

Titulació	Tipus	Curs
2503852 Estadística Aplicada	OB	3

Professor/a de contacte

Nom: Maria Merce Farre Cervello

Correu electrònic: merce.farre@uab.cat

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

És imprescindible un curs previ d'àlgebra lineal, així com els cursos de probabilitats, dustrubucions multidimensionals i inferència estadística. També se suposa que sap utilitzar amb agilitat el llenguatge R.

Objectius

La necessitat de processar gran quantitat de dades amb moltes variables de naturalesa diversa, tot i reduint la informació que no sigui rellevant i descobrint patrons d'associació entre variables i/o casos, han ocasionat el desenvolupament d'un gran nombre de procediments que se situen en l'escenari multivariant. En Aprenentatge No Supervisat es tracten els mètodes que són més apropiats per a descriure, explorar i modelitzar dades vectorials. L'assignatura es planteja com un primer contacte de l'estudiant amb el món de l'anomenat "aprenentatge estadístic" ("statistical learning"), per tal que compregui la potència i aplicabilitat, i alhora les limitacions, dels mètodes, alguns dels quals es basen en idees heurístiques força intuïtives. La majoria dels mètodes treballats al curs són de tipus no-supervisat, és a dir, no es disposa d'un conjunt de casos amb respostes conegudes que permetin avaluar el mètode. L'enfocament de l'assignatura és eminentment aplicat pel que fa al treball amb dades usant el potencial del programari lliure R, acompanyat del rigor i generalitat adequats en la definició del models teòrics i els corresponents mètodes d'anàlisi i validació de resultats.

Resultats d'aprenentatge

1. CM11 (Competència) Crear nous models d'aprenentatge automàtic, executant experiments per a demostrar-ne la viabilitat i millora del rendiment respecte a l'estat de l'art.
2. CM12 (Competència) Valorar l'existència de desigualtats per raó de gènere en les bases de dades, per a evitar els biaixos en la presa de decisions automàtica (algorísmica).
3. KM16 (Coneixement) Reconèixer models d'aprenentatge automàtic, supervisat i no supervisat, profund i generatiu, fomentant la innovació en l'àmbit de l'estadística.
4. KM16 (Coneixement) Reconèixer models d'aprenentatge automàtic, supervisat i no supervisat, profund i generatiu, fomentant la innovació en l'àmbit de l'estadística.

Continguts

Aprenentatge estadístic (*statistical learning*) i reducció de la dimensió

- L'aprenentatge supervisat i no-supervisat. Classificació dels mètodes multivariants. Exemples.
- Vectors aleatoris. El vector d'esperances i les matrius de covariàncies i correlacions. Propietats.
- Dades multivariants. El vector de mitjanes i les matrius de covariància i correlació mostrals.
- Estimadors de màxima versemblança en el cas Gaussià.
- Les descomposicions espectral (SD) i en valors singulars (SVD).
- Teorema fonamental de maximització de formes quadràtiques amb restriccions.

Mètodes factorials I: Anàlisi de components principals (ACP)

- Introducció a l'ACP. Definició de components. El resultat fonamental.
- Criteris per decidir el nombre de components: Les components principals.
- Gràfiques de variables i d'individus. Estandarditzacions.
- Anàlisi de la matriu de vectors propis, i de matrius derivades, per files i per columnes.
- Punt de vista geomètric de l'ACP.

Mètodes factorials II: Anàlisi factorial (AF)

- El model factorial. Comunalitats i especificitats.
- Teorema de descomposició de la matriu de covariàncies.
- Existència i unicitat del model factorial? Rotacions.
- Mètodes d'estimació dels paràmetres del model factorial. Les puntuacions factorials (*scores*): estimació o predicció.
- Interpretació dels resultats. Comparació de l'ACP i l'AF.

Mètodes factorials III: Escalament multidimensional (EMD) i anàlisi de correspondències (AC)

- Objectius i mètodes.
- EMD: clàssic i mètric.
- EMD no mètric.
- Distàncies, proximitats i dissimilaritats.
- Dades categòriques: distància *khi-quadrat* i altres.
- Anàlisi de correspondències (AC) com a cas particular d'EMD.
- Perfils i inèrcia. Descomposicions de la inèrcia.
- Representació gràfica i interpretació dels resultats de l'AC.

Anàlisi de clústers (ACL)

- Comparació de diferents mètodes de clusterització. Exemples.
- Anàlisi i validació dels resultats dels procediments de clusterització.
- Clústers jeràrquics: Funcions d'enllaç.
- Mètodes basats en centroides: El mètode de les k-mitjanes.
- Mètodes basats en models probabilístics: Esperança i maximització (EM).

Bàsics en inferència multivariant

- Test de la raó de versemblances.
- Tests per a vectors de mitjanes.
- Tests per a matrius de covariàncies. ANOVA i MANOVA.

Anàlisi discriminant (AD)

- Objectius i criteris d'AD.
- Anàlisi discriminant en models Gaussians.
- Discriminant lineal de Fisher.

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de pràctiques amb ordinador	26	1,04	
Classes de teoria	26	1,04	
Tipus: Autònomes			
Estudi i consultes	42	1,68	
Resolució i lliurament d'exercicis	44	1,76	

Classes de teoria, on s'exposaran i es discutiran els continguts en base a exemples d'aplicació convenientment interpretats. Els resums dels apunts es penjaran al campus virtual. S'encoratjarà la revisió i ampliació de continguts amb la bibliografia bàsica i complementària.

Classes pràctiques pensades per implementar amb el llenguatge estadístic R. Els guions de les pràctiques i d'altre material auxiliar es penjaran al Campus Virtual. A les classes pràctiques es proposaran exercicis d'ampliació que caldrà resoldre autònomament.

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.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Lliurament de tasques (problemes i pràctiques resoltes)	0,2	4	0,16	CM11, CM12, KM16
Primer parcial (teoria i pràctiques)	0,35	4	0,16	CM11, KM16
Segon parcial (teoria i pràctiques)	0,45	4	0,16	CM11, KM16

Avaluació ordinària

La nota de curs (NC) es calcularà a partir de les entregues d'exercicis i pràctiques i de les notes dels exàmens parcials (P1 i P2) basats en qüestions de teoria i pràctiques, i de la nota dels lliuraments. Detall de la nota de curs:

$$NC = 0,35 \cdot P1 + 0,45 \cdot P2 + 0,20 \cdot Lli$$

on P1 és la nota del primer parcial, P2 del segon parcial i Lli és la nota dels lliuraments que es proposaran al llarg del curs. L'avaluació d'alguns d'aquests lliuraments es farà amb un test a la mateixa aula de pràctiques els dies indicats.

L'aprovat per curs requereix que NC sigui igual o més gran que 5 i que les notes de cada parcial siguin més grans que 3.5 (sobre 10). L'examen final de recuperació (F) serà una prova de síntesi de les competències de tot el curs i tindrà una part teòrica i una de computacional. Si l'estudiant vol millorar la nota de curs, també es pot presentar a la recuperació. No obstant l'anterior, el professorat es reserva el dret a decidir les matrícules d'honor, si n'hi ha, abans de l'examen de recuperació. La nota final serà:

$$NF = \text{Max}(NC, 0,80 \cdot F + 0,20Li)$$

Avaluació única

L'avaluació única serà una prova de síntesi de les competències dels dos parcials, en base a: (1) Un examen amb qüestions de teoria i de practiques (pes: 50%). (2) Una prova de practiques davant de l'ordinador (pes: 40%). (3) El lliurament de les tasques programades que s'indiquin, amb la possibilitat de que el professorat demani que l'estudiant expliqui detalls d'aquests lliuraments (pes: 10%).

Bibliografia

Everitt, B., Hothorn, T. ; An introduction to Applied Multivariate Analysis with R. Springer, 2011.

Härdle, W., Simar, L.; Applied Multivariate Statistical Analysis. Springer,2007.

Peña, D.; Análisis de datos multivariantes. McGraw Hill, 2002.

Rencher, A., Christensen, W.; Methods of Multivariate Analysis. Wiley Series in Probability and Mathematical Statistics, 2012.

Wehrens, R. (2020). Chemometrics with R: Multivariate data analysis in the natural sciences and life sciences. Heidelberg: Springer. <https://link-springer-com.are.uab.cat/book/10.1007/978-3-662-62027-4>

Bibliografia complementària

Coghlan, A.; Little book of R for Multivariate Analysis.

<https://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/>

Cuadras, C.; Nuevos Métodos de Análisis Multivariante (web), 2014.

Greenacre, M.; La pràctica del anàlisi de correspondències. Fundacion BBA, 2003.

James, G., Witten, D., Hastie, T., Tibshirani, R.; An Introduction to Statistical Learning. Springer, 2014.

Mardia, K.V, Kent, J.T., Bibby, J.M.; Multivariate Analysis. Academic Press, 2003.

Rencher, A.; Multivariate Statistical Inference and Applications. John Wiley & Sons, 1998.

Programari

Programari lliure R, en l'entorn de treball RStudio.

Llista d'idiomes

La informació sobre els idiomes d'impartició de la docència es pot consultar a la part de CONTINGUTS de la guia.

PROVISIONAL