

Guia docent

270652 - MVA - Anàlisi Multivariant

Última modificació: 14/02/2020

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

Titulació: MÀSTER UNIVERSITARI EN INNOVACIÓ I RECERCA EN INFORMÀTICA (Pla 2012). (Assignatura optativa).

Curs: 2019 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable:

Altres:

CAPACITATS PRÈVIES

L'assignatura suposa haver efectuat previament un curs basic d'estadística, programació i matemàtiques, en particular tenir adquirits els conceptes següents

- Concepte de mitjana, matriu de covariancies i correlacions.
- Concepte de prova de hipòtesis.
- Operacions d'àlgebra matricial, valors i vectors propis.
- Programació d'algorismes.
- Regressió lineal múltiple

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CEC1. Capacitat per aplicar el mètode científic en l'estudi i anàlisi de fenòmens i sistemes en qualsevol àmbit de la Informàtica, així com en la concepció, disseny i implantació de solucions informàtiques innovadores i originals.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

Genèriques:

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Transversals:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

METODOLOGIES DOCENTS

L'objectiu de l'assignatura es donar els fonaments estadístics subjacents a la Mineria de Dades. L'aprenentatge es farà mitjançant la combinació de l'explicació teòrica i la seva aplicació a un cas real. En les classes de teoria es desenvoluparan els coneixements científics necessaris, mentre que en les classes de laboratori es veurà la seva aplicació per a la resolució de problemes de Mineria de Dades.

Aquests problemes constituïran les practiques de l'assignatura, que es desenvoluparan en part durant les classes de laboratori. La realització de les pràctiques fomenta les competències transversals lligades al treball en equip i presentació de resultats i serveixen per integrar els diferents coneixements de l'assignatura.

El software a utilitzar sera fonamentalment l'R.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	96,0	64.00
Hores activitats dirigides	2,0	1.33
Hores grup gran	26,0	17.33
Hores grup petit	26,0	17.33

Dedicació total: 150 h

CONTINGUTS

(CAT) Multivariate Data Analysis

Descripció:

(CAT) Advantages of the multivariate treatment. Examples of multivariate data. Probabilistic and distribution free methods. Exploratory versus modeling approach.

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Principal Component Analysis

Descripció:

(CAT) Analysis of individuals. Analysis of variables. Visual representation of the information. Dimensionality reduction. Supplementary information

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)



(CAT) Singular Value Decomposition. Biplots

Descripció:

(CAT) Simultaneous representation of the rows and columns of a data table.

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Factor Analysis

Descripció:

(CAT) Latents constructs. Measurement model.

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Multidimensional Scaling

Descripció:

(CAT) Visualisation of link matrices

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Correspondence Analysis

Descripció:

(CAT) Analysis of frequency data

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Multiple Correspondence Analysis

Descripció:

(CAT) Analysis of categorical data

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)



(CAT) Hierarchical clustering

Descripció:

(CAT) Synthesis of the represented information. Consolidation of the partition

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Multivariate normal distribution

Descripció:

(CAT) Definition and properties

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Sampling distributions of the normal multivariate distribution

Descripció:

(CAT) Inferences respect to the covariance matrix. Inferences respect to the centroid of the distribution. Whishart distribution. T2 of Hotelling, Wilks lambda.

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Discriminant Analysis

Descripció:

(CAT) With the assumption of multivariate normal distribution. Linear discriminant analysis. Quadratic discriminant analysis.

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Naive Bayes

Descripció:

(CAT) Simplifying the linear discriminant analysis

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)



(CAT) Discriminant analysis without probabilistic assumptions

Descripció:

(CAT) K nearest neighbor classifier

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Decision trees

Descripció:

(CAT) Classification and regression trees

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

(CAT) Association rules

Descripció:

(CAT) Apriori algorithm

Objectius específics:

(CAT)

Activitats vinculades:

(CAT)

ACTIVITATS

Anàlisi Multivariant de Dades

Objectius específics:

1, 2

Dedicació: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 2h



Anàlisi de Components Principals

Objectius específics:

1, 2

Dedicació: 11h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 7h

Descomposició en Valors Singulars. Biplots

Objectius específics:

1, 2

Dedicació: 7h

Grup gran/Teoria: 1h

Grup petit/Laboratori: 1h

Aprenentatge autònom: 5h

Anàlisi Factorial

Objectius específics:

1, 2

Dedicació: 7h

Grup gran/Teoria: 1h

Grup petit/Laboratori: 1h

Aprenentatge autònom: 5h

Escalat Multidimensional

Objectius específics:

1, 2

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Anàlisi de Correspondències

Objectius específics:

1, 2

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h



Anàlisi de Correspondències Múltiples

Objectius específics:

1, 2

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Classificació Jeràrquica

Objectius específics:

1, 2

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Distribució normal multivariant

Objectius específics:

3

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Distribucions muestrals.

Objectius específics:

3

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Proves estadístiques multivariants

Objectius específics:

4

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h



T2 de Hotelling

Objectius específics:

4

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Classificació per arbres de decisió

Objectius específics:

4

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Regles d'associació

Objectius específics:

4

Dedicació: 9h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Pràctiques sobre Anàlisi Exploratori de Dades

Dedicació: 10h

Activitats dirigides: 1h

Aprenentatge autònom: 9h

Pràctica Final

Dedicació: 9h

Activitats dirigides: 1h

Aprenentatge autònom: 8h

Examen de conceptes

Dedicació: 10h

Aprenentatge autònom: 10h

SISTEMA DE QUALIFICACIÓ

L'avaluació de l'assignatura es realitzarà a partir de la nota obtinguda en els exercicis pràctics realitzats durant el curs, la nota d'un examen i la nota obtinguda en la pràctica final.

Cada pràctica comportarà la redacció del corresponent informe escrit i podrà ser efectuada conjuntament, fins un màxim de dos alumnes per grup.

Els exercicis efectuats al llarg del curs tenen com a finalitat consolidar l'aprenentatge de les tècniques multivariants.

La pràctica final tracta de que l'alumne mostri la seva maduresa per resoldre un problema real, utilitzant les tècniques de visualització multivariant, "clustering", interpretació de la partició obtinguda i predicció. L'alumne escollirà el problema a resoldre entre diferents alternatives. Aquesta pràctica haurà de ser presentada i defensada públicament i l'estudiant haurà de respondre les eventuales preguntes teòriques sobre els models i mètodes usats a la seva solució. Les pràctiques es realitzaran mitjançant el software R.

La prova escrita es realitzarà el darrer dia de classe i avaluarà l'assimilació dels conceptes bàsics de l'assignatura. Mentre que la presentació de la segona practica es fara durant el periode d'examens.

Els exercicis es ponderaran amb un 30%, l'examen en un 40% respectivament i la pràctica final amb un 30%.

BIBLIOGRAFIA

Bàsica:

- Aluja Banet, T.; Morineau, A. Aprender de los datos: el análisis de componentes principales: una aproximación desde el data mining. Barcelona: EUB, 1999. ISBN 8483120224.
- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction [en línia]. 2nd ed. New York: Springer, 2009 [Consulta: 11/03/2020]. Disponible a: <http://dx.doi.org/10.1007/978-0-387-84858-7>. ISBN 9780387848570.
- Johnson, R.A.; Wichern, D.W. Applied multivariate statistical analysis [en línia]. 6th ed. Harlow, Essex: Prentice Hall, 2014 [Consulta: 11/03/2020]. Disponible a: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=5174865>. ISBN 9781292037578.
- Lebart, L.; Morineau, A.; Warwick, K.M. Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices. New York: John Wiley and Sons, 1984. ISBN 0471867438.
- Hand, D.J. Construction and assessment of classification rules. Chichester: Wiley, 1997. ISBN 0471965839.
- Husson, F.; Le, S.; Pagès, J. Exploratory multivariate analysis by example using R. Boca Raton: CRC Press, Taylor & Francis Group, 2017. ISBN 9781315301860.

Complementària:

- Peña, D. Análisis de datos multivariantes [en línia]. Madrid: McGraw Hill, cop. 2002 [Consulta: 11/03/2020]. Disponible a: http://www.ingebook.com/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203. ISBN 9788448136109.
- Everitt, B.S. An R and S-PLUS companion to multivariate analysis [en línia]. London: Springer-Verlag London Limited, 2005 [Consulta: 11/03/2020]. Disponible a: <http://dx.doi.org/10.1007/b138954>. ISBN 9781846281242.
- Lebart, L.; Morineau, A.; Piron, M. Statistique exploratoire multidimensionnelle: visualisations et inférences en fouille de données. 4e éd. Paris: Dunod, 2006. ISBN 9782100496167.

RECURSOS

Enllaç web:

- <https://cran.r-project.org/>. Homepage of R
- http://www.dtmvic.com/05_SoftwareE.html. Software DtmVic: Exploratory statistical processing of complex data sets comprising both numerical and textual data.