Course guides
270655 - ASM - Advanced Statistical Modelling

Last modified: 12/07/2021

Unit in charge: Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.

Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).

Academic year: 2021 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer: JOSE ANTONIO SÁNCHEZ ESPIGARES

Others: Primer quadrimestre:
XAVIER PUIG ORIOL - 10
JOSE ANTONIO SÁNCHEZ ESPIGARES - 10

PRIOR SKILLS

Not specified

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.

Generical:
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

TEACHING METHODOLOGY

There is a weekly 3 hours session. The first two hours are devoted to the exposition of the theoretical subjects by the teacher. The last hour is dedicated to implement these contents: Each student has his laptop in class and he or she performs the tasks proposed by the teacher. Each session ends with an assigment to students who must be delivered the following session.

LEARNING OBJECTIVES OF THE SUBJECT

1.At the end of the course the student will be able to propose, estimate, interpret and validate generalized linear models.
2.At the end of the course the student will be able to propose, estimate, interpret and validate non-parametric versions of linear regression models and generalized linear models.
3.At the end of the course the student will know properly how to choose the smoothing parameters which in nonparametric regression models control the trade-off between good fit to the observed sample and good generalization.
4.At the end of the course the student, facing a real problem of modeling and / or prediction, will know to choose the most suitable regression model (parametric, non-parametric, semi-parametric or Bayesian).
5.At the end of the course the student will be able to distinguish the difference between Bayesian and non-Bayesian statistical modelling
6.At the end of the course the student will be able to define a prior distribution, and go from prior to posterior distributions
7.At the end of the course the student will be able to understand the difference between hierarchical and non-hierarchical Bayesian models
8.At the end of the course the student will be able to check a Bayesian model, compare Bayesian models and use them for prediction
9.At the end of the course the student will be able to simulate from the posterior distribution by means of the suitable software

STUDY LOAD

Type	Hours	Percentage
Hours large group	54,0	36.00
Self study	96,0	64.00

Total learning time: 150 h

CONTENTS

Parametric Modelling

Description:
1. Introduction. Deterministic models and statistical models. Parametric, nonparametric and semiparametric models.

2. Generalized linear models. Models for binary response data. Models for count data and contingency tables. Estimation by maximum likelihood and through the Xi^2 statistic. Inference. Model checking.

3. Regularized estimation of LM and GLM. Ridge regression. LASSO estimation

Nonparametric Modelling

Description:
1. Nonparametric regression model. Local polynomial regression. Kernels. Linear smoothers. Choosing the smoothing parameter: Cross validation, plug-in methods, varying windows.

2. Generalized nonparametric regression model. Estimation by maximum local likelihood.

3. Inference with nonparametric regression. Variability bands. Testing for no effects. Checking a parametric model. Comparing curves.

4. Spline smoothing. Penalized least squares nonparametric regression. Cubic splines and interpolation. Smoothing splines. B-splines and P-splines. Fitting generalized nonparametric regression models with splines.

5. Generalized additive models and Semiparametric models. Multiple nonparametric regression. The curse of dimensionality. Generalized additive models. Semiparametric models.

Bayesian Data Analysis

Description:
1. Bayesian Model. The statistical model. The Likelihood function. The Bayesian model

2. Bayesian Inference. Point and Interval estimation.Hypothesis Test

3. Bayesian Computation. Markov Chain Montecarlo simulation. Monitoring convergence

4. Hierarchical Models

5. Checking and defining the model

ACTIVITIES

Presentation of Theme 1 (parametric regression models) in class

Description:
Presentation of Theme 1 (parametric regression models) in class

Specific objectives:
1, 4

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.

Full-or-part-time: 20h
Theory classes: 6h
Self study: 14h

Presentation of Theme 2 (non-parametric regression models) in class

Description:
Presentation of Theme 2 (non-parametric regression models) in class

Specific objectives:
2, 3, 4

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.

Full-or-part-time: 55h
Theory classes: 16h 30m
Self study: 38h 30m

Presentation of theme 3 (Bayesian models) in class

Description:
Presentation of theme 3 (Bayesian models) in class

Specific objectives:
4, 5, 6, 7, 8, 9

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.

Full-or-part-time: 75h
Theory classes: 22h 30m
Self study: 52h 30m

GRADING SYSTEM

Homeworks will be assigned during the course. Homework grades will be worth 50% of your course grade.

There will be an exam for the first part of the course (first and second themes), during the partial exams week, and another one for the second part (third theme), each one with a weight of 25%.

Course Grade = 0.5 * Hwk Grade + 0.25 * 1st part Exam Grade + 0.25 * 2nd part Exam Grade

BIBLIOGRAPHY

Basic:
- Weisberg, S. Applied linear regression [on line]. 4th ed. Hoboken, New Jersey: Wiley, 2014 [Consultation: 14/07/2021]. Available on: https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=1574352. ISBN 9781118594858.
- Chambers, J.M.; Hastie, T.J. (eds.). Statistical models in S. London: Chapman & Hall, 1992. ISBN 041283040X.
- Bolstad, William M. Introduction to Bayesian statistics. 2nd ed. Hoboken, N.J: John Wiley, 2007. ISBN 9780470141151.
- Gelman, Andrew. Bayesian data analysis [on line]. 3rd. ed. London, [etc.]: Chapman & Hall, 2014 [Consultation: 14/07/2021]. Available on: https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=1438153. ISBN 9781439840955.
- Kruschke, John K. Doing Bayesian data analysis : a tutorial with R, JAGS, and Stan. 2nd ed. Amsterdam: Academic Press, 2015. ISBN 9780124058880.
- McCullagh, P.; Nelder, J.A. Generalized Linear Models. 2nd. ed. London [etc.]: Chapman and Hall, 1989. ISBN 0412317605.
- Venables, W.N.; Ripley, B.D. Modern applied statistics with S-PLUS. 3rd. ed. New York: Springer-Verlag, 1999. ISBN 0387988254.
- Hastie, T.J.; Tibshirani, R.J.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction [on line]. 2nd. ed. New York [etc.]: Springer, 2009 [Consultation: 14/07/2021]. Available on: http://dx.doi.org/10.1007/978-0-387-84858-7. ISBN 9780387848570.
- Hastie, T.J.; Tibshirani, R.J. Generalized additive models. Chapman and Hall, 1990. ISBN 0412343908.
- Bowman, A.W.; Azzalini, A. Applied smoothing techniques for data analysis: the Kernel approach with S-Plus illustrations. Oxford: Clarendon Press, 1997. ISBN 0198523963.
- Ruppert, D.; Wand, M.P.; Carroll, R.J. Semiparametric regression. Cambridge University Press, 2003. ISBN 0521785162.
- Wasserman, L. All of nonparametric statistics [on line]. Springer, 2010 [Consultation: 14/07/2021]. Available on: http://dx.doi.org/10.1007/0-387-30623-4. ISBN 9781441920447.
- Wood, S.N. Generalized additive models: an introduction with R. 2nd ed. Boca Raton: CRC Press/Taylor & Francis Group, 2017. ISBN 9781498728331.