

Mineria de Dades

2012/2013

Codi: 103205

Crèdits ECTS: 6

Titulació	Pla	Tipus	Curs	Semestre
2501919 Graduat en Estadística Aplicada	973 Graduat en Estadística Aplicada	OB	3	2

Professor de contacte

Nom: Joan Valls Marsal

Correu electrònic: Joan.Valls@uab.cat

Utilització d'idiomes

Llengua vehicular majoritària: català (cat)

Algun grup íntegre en anglès: No

Algun grup íntegre en català: Sí

Algun grup íntegre en espanyol: No

Prerequisits

Inferència estadística, estadística descriptiva i anàlisi multivariant

Objectius

Conèixer i aprofundir en les tècniques de mineria de dades

Competències

- Analitzar les dades mitjançant l'aplicació de mètodes i tècniques estadístiques i treballar amb dades qualitatives i quantitatives.
- Desenvolupar estratègies d'aprenentatge autònom.
- Dissenyar un estudi estadístic o de recerca operativa per a la resolució d'un problema real.
- Identificar els models estadístics i de recerca operativa més adequats per a cada context i que permetin la presa de decisions.
- Interpretar resultats, extreure conclusions i elaborar informes tècnics.
- Reconèixer els avantatges i els inconvenients dels procediments estudiats.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.
- Utilitzar bibliografia o eines d'Internet específiques de l'estadística i d'altres ciències afins, tant en llengua anglesa com en la llengua pròpia.
- Utilitzar correctament una bona part del programari estadístic i de recerca operativa existent, escollir el més apropiat per a cada anàlisi estadística i ser capaç d'adaptar-lo a les noves necessitats.

Resultats d'aprenentatge

1. Analitzar dades mitjançant metodologia de mineria de dades.
2. Conèixer programari estadístic per als procediments de mineria de dades.
3. Descobrir comportaments i tipologies d'individus mitjançant tècniques de mineria de dades.
4. Desenvolupar estratègies d'aprenentatge autònom.
5. Identificar el model o els models més apropiats utilitzant tècniques de mineria de dades.
6. Identificar l'existència de recursos específics per a l'ús i l'aprenentatge del programari estadístic i familiaritzar-s'hi.
7. Projectar un estudi a partir de metodologies multivariants i/o una mineria de dades per resoldre un problema contextualitzat en la realitat experimental.
8. Redactar l'informe tècnic a partir d'una anàlisi estadística avançada.

9. Utilitzar mètodes de mineria de dades per validar i comparar possibles models.

Continguts

Introducció a R

Qüestions bàsiques. Escalars, vectors, factors i matrius
Lectura i escriptura de fitxers de dades
Programació de funcions i gràfics

Mètodes basats en arbres

Arbres de classificació
Arbres de regressió
Poda d'arbres (tree pruning)

Tècniques de Support Vector Machine i discriminació flexible

Support vector classifier
Support Vector Machines i Kernels
Mètodes de discriminació flexible

Xarxes Neuronals

Models estadístics i xarxes neuronals
Ajust de xarxes neuronals: valors inicials i entrenament

Mineria de textos

Taules lèxiques i construcció del diccionari
Anàlisi de correspondències múltiples aplicada a la mineria de textos

Problemes $p \gg n$

Quan p és molt més gran que n
False Discovery Rate

Metodologia

Classes teòriques i pràctiques amb R

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes teòriques i pràctiques amb R	130	5,2	

Avaluació

Examens parcials i entrega de treballs

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examens parcials i entrega de treballs	100%	20	0,8	1, 2, 3, 4, 5, 6, 7, 8, 9

Bibliografia

T Hastie, R Tibshirani, J Friedman. (2009) *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, Springer, New York.

L Torgo. (2010) *Data Mining with R. Learning with Case Studies*, Chapman & Hall, Miami.

J Faraway. (2006) *Extending de Linear Model with R*, Chapman & Hall, Miami.

J Faraway. (2005) *Linear Models with R*, Chapman & Hall, Boca Raton.

W Venables, B Ripley. (2002) *Modern Applied Statistics with S-PLUS*, Springer, New York.