

Aprenentatge No Supervisat

Codi: 104869

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2503852 Estadística Aplicada	OB	2	2

La metodologia docent i l'avaluació proposades a la guia poden experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries.

Professor/a de contacte

Nom: Mercè Farré Cervelló

Correu electrònic: Merce.Farre@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: català (cat)

Grup íntegre en anglès: No

Grup íntegre en català: Sí

Grup íntegre en espanyol: No

Prerequisits

És imprescindible un curs previ d'Àlgebra Lineal, així com cursos bàsics de Probabilitats i d'Inferència Estadística. També se suposa que sap utilitzar amb agilitat el programari R.

Objectius

La majoria de dades recollides són multivariants, és a dir, per a una mateixa unitat experimental (que pot ser un objecte de naturalesa complexa) observem de manera simultània els valors de diverses variables. L'Anàlisi Multivariant tracta els mètodes que són més apropiats per a descriure, explorar i modelitzar aquestes dades vectorials, així com per a aplicar-hi la inferència estadística. La necessitat de processar gran quantitat de dades amb moltes variables de naturalesa diversa, amb la voluntat de reduir la informació que no sigui rellevant i de descobrir patrons d'associació entre variables, així com l'interès en la classificació supervisada o autònoma de casos, han fet desenvolupar recentment un gran nombre de procediments que se situen en l'escenari multivariant. 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. 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 dels models teòrics i els corresponents mètodes d'anàlisi i validació dels resultats.

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.
- Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia.
- 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'indole 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.
- Seleccionar les fonts i tècniques d'adquisició i gestió de dades adequades per a fer-ne un tractament estadístic.
- Treballar cooperativament en un context multidisciplinari assumint i respectant el rol dels diferents membres de l'equip.
- Utilitzar eficaçment la bibliografia i els recursos electrònics per obtenir informació.

Resultats d'aprenentatge

1. Analitzar dades mitjançant la metodologia d'aprenentatge automàtic.
2. Aplicar l'esperit crític i el rigor per validar o refutar arguments, tant propis com d'altres.
3. Avaluar de manera crítica i amb criteris de qualitat la feina feta.
4. Caracteritzar grups homogenis d'individus mitjançant una anàlisi multivariant.
5. Descriure els avantatges i els inconvenients dels mètodes algorítmics vers els mètodes convencionals de la inferència estadística.
6. Fer servir gràfics de resum de dades multivariades o més complexes.
7. Identificar les suposicions estadístiques associades a cada procediment avançat.
8. Identificar, fer servir i interpretar els criteris per avaluar el grau de compliment dels requisits necessaris per aplicar cada procediment avançat.
9. Obtenir i gestionar bases de dades complexes per analitzar-les posteriorment.
10. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia.
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'indole social, científica o ètica.
14. Treballar cooperativament en un context multidisciplinari assumint i respectant el rol dels diferents membres de l'equip.
15. Utilitzar eficaçment bibliografia i recursos electrònics per obtenir informació.

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) i altres mètodes supervisats

- Objectius i criteris d'AD.
- Anàlisi discriminant en models Gaussians.
- Discriminant lineal de Fisher.
- Regressió per mínims quadrats parcials (PLS). Altres mètodes.

Metodologia

Classes de teoria, on s'exposaran detalladament 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, enunciats de problemes i l'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.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de pràctiques amb ordinador	26	1,04	1, 5, 6, 7, 8, 9, 12, 13, 14, 15
Classes de teoria	26	1,04	1, 2, 3, 4, 5, 6, 7, 8, 9
Tipus: Autònomes			
Estudi i consultes	42	1,68	2, 5, 7, 8, 10, 15
Resolució i lliurament d'exercicis	44	1,76	1, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15

Avaluació

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), amb preguntes de teoria i de 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 i que no es podran recuperar.

L'aprobat 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 tot el curs i tindrà una part teòrica i una de computacional. Si un alumne vol millorar la nota de curs, també es pot presentar a la recuperació. No obstant l'anterior, el professor es reserva el dret a decidir les matrícules d'honor, si n'hi ha, abans de l'examen de recuperació. La nota final serà:

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

Observació: Només es podran presentar a la prova final aquells estudiants que hagin participat a 2/3 parts de les activitats d'avaluació continuada.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Lliurament de tasques (problemes i pràctiques resoltes)	0,2	4	0,16	2, 3, 9, 10, 11, 12, 13, 14, 15
Primer parcial (teoria i pràctiques)	0,35	4	0,16	1, 6, 7, 8, 10, 11
Segon parcial (teoria i pràctiques)	0,45	4	0,16	1, 4, 5, 7, 8, 9, 10, 11, 12, 15

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 anàlisis 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.