

Guia docent

200644 - APE - Aprenentatge Estadístic

Última modificació: 09/05/2024

Unitat responsable: Facultat de Matemàtiques i Estadística
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.
1004 - UB - Universitat de Barcelona.

Titulació: MÀSTER UNIVERSITARI EN ESTADÍSTICA I INVESTIGACIÓ OPERATIVA (Pla 2013). (Assignatura optativa).

Curs: 2024 **Crèdits ECTS:** 5.0 **Idiomes:** Castellà

PROFESSORAT

Professorat responsable: PEDRO FRANCISCO DELICADO USEROS

Altres: Segon quadrimestre:
PEDRO FRANCISCO DELICADO USEROS - A
ÀLEX SÁNCHEZ PLA - A

CAPACITATS PRÈVIES

Familiaritat amb els conceptes bàsics de càlcul en una o més variables (materials per preparació prèvia:). Formació de nivell mitja en probabilitats i inferència. Domini de l'entorn de treball estadístic i programació R (materials per preparació prèvia: qualsevol bon curs d'autoaprenentatge de R). És útil tenir nocions bàsiques de regressió lineal múltiple i de regressió logística.

REQUISITS

"Fundamentos de Inferencia Estadística" o "Inferencia Estadística Avanzada"
"Software estadístic: R i SAS"

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

MESIO-CE2. 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.

MESIO-CE3. 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.

MESIO-CE6. CE-6. Capacitat per a fer servir el software més adequat per a realitzar els càlculs necessaris a la resolució d'un problema.

MESIO-CE8. CE-8. Capacitat de discutir la validesa, l'abast i la rellevància d'aquestes solucions i saber presentar i defensar les conclusions.

MESIO-CE9. CE-9. Capacitat per a implementar algoritmes d'estadística i investigació operativa.

MESIO-CE4. CE-4. Capacitat de fer servir els diferents procediments d'inferència per a respondre preguntes, identificant les propietats dels diferents mètodes d'estimació i els seus avantatges i inconvenients, adaptats a una situació concreta i en un context específic.

Transversals:

CT1a. 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.

CT3. 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.

METODOLOGIES DOCENTS

L'aprenentatge s'organitza en sessions teòric-pràctiques, que combinen una part expositiva i una altra d'il·lustrativa. A la part expositiva de les sessions, els aspectes teòrics es presenten i discuteixen, acompanyats d'exemples pràctics utilitzant diapositives que es lliuraran prèviament als estudiants. A la part il·lustrativa, l'entorn de treball fonamental serà R, del qual es presumeix un coneixement intermedi (ús de l'entorn i programació bàsica). També es farà servir Python en algunes sessions.

L'aprenentatge autònom consistirà en l'estudi i resolució de problemes pràctics que l'estudiant ha de lliurar al llarg del curs.

Concretament, les activitats previstes són:

- Anàlisi detallada de diversos conjunts de dades. S'intentarà que cada conjunt de dades serveixi de base per a un cas d'estudi en diversos mètodes.
- La realització d'exercicis pràctics sobre els mètodes estudiats. Els exercicis pràctics requeriran completar les tasques de programació en R.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

Conèixer l'estructura dels problemes d'aprenentatge supervisats i no supervisats.

Ser capaç d'ajustar un model de regressió lineal múltiple, i també un GLM, utilitzant la versió penalitzada dels mínims quadrats ordinaris (OLS) i dels estimadors de màxima versemblança.

Conèixer les característiques comunes essencials dels estimadors de regressió no paramètrics (disjuntiva biaix-variància, selecció del paràmetre de suavitzat, nombre efectiu de paràmetres, etc.) i els detalls de tres d'ells: regressió polinòmica local, suavitzat per splines, models additius generalitzats (GAM).

Conèixer els principals mètodes basats en arbres i poder aplicar aquests mètodes en conjunts de dades reals.

Comprendre els fonaments de les xarxes neuronals artificials (incloent-hi models de deep-learning i les xarxes neuronals convolucionals) i adquirir les habilitats necessàries per aplicar-les

Conèixer els principals procediments de validació creuada per avaluar la precisió d'un model de predicció.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANT

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

Dedicació total: 125 h

CONTINGUTS

Introducció a l'aprenentatge estadístic

Descripció:

1. Aprenentatge supervisat i no supervisat.

Dedicació: 1h 30m

Grup gran/Teoria: 1h

Grup petit/Laboratori: 0h 30m

Estimadors de regressió penalitzats: Regressió ridge i Lasso

Descripció:

1. Regressió ridge.
2. Validació creuada.
3. Lasso en el model de regressió lineal múltiple. Optimització cíclica coordinada a coordinada.
4. Lasso al GLM.
5. Comparació de les regles de classificació. Corba ROC.

Dedicació: 6h

Grup gran/Teoria: 4h

Grup petit/Laboratori: 2h

Regressió no paramètrica. Models Additius Generalitzats

Descripció:

1. Introducció al modelatge no paramètric.
2. Regressió polinòmica local. La disjuntiva biaix-variància. Suavitzaadors lineals. Selecció del paràmetre de suavitzat.
3. Regressió no paramètrica amb resposta binària. Model de regressió no paramètrica generalitzat. Estimació per màxima versemblança local.
4. Suavitzat per splines. Regressió no paramètrica de mínims quadrats penalitzada. Splines cúbics, interpolació i suavitzat. B-splines. Ajust de models de regressió no paramètrics generalitzats amb splines.
5. Models d'additius generalitzats (GAM). Regressió no paramètrica múltiple. La maledicció de la dimensionalitat. Models additius. Models additius generalitzats.

Dedicació: 15h

Grup gran/Teoria: 10h

Grup petit/Laboratori: 5h

Mètodes basats en arbres

Descripció:

1. Els fonaments dels arbres de decisió. Arbres de regressió. Arbres de classificació.
2. Ensemble Learning. Bagging. Random forests. Boosting.

Dedicació: 10h 30m

Grup gran/Teoria: 7h

Grup petit/Laboratori: 3h 30m

Xarxes Neuronals Artificials

Descripció:

1. Xarxes feed-forward.
2. Entrenament d'una xarxa.
3. Retro-propagació de l'error.
4. Models d'aprenentatge profund (Deep Learning).
5. Xarxes neuronals convolucionals.

Dedicació: 12h

Grup gran/Teoria: 8h

Grup petit/Laboratori: 4h

SISTEMA DE QUALIFICACIÓ

Es basa en dues parts:

- 1) Exercicis pràctics realitzats al llarg del curs: 50%
- 2) Examen final: 50%

BIBLIOGRAFIA

Bàsica:

- Bowman, A. W; Azzalini, Adelchi. Applied smoothing techniques for data analysis : the Kernel approach with S-Plus illustrations. Oxford: Clarendon Press, 1997. ISBN 0198523963.
- Chollet, François; Allaire, J. J. Deep Learning with R. Manning Publications, 2018. ISBN 9781617295546.
- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction [en línia]. 2nd ed. New York [etc.]: Springer, cop. 2009 [Consulta: 04/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7>. ISBN 9780387848570.
- Hastie, Trevor; Tibshirani, Robert; Wainwright, Martin. Statistical learning with sparsity: the lasso and generalizations [en línia]. Boca Raton, FL: Chapman & Hall/CRC, 2015 [Consulta: 04/07/2023]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=4087701>. ISBN 9781498712163.
- James, Gareth. An Introduction to statistical learning : with applications in R [en línia]. New York: Springer, 2013 [Consulta: 04/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-1-4614-7138-7>. ISBN 9781461471370.
- Lantz, Brett. Machine learning with R : discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R [en línia]. 2nd ed. Birmingham: Packt Pub, 2015 [Consulta: 24/05/2024]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=2122139>. ISBN 9781784393908.
- Wood, Simon N. Generalized additive models : an introduction with R. Boca Raton, Fla. [etc.]: Chapman & Hall/CRC, 2006. ISBN 9781584884743.

Complementària:

- Bishop, Christopher M. Pattern recognition and machine learning. New York: Springer, cop. 2006. ISBN 9780387310732.
- Haykin, Simon S. Neural networks and learning machines. 3rd. Upper Saddle River: Prentice Hall, cop. 2009. ISBN 9780131471399.
- Wasserman, Larry. All of nonparametric statistics [en línia]. New York: Springer, 2006 [Consulta: 04/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/0-387-30623-4>. ISBN 9780387251455.
- Fan, Jianqing; Li, Runze; Zhang, Cun-Hui; Zou, Hui. Statistical foundations of data science. Boca Raton;: CRC Press, 2020. ISBN 9781466510845.

RECURSOS

Altres recursos:

ATENEA