

Guia docent

270655 - ASM - Modelat Estadístic Avançat

Última modificació: 20/07/2020

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.

Titulació: MÀSTER UNIVERSITARI EN INNOVACIÓ I RECERCA EN INFORMÀTICA (Pla 2012). (Assignatura optativa).

Curs: 2020 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: JOSEP ANTON SÁNCHEZ ESPIGARES - PEDRO FRANCISCO DELICADO USEROS

Altres: Primer quadrimestre:
PEDRO FRANCISCO DELICADO USEROS - 10
JOSEP ANTON SÁNCHEZ ESPIGARES - 10

CAPACITATS PRÈVIES

No especificades

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

Genèriques:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informàtica.

Transversals:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR5. ACTITUD ADEQUADA DAVANT EL TREBALL: Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

METODOLOGIES DOCENTS

Hi ha una sessió setmanal de 3 hores. Les 2 primeres hores es dediquen a la presentació, per part del professor, dels continguts teòrics de l'assignatura. L'última hora es dedica a posar en pràctica aquests continguts: cada alumne té a classe el seu ordinador portàtil i realitza les tasques que el professor proposa.

Cada sessió finalitza amb la proposta d'una feina als alumnes que han de lliurar resolta la següent sessió.



OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Al final del curs l'estudiant serà capaç de proposar i estimar models de regressió lineal simples i múltiples. També podrà interpretar els models estimats i validar-los.
2. Al final del curs l'estudiant serà capaç de proposar, estimar, interpretar i validar models lineals generalitzats.
3. Al final del curs l'estudiant serà capaç de proposar, estimar, interpretar i validar versions no paramètriques dels models lineals de regressió i dels models lineals generalitzats.
4. Al final del curs l'alumne coneixerà adequadament la forma de triar els paràmetres de suavitzat que en models de regressió no paramètrics controlen l'equilibri entre bon ajust a la mostra observada i bona generalització.
5. Al final de curs l'alumne, enfrontat a un problema real de modelització i/o predicció, sabrà triar el model de regressió més adequat (paramètric, no paramètric o semiparamètric).

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANT

Tipus	Hores	Percentatge
Grup petit/Laboratori	12,0	8.00
Aprenentatge autònom	96,0	64.00
Grup gran/Teoria	24,0	16.00
Grup mitjà/Pràctiques	12,0	8.00
Activitats dirigides	6,0	4.00

Dedicació total: 150 h

CONTINGUTS

Models paramètrics

Descripció:

1. Introducció. Models deterministes i models estadístics. Models paramètrics, no paramètrics i semiparamètrics. La construcció de models estadístics. Exemples. Software.
2. Models lineals normals. Descripció del model lineal normal. Estimació per mínims quadrats. Taula ANOVA. Inferència. Validació del model. L'ús de variables explicatives categòriques. Selecció de model. Predicció. Interpretació de la model i col·linealitat. Regressió robusta i detecció de valors atípics. Model no lineal normal.
3. Models lineals generalitzats. Descripció dels models lineals generalitzats. Models per a la variable de resposta binària. Models per a dades de recompte i les taules de contingència. Models per a dades de resposta de temps de vida. Estimació per màxima versemblança i per mitjà de l'estadístic χ^2 . Inferència. Validació del model.
4. Models bayesians. Inferència Freqüentista i inferència basada en la funció de versemblança. Què és un model bayesià? Distribució a posteriori. Distribució predictiva a priori distribució predictiva a posteriori. La selecció d'una distribució a priori. Inferència bayesiana.



Models no paramètrics

Descripció:

1. Model de regressió no paramètrica. Introducció al modelatge no paramètric. Regressió polinòmica local. El trade-off entre biaix i variància. Nuclis. Suavitadors lineals. Elecció del grau del polinomi local. Elecció del paràmetre de suavitzat: validació creuada, mètode plug-in, finestres variables.
2. Model de regressió no paramètrica generalitzat. Regressió no paramètrica amb resposta binària. Model de regressió no paramètrica generalitzat. Estimació per màxima versemblança local.
3. Inferència en la regressió no paramètrica. Bandes de variabilitat. Proves d'absència d'efectes. Prova d'un model paramètric. Comparació de corbes.
4. Spline suavitzador. Ajust per mínims quadrats penalitzats de la regressió no paramètrica. Splines cúbics i interpolació. Suavitat mitjançant splines. B-splines i P-splines. Regressió spline. Ajust de models de regressió no paramètrics generalitzats, mitjançant splines.
5. Models additius generalitzats i models semiparamètrics. Regressió múltiple no paramètrica. La maledicció de la dimensionalitat. Models additius. Models additius generalitzats. Models semiparamètrics.

ACTIVITATS

Desenvolupament del Tema 1 (models paramètrics de regressió) a classe

Descripció:

Desenvolupament del Tema 1 (models paramètrics de regressió) a classe

Objectius específics:

1, 2, 5

Competències relacionades:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi.

Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

CTR5. ACTITUD ADEQUADA DAVANT EL TREBALL: Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

Dedicació: 75h

Grup gran/Teoria: 22h 30m

Aprenentatge autònom: 52h 30m



Presentació del Tema 2 (models de regressió no paramètrica) a classe

Descripció:

Presentació del Tema 2 (models de regressió no paramètrica) a classe

Objectius específics:

3, 4, 5

Competències relacionades:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi.

Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

CTR5. ACTITUD ADEQUADA DAVANT EL TREBALL: Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

Dedicació: 75h

Grup gran/Teoria: 22h 30m

Aprenentatge autònom: 52h 30m

SISTEMA DE QUALIFICACIÓ

S'assignaran tasques per fer a casa. La nota de les tasques valdrà el 50% de la nota final.

Hi haurà un examen de la primera part de l'assignatura, fet a la setmana de parcials, i un altre examen de la segona part fet com examen final, tots dos amb un pes del 25%.

Nota del curs = $0.5 * \text{Nota Tasques} + 0.25 * \text{Nota Examen 1a part} + 0.25 * \text{Nota Examen 2a part}$

BIBLIOGRAFIA

Bàsica:

- Bowman, A.W.; Azzalini, A. Applied smoothing techniques for data analysis: the Kernel approach with S-Plus illustrations. Clarendon Press, 1997. ISBN 0198523963.

- Wood, S.N. Generalized additive models: an introduction with R. 2nd ed. CRC Press/Taylor & Francis Group, 2017. ISBN 9781498728331.

- Chambers, J.M.; Hastie, T.J. (eds.). Statistical models in S. Chapman & Hall, 1992. ISBN 041283040X.

- Hastie, T.J.; Tibshirani, R.J. Generalized additive models. Chapman and Hall, 1990. ISBN 0412343908.

- Hastie, T.J.; Tibshirani, R.J.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction [en línia]. 2nd ed. Springer, 2009 Disponible a: <http://dx.doi.org/10.1007/978-0-387-84858-7>. ISBN 9780387848570.

- McCullagh, P.; Nelder, J.A. Generalized Linear Models. 2nd ed. Chapman and Hall, 1989. ISBN 0412317605.

- Ruppert, D.; Wand, M.P.; Carroll, R.J. Semiparametric regression. Cambridge University Press, 2003. ISBN 0521785162.

- Venables, W.N.; Ripley, B.D. Modern applied statistics with S-PLUS. 3rd ed. Springer-Verlag, 1999. ISBN 0387988254.

- Wasserman, L. All of nonparametric statistics [en línia]. Springer, 2010 Disponible a: <http://dx.doi.org/10.1007/0-387-30623-4>. ISBN 9781441920447.

- Weisberg, S. Applied linear regression [en línia]. 4th ed. Wiley, 2014 Disponible a: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=1574352>. ISBN 9781118594858.