

Guia docent

240615 - 240615 - Introducció a Data Science

Última modificació: 16/04/2024

Unitat responsable: Escola Tècnica Superior d'Enginyeria Industrial de Barcelona
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.

Titulació: GRAU EN ENGINYERIA EN TECNOLOGIES INDUSTRIALS (Pla 2010). (Assignatura optativa).

Curs: 2024 **Crèdits ECTS:** 4.5 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: JOSEP GINEBRA

Altres: JOSEP GINEBRA

CAPACITATS PRÈVIES

Haver aprovat l'assignatura d'Estadística

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

1. Coneixements bàsics sobre l'ús i programació d'ordinadors, sistemes operatius, bases de dades i programes informàtics amb aplicacions en enginyeria.
2. Coneixements i capacitats per organitzar i gestionar projectes. Conèixer l'estructura organitzativa i les funcions d'una oficina de projectes.

Transversals:

3. COMUNICACIÓ EFICAÇ ORAL I ESCRITA: Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.
4. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip, ja sigui com un membre més, o realitzant tasques de direcció amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

METODOLOGIES DOCENTS

Totes les classes són pràctiques, a davant de l'ordinador. L'anàlisi de les dades es farà amb MINITAB i amb R. Cada setmana es proposaran petits exercicis d'anàlisi de dades per fer a casa. Hi haurà un treball final d'assignatura. Al Q1 del curs 2020/21 les classes es podran seguir online, tot i que també hi haurà la opció d'assistir a classe presencialment si així volen.

OBJECTIUS D'APRENENTATGE DE L'ASSIGNATURA

A l'acabar l'assignatura l'estudiant ha de ser capaç d'identificar situacions en les que es profita analitzar dades, d'identificar el model o mètode d'anàlisi més adient per a les seves dades, de construir un model que sintetitzi la informació a les dades i permeti fer prediccions, de reduir la dimensionalitat i visualitzar dades multivariants, de implementar algorismes de classificació supervisada i no supervisada, i d'avaluar la qualitat dels resultats obtinguts.

HORES TOTS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	67,5	60.00
Hores grup mitjà	45,0	40.00

Dedicació total: 112.5 h

CONTINGUTS

Tema 1: Introducció

Descripció:

1.- Modelització estadística. 2.- Anàlisi multivariant.

Dedicació: 3h 30m

Grup gran/Teoria: 1h 30m

Activitats dirigides: 1h

Aprenentatge autònom: 1h

Tema 2: Models lineals

Descripció:

1.- Model lineal normal. 2.- Ajust per mínims quadrats i altres criteris. 3.- Taula ANOVA i bondat de l'ajust. 4.- Intervalls de confiança i tests sobre els coeficients. 5.- Predicció. 6.- Validació del model. 7.- Selecció del model i validació creuada. 8.- Interpretació del model; Colinearitat, biaix i causalitat. 9.- Us de variables explicatives categòriques.

Dedicació: 30h

Grup gran/Teoria: 6h

Grup petit/Laboratori: 6h

Activitats dirigides: 6h

Aprenentatge autònom: 12h

Tema 3: Models no-lineals

Descripció:

1.- Model no-lineal normal teòric i ajustat. 2.- Ajust del model. 3.- Intervalls de confiança i tests. 4.- Validació del model.

Dedicació: 6h

Grup gran/Teoria: 1h 30m

Grup petit/Laboratori: 1h 30m

Aprenentatge autònom: 3h

Tema 4: Models per a resposta categòrica i per competjos

Descripció:

1.- Model per a resposta binària. 2.- Ajust. 3.- Intervals de confiança i tests sobre els coeficients. 4.- Validació. 5.- Selecció de models i validació creuada. 6.- Interpretació del model. 7.- Model per a resposta politòmica. 8.- Taules de contingència i model logístic. 9.- Model per a competjos.

Dedicació: 22h 30m

Grup gran/Teoria: 4h 30m

Grup petit/Laboratori: 4h 30m

Activitats dirigides: 4h 30m

Aprenentatge autònom: 9h

Tema 5: Models de sèries temporals

Descripció:

1.- Descripció d'una sèrie temporal; Estacionarietat i estacionalitat. 2.- Models AR. 3.- Models MA. 4.- Models ARIMA. 5.- Models ARIMA estacionals.

Dedicació: 13h

Grup gran/Teoria: 3h

Grup petit/Laboratori: 3h

Activitats dirigides: 3h

Aprenentatge autònom: 4h

Tema 6: Visualització de dades multivariants (Reducció de la dimensionalitat)

Descripció:

1.- Anàlisi de components principals. 2.- Anàlisi de correspondències.

Dedicació: 6h

Grup gran/Teoria: 1h 30m

Grup petit/Laboratori: 1h 30m

Aprenentatge autònom: 3h

Tema 7: Anàlisi clúster (Classificació no supervisada)

Descripció:

1.- Mètodes jeràrquics. 2.- Mètodes de partició (Algoritmes k-means). 3.- Anàlisi cluster per variables.

Dedicació: 6h 30m

Grup gran/Teoria: 1h 30m

Grup petit/Laboratori: 1h 30m

Activitats dirigides: 1h 30m

Aprenentatge autònom: 2h

Tema 8: Anàlisi discriminant (Classificació supervisada)

Descripció:

1.- Anàlisi discriminant lineal. 2.- Comparació de mètodes de classificació. 3.- Anàlisi discriminant quadràtic. 4.- Anàlisi discriminant logístic. 5.- Mètodes de classificació en arbre. 6.- Nearest neighbor classifier. 7.- Sensitivitat, especificitat i la corba ROC.

Dedicació: 13h

Grup gran/Teoria: 3h

Grup petit/Laboratori: 3h

Activitats dirigides: 3h

Aprenentatge autònom: 4h

SISTEMA DE QUALIFICACIÓ

Hi haurà un examen parcial fet a casa i un examen final fet a classe.

Nota Assignat. = 0,1 Exercicis + 0,3 Projecte + 0,1 Examen Parcial + 0,5 Examen Final

Durant el quadrimestre de primavera del curs 2019-20, i com a conseqüència de la crisi del covid 19, el sistema de qualificació serà el mateix, amb la única diferència de que l'examen final serà un take home, en comptes de fer-lo a classe.

BIBLIOGRAFIA

Bàsica:

- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning : data mining, inference and prediction [en línia]. 2nd ed. New York [etc.]: Springer, cop 2009 [Consulta: 25/08/2022]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7>. ISBN 0387952845.
- Peña, Daniel. Regresión y diseño de experimentos. Madrid: Alianza, 2002. ISBN 9788420693897.
- Venables, William N; Ripley, B.D. Modern Applied Statistics with S. 4th ed. New York: Springer Verlag, 2003. ISBN 0387954570.
- Peña, Daniel. Análisis de datos multivariantes [en línia]. 1a ed. Madrid: McGraw-Hill/Interamericana de España, S.L., 2013 [Consulta: 22/10/2024]. Disponible a: https://www-ingebook-com.recursos.biblioteca.upc.edu/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203. ISBN 9788448191849.
- Weisberg, Sanford. Applied linear regression. 4th ed. Hoboken, New Jersey: Wiley, 2014. ISBN 9781118594858.
- Everitt, B.S.; Dunn, G. Applied multivariate data analysis [en línia]. 2nd. Chichester, West Sussex: Wiley, 2001 [Consulta: 24/10/2024]. Disponible a: <https://onlinelibrary-wiley-com.recursos.biblioteca.upc.edu/doi/epdf/10.1002/9781118887486>. ISBN 9781118887486.
- Greenacre, Michael J. Correspondence analysis in practice [en línia]. 3rd ed. Boca Raton: CRC Press, Taylor & Francis Group, 2017 [Consulta: 20/02/2025]. Disponible a: <https://www-taylorfrancis-com.recursos.biblioteca.upc.edu/books/mono/10.1201/9781315369983/correspondence-analysis-practice-michael-greenacre>. ISBN 9781498731782.
- Gareth James,... [et al.]. An Introduction to statistical learning : with applications in R [en línia]. 2nd ed. New York: Springer, 2021 [Consulta: 01/10/2024]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6686746>. ISBN 1071614177.

Complementària:

- Clarke, B.; Fokoue, E.; Zhang, H.H. Principles and theory for data mining and machine learning [en línia]. 1st ed. New York: Springer, 2009 [Consulta: 26/08/2022]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-98135-2>. ISBN 9780387981345.
- Dobson, Annette J; Barnett, Adrian G. An Introduction to generalized linear models. 4th ed. Boca Raton: CRC Press / Chapman & Hall, 2018. ISBN 9781138741515.
- Johnson, Richard; Wichern, Dean. Applied multivariate statistical analysis [en línia]. Pearson new international edition. Harlow, Essex: Pearson, 2014 [Consulta: 24/10/2024]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=5174>

[865](#). ISBN 9781292024943.

- Peña, Daniel. Análisis de series temporales. 2nd ed. Madrid: Alianza, cop. 2010. ISBN 9788420669458.

- Wakefield, Jon. Bayesian and frequentist regression methods [en línia]. 1st ed. New York, NY: Springer, 2013 [Consulta: 25/10/2024]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-1-4419-0925-1>. ISBN 9781441909244.