

Mineria de Dades

2015/2016

Codi: 103205

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2501919 Estadística Aplicada	OB	3	2

Professor de contacte

Nom: Joan Valls Marsal

Correu electrònic: Joan.Valls@uab.cat

Utilització de llengües

Llengua vehicular majoritària: català (cat)

Prerequisites

Assignatures d'estadística inferencial i de modelització prèvies

Objectius

Conèixer alguns de les tècniques de mineria de dades emprades en la modelització de dades

Adquirir les habilitats necessàries per a implementar els mètodes estadístics en R

Competències

- Analitzar les dades mitjançant l'aplicació de mètodes i tècniques estadístiques i treballar amb dades qualitatives i quantitatives.
- Demostrar iniciativa i inquietud per tal d'actualitzar els coneixements.
- Demostrar que es té un pensament lògic, un raonament estructurat i capacitat de síntesi.
- Desenvolupar estratègies d'aprenentatge autònom.
- Desenvolupar l'interès pel coneixement i la interpretació de fenòmens susceptibles de quantificació.
- Desenvolupar un pensament i un raonament crítics i saber comunicar-los de manera efectiva, tant en les llengües pròpies com en una tercera llengua.
- Dissenyar un estudi estadístic o de recerca operativa per a la resolució d'un problema real.
- Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes
- Identificar els models estadístics i de recerca operativa més adequats per a cada context i que permetin la presa de decisions.
- Identificar la utilitat i la potencialitat de l'estadística en les diferents àrees de coneixement i saber aplicar-la adequadament per a extreure'n conclusions rellevants.
- 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. Conèixer programari estadístic per programar funcions i procediments avançats.
4. Demostrar iniciativa i inquietud per tal d'actualitzar els coneixements.
5. Demostrar que es té un pensament lògic, un raonament estructurat i capacitat de síntesi.
6. Descobrir comportaments i tipologies d'individus mitjançant tècniques de mineria de dades.
7. Descriure els avantatges i inconvenients dels mètodes algorítmics enfront dels mètodes convencionals de la inferència estadística.
8. Desenvolupar estratègies d'aprenentatge autònom.
9. Desenvolupar l'interès pel coneixement i la interpretació de fenòmens susceptibles de quantificació.
10. Desenvolupar un pensament i un raonament crítics i saber comunicar-los de manera efectiva, tant en les llengües pròpies com en una tercera llengua.
11. Emprar índexs de resum de dades multivariants, sèries temporals i totes les altres tècniques avançades.
12. Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes.
13. Extreure conclusions rellevants a problemes aplicats, mitjançant l'aplicació de mètodes estadístics avançats.
14. Fer consultes bibliogràfiques en l'àmbit dels mètodes avançats, també en llengua anglesa.
15. Identificar el model o els models més apropiats utilitzant tècniques de mineria de dades.
16. Identificar, emprar i interpretar els criteris per avaluar el grau de compliment dels requisits necessaris per aplicar cada procediment avançat.
17. Identificar les característiques metodològiques especials en l'anàlisi estadística segons les diferents àrees d'aplicació.
18. Identificar les diferents àrees d'aplicació més comunes per a cada metodologia avançada.
19. Identificar les suposicions estadístiques associades a cada procediment avançat.
20. Identificar l'existència de recursos específics d'aplicació de metodologies estadístiques i practicar-hi.
21. Identificar l'existència de recursos específics per a l'ús i l'aprenentatge del programari estadístic i familiaritzar-s'hi.
22. Interpretar resultats amb metodologies avançades i extreure'n conclusions.
23. Projectar un estudi a partir de metodologies multivariants i/o una mineria de dades per resoldre un problema contextualitzat en la realitat experimental.
24. Redactar l'informe tècnic a partir d'una anàlisi estadística avançada.
25. Utilitzar mètodes de mineria de dades per validar i comparar possibles models.

Continguts

1. Introducció a R

Qüestions bàsiques. Escalars, vectors, factors i matrius

Lectura i escriptura de fitxers de dades

Programació de funcions i gràfics

2. Mètodes basats en arbres

- Arbres de classificació

- Arbres de regressió

- Poda d'arbres

3. Tècniques de Support Vector Machine i discriminació flexible

- Support vector classifier

- Support Vector Machines i discriminació emprant Kernels

- Mètodes de discriminació flexible

4. Xarxes Neuronals

- Models estadístics i xarxes neuronals

- Ajust de xarxes neuronals: valors inicials i entrenament

5. Minería de textos

- Taules lèxiques i construcció del diccionari

- Anàlisi de correspondències múltiples aplicada a la mineria de textos

6. Problemes $p \gg n$

- Quan p és molt més gran que n

- False Discovery Rate

Metodologia

Classes magistrals i classes pràctiques amb R.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
classe magistral i classes pràctiques	150	6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25

Avaluació

Mitjantçant examens parcials i realització d'un projecte d'anàlisi de dades

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
examens	60%	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
projecte d'anàlisi de dades	40%	0	0	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 14, 20, 21, 23, 24, 25

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.