

Guia docent

200642 - ODS - Optimització en Ciència de Dades

Última modificació: 29/05/2025

Unitat responsable: Facultat de Matemàtiques i Estadística
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.
Titulació: MÀSTER UNIVERSITARI EN ESTADÍSTICA I INVESTIGACIÓ OPERATIVA (Pla 2013). (Assignatura optativa).
Curs: 2025 **Crèdits ECTS:** 5.0 **Idiomes:** Castellà, Anglès

PROFESSORAT

Professorat responsable: JORDI CASTRO PÉREZ
Altres: Primer quadrimestre:
DANIEL BAENA MIRABETE - A
JORDI CASTRO PÉREZ - A

CAPACITATS PRÈVIES

* Conceptes bàsics d'estadística i d'optimització/investigació operativa.

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

6. CE-2. Capacitat per a dominar la terminologia pròpia d'algun àmbit en el que sigui necessària l'aplicació de models i mètodes estadístics o d'investigació operativa per a resoldre problemes reals.
7. CE-3. Capacitat per a formular, analitzar i validar models aplicables a problemes d'índole pràctica. Capacitat de seleccionar el mètode i/o tècnica estadística o d'investigació operativa més adequada per aplicar aquest model a cada situació o problema concret.
8. CE-5. Capacitat per a formular i resoldre problemes reals de presa de decisions als diferents àmbits d'aplicació sabent triar el mètode estadístic i l'algoritme d'optimització més adequat a cada ocasió.
9. CE-6. Capacitat per a fer servir el software més adequat per a realitzar els càlculs necessaris a la resolució d'un problema.
10. CE-7. Capacitat per a comprendre articles d'estadística i investigació operativa de nivell avançat. Conèixer els procediments d'investigació tant per a la producció de nous coneixements com per a la seva transmissió.
11. CE-8. Capacitat de discutir la validesa, l'abast i la rellevància d'aquestes solucions i saber presentar i defensar les conclusions.
12. CE-9. Capacitat per a implementar algoritmes d'estadística i investigació operativa.

Transversals:

1. EMPRENEDORIA I INNOVACIÓ: Conèixer i comprendre l'organització d'una empresa i les ciències que en regeixen l'activitat; tenir capacitat per comprendre les regles laborals i les relacions entre la planificació, les estratègies industrials i comercials, la qualitat i el benefici.
2. SOSTENIBILITAT I COMPROMÍS SOCIAL: Conèixer i comprendre la complexitat dels fenòmens econòmics i socials típics de la societat del benestar; tenir capacitat per relacionar el benestar amb la globalització i la sostenibilitat; assolir habilitats per usar de forma equilibrada i compatible la tècnica, la tecnologia, l'economia i la sostenibilitat.
3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme 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.
4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.
5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

METODOLOGIES DOCENTS

Teoria:

Es presenten i discuteixen els continguts de l'assignatura combinant explicacions a la pissarra i transparències.

Pràctiques:

Sessions de laboratori en que es mostra l'ús de software.

OBJECTIUS D'APRENENTATGE DE L'ASSIGNATURA

L'objectiu de l'assignatura és introduir a l'alumne en algunes aplicacions en "data science" que poden ser formulades o solucionades per tècniques d'optimització. L'assignatura té tres parts, cada una representa aproximadament 1/3 del total:

1. Solució de problemes estadístics i d'aprenentatge automàtic usant optimització entera i combinatòria: disseny d'experiments (quadrats llatins ortogonals), clustering òptim (k-medoid), clustering heurístic (k-means).
2. Mètodes d'optimització contínua per a la solució de problemes d'aprenentatge automàtic: regressió, "support vector machines" i xarxes neuronals.
3. Introducció al camp del control de la revelació estadística o protecció de dades estadístiques: mètodes per a microdades i per a taules de dades. Aquesta disciplina proposa un conjunt de mètodes per garantir la confidencialitat de dades individuals en disseminar dades estadístiques, siguin microdades o dades agregades en forma tabular. Aquest problema és de gran importància per a Instituts Nacionals d'Estadística, i, en general, qualsevol entitat privada o organisme oficial que hagi de divulgar dades.

Capacitats a adquirir:

- * Formular problemes en "data science" com a problemes d'optimització (clustering, support vector machines, xarxes neuronals...)
- * Saber solucionar els problemes de "data science" formulats usant software d'optimització i d'aprenentatge automàtic (scikit-learn, tensorflow).
- * Saber què és el camp del control de la revelació estadística o protecció de dades estadístiques.
- * Conèixer software per a protecció de dades i ser capaç de protegir dades usant alguna tècnica existent.
- * Familiaritzar-se amb la literatura d'optimització en "data science".

HORES TOTALS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	80,0	64.00
Hores grup gran	30,0	24.00
Hores grup petit	15,0	12.00

Dedicació total: 125 h

CONTINGUTS

Optimització entera i combinatòria en problemes estadístics i de "data science".

Descripció:

Conceptes bàsics d'optimització entera i combinatòria. Modelització de problemes d'optimització entera i combinatoria.

Aplicacions: disseny d'experiments (quadrats llatins ortogonals), clustering òptim (k-medoids), clustering heurístic (k-means).

Dedicació: 15h

Grup gran/Teoria: 10h

Grup mitjà/Pràctiques: 5h

Optimització contínua en problemes de data science.

Descripció:

Regressió ridge i LASSO. Support Vector machines (SVMs): formulació primal, condicions d'optimalitat KKT, dualitat en SVMs, formulació dual, mètodes d'optimització per a SVMs, software per a SVMs. Xarxes neuronals: estructura i modelització de XNs com a problema d'optimització, mètodes d'optimització per a XNs, software per a XNs (scikit-learn, tensorflow).

Dedicació: 15h

Grup gran/Teoria: 5h

Grup mitjà/Pràctiques: 10h

Introducció a la protecció de dades estadístiques.

Descripció:

Introducció. Definicions. Tipus de dades i mètodes. Mètodes de protecció per a microdades. Mètodes de protecció per a dades tabulars. Software de protecció de dades.

Dedicació: 15h

Grup gran/Teoria: 10h

Grup mitjà/Pràctiques: 5h

SISTEMA DE QUALIFICACIÓ

Un examen parcial de la primera part de l'assignatura (40% de la nota) i realització de treballs pràctics (60% de la nota).

BIBLIOGRAFIA

Bàsica:

- Cristianini, Nello; Shawe-Taylor, John. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press, 2000. ISBN 0521780195.
- Willenborg, Leon; Waal, Ton de. Elements of statistical disclosure control. New York: Springer, 2001. ISBN 0387951210.
- Arthanari, T.S. Mathematical programming in statistics. Wiley, 1993. ISBN 0471592129.
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron. Deep learning [en línia]. Cambridge, Massachusetts: The MIT Press, [2016] [Consulta: 05/07/2023]. Disponible a: <https://www.deeplearningbook.org/>. ISBN 9780262035613.
- Nocedal, Jorge; Wright, Stephen J. Numerical optimization [en línia]. New York: Springer, 1999 [Consulta: 05/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-40065-5>. ISBN 9780387987934.
- Luenberger, David G; Ye, Yinyu. Linear and nonlinear programming [en línia]. 3rd ed. New York: Springer, cop. 2008 [Consulta: 05/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-74503-9>. ISBN 0387745033.