario.delgado@uab.cat

Idiomes dels grups

Podeu accedir-hi des d'aquest [enllaç](#). Per consultar l'idioma us caldrà introduir el CODI de l'assignatura. Tingueu en compte que la informació és provisional fins a 30 de novembre de 2023.

Equip docent

Rosario Delgado de la Torre

Prerequisits

Es suposa que l'alumne que cursa aquesta assignatura ha adquirit les competències pròpies de les assignatures de

- Càlcul 1,
- Eines informàtiques per a l'Estadística i Introducció a la Programació,
- Introducció a la Probabilitat i Inferència Estadística 1, i
- Aprenentatge Automàtic 1.

Caldrà un bon nivell i pràctica en la programació amb R.

Objectius

Aprendre què són i com es fan servir les Xarxes Bayesianes: són un model probabilístic que es fa servir a l'Aprenentatge Automàtic Supervisat i que descriu les relacions de tipus probabilístic entre les variables que afecten determinat fenomen d'interès (que pot ser un sistema complex) i es poden fer servir com a classificadors.

Entendre com les Xarxes Bayesianes es fan servir per avaluar i quantificar riscos, entre altres aplicacions.

Conèixer diferents metodologies que s'hauran d'aplicar, o no, quan es treballi amb aquests models, dependent de les seves característiques, en la fase de pre-procés de la base de dades o bé en la fase de construcció del model predictiu.

Conèixer diferents mètriques de comportament per a fer la validació del model i entendre la seva utilitat i adequació, en funció de les característiques de la base de dades.

Aprendre com construir scripts de R que permetin aprendre aquests models a partir d'una base de dades i fer la seva validació, fent servir les llibreries pertinents. Aplicar-ho amb dades reals.

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.
- Dissenyar un estudi estadístic o de recerca operativa per a la resolució d'un problema real.
- Formular hipòtesis estadístiques i desenvolupar estratègies per confirmar-les o refutar-les.
- Interpretar resultats, extreure conclusions i elaborar informes tècnics en el camp de l'estadística.
- 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'índole social, científica o ètica.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.
- Seleccionar i aplicar procediments més apropiats per a la modelització estadística i l'anàlisi de dades complexes.
- 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 altres models per a dades complexes (dades funcionals, dades de recompte, etc.).
2. Analitzar dades mitjançant tècniques d'inferència utilitzant programari estadístic.
3. Aplicar l'esperit crític i el rigor per validar o refutar arguments, tant propis com d'altres.
4. Avaluar de manera crítica i amb criteris de qualitat la feina feta.
5. Elaborar informes tècnics específics de l'àmbit de la modelització estadística.
6. Establir les hipòtesis experimentals de la modelització.
7. Fer servir gràfics de visualització de l'ajustament i de l'adequació del model.
8. Identificar les etapes en els problemes de modelització.
9. Identificar les suposicions estadístiques associades a cada procediment.
10. Modificar lleugerament el programari existent si el model estadístic proposat ho requereix.
11. Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
12. 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.
13. 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'índole social, científica o ètica.
14. Utilitzar eficaçment bibliografia i recursos electrònics per obtenir informació.
15. Validar els models utilitzats mitjançant tècniques d'inferència adequades.

Continguts

1. Introducció a les Xarxes Bayesianes (XBs).
Definició.
Inferència amb XBs.
Aprenentatge de XBs (tant d'estructura com de paràmetres).
2. Les XB com a classificadors.
La tasca de classificació dins l'Aprenentatge Automàtic Supervisat.
El criteri MAP.
Tipus de XB (*Naive Bayes*, *Augmented Naive*, *TAN*).
Tipus de classificació: binària, multi-class, multi-label.
3. La validació i les mètriques que comportament.
Validació creuada.
Mètriques pel cas binari i multi-class.
Mètriques pel cas de classificació ordinal.
4. Altres aspectes.
Classificació multi-dimensional.
Ensembles de classificadors.
El "*concept drive*" i les XBs dinàmiques.
XBs gaussianes i híbrides.
Classificació *multi-instance*.

Metodologia

L'assignatura s'estructura a partir de classes teòriques, de problemes i de pràctiques. El seguiment de l'assignatura és presencial, però caldrà ampliar les explicacions del professor amb l'estudi autònom de l'alumne, amb el suport de la bibliografia de referència i del material facilitat pel professor.

La classe de problemes es dedicaran a la resolució orientada d'alguns problemes proposats. A les classes pràctiques es treballarà amb R i les seves llibreries. Es valorarà especialment la participació dels estudiants a les classes de problemes i pràctiques.

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.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Problemes	14	0,56	1, 6, 7, 8, 9, 11, 15
Pràctiques (lliuraments, controls)	12	0,48	2, 4, 5, 10
Teoria	26	1,04	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Tipus: Supervisades			
Tutories	10	0,4	3, 4, 12, 13, 14
Tipus: Autònomes			

Estudi i pensar problemes	40	1,6	6, 7, 8, 9, 11, 15
Treball pràctic amb instruments informàtics	30	1,2	2, 4, 5, 10, 15

Avaluació

La qualificació final d'aquesta assignatura s'obté com la mitjana ponderada de les notes de:

- PAC1 (20%)
- PAC2 (20%)
- Examen (60%)

Les proves d'avaluació contínua PAC1 i PAC2 consisteixen en un lliurament de problemes/exercicis de pràctiques/treballs amb R, que es concretaran al llarg del curs, i en el seu desenvolupament en les classes presencials al llarg del semestre.

Només es tindran en compte en el càlcul de la mitjana ponderada aquelles notes que siguin com a mínim 3.5 sobre 10 (les que no ho compleixin ponderaran 0).

Per a superar l'assignatura cal que aquesta mitjana sigui almenys 5.0 sobre 10.

Si no es supera l'assignatura en primera convocatòria, l'alumne es pot presentar a recuperació. L'examen de recuperació representa el 100% de la nota final per a aquells alumnes que es presenten a la recuperació, que només poden ser alumnes que no han superat l'assignatura a la primera convocatòria (l'examen de recuperació no serveix per a millorar la nota d'alumnes que ja han aprovat).

Es considera avaluable l'alumne que hagi presentat els lliuraments PAC1 o PAC2, o s'hagi presentat a l'examen o l'examen de recuperació. En cas contrari constarà a l'acta com a No Avaluable.

Per a l'eventual assignació de Matricules d'Honor no es tindran en compte les notes de la segona convocatòria.

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen	60%	3	0,12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
PAC1	20%	6	0,24	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
PAC2	20%	9	0,36	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

Bibliografia

- Norman Fenton and Martin Neil, "Risk Assessment and Decision Analysis with Bayesian Networks", CRC Press. A Chapman & Hall Book, 2013. (Disponible en línia)
- Radhakrishnan Nagarajan, Marco Scutari and Sophie Lèbre, "Bayesian Networks in R with applications in Systems Biology", Springer, 2013. (Disponible en línia)
- Oliver Porret, Patrick Naïm and Bruce Marcot, "Bayesian Networks. A practical guide to applications". Series: Statistics in Practice. Wiley, 2008. (Disponible en línia)
- Richard E. Neapolitan, "Learning Bayesian Networks", Prentice Hall Series in Artificial Intelligence, 2004.

Programari

Es farà servir el programari R amb algunes llibreries que s'indican oportunament al llarg del curs. Preferiblement en l'entorn RStudio.