UAB
Universitat Autònoma de Barcelona

2018/2019

Data Visualisation and Modelling

Code: 43482
ECTS Credits: 6

Degree	Type	Year	Semester
4313136 Modelling for Science and Engineering	OT	0	1

Contact

Name: Pere Puig Casado

Email: Pere.Puig@uab.cat

Teachers

Rosa Camps Camprubí

Rosario Delgado de la Torre

Juan Ramón González Ruíz

Use of languages

Principal working language: english (eng)

Prerequisites

An elementary knowledge in Probability Theory and Statistical Inference.

Objectives and Contextualisation

Course of R All the practical exercises will be solved using the statistical package R. This introductory course is basic for all the posterior developements.

Visualization of large-scale datasets with R GViz, Maps and Tabplot.

Data Simulation, Boostrapping and Permutation testing These methodologies allow a fast solution to complex statistical models without a deep knowledge of the general and classical statistical topics. They are indispensable tools in the current statistical modelling techniques. The students will complete a basic training program, including the use of an appropriate software, and they will learn how to attack several real data analysis problems.

Bayesian networks, in the opinion of many researchers one of the most significant contribution in AI in this century, are graphical structures for representing the probabilistic relationships among a large number of variables and for doing probabilistic inference with those variables, with a huge number of application fields. One of the objectives of this course is to introduce them and develop in students some skill in their use in modelling, both from a theoretical and applied point of view, with particular emphasis on the use of appropriate software.

Skills

- Analyse complex systems in different fields and determine the basic structures and parameters of their workings.
- Analyse, synthesise, organise and plan projects in the field of study.

- Apply logical/mathematical thinking: the analytic process that involves moving from general principles to particular cases, and the synthetic process that derives a general rule from different examples.
- Apply specific methodologies, techniques and resources to conduct research and produce innovative results in the area of specialisation.
- Apply techniques for solving mathematical models and their real implementation problems.
- Conceive and design efficient solutions, applying computational techniques in order to solve mathematical models of complex systems.
- Formulate, analyse and validate mathematical models of practical problems in different fields.
- Isolate the main difficulty in a complex problem from other, less important issues.

Learning outcomes

1. Analyse, synthesise, organise and plan projects in the field of study.
2. Apply logical/mathematical thinking: the analytic process that involves moving from general principles to particular cases, and the synthetic process that derives a general rule from different examples.
3. Apply specific methodologies, techniques and resources to conduct research and produce innovative results in the area of specialisation.
4. Apply temporary-series techniques to predict the behaviour of certain phenomena.
5. Apply temporary-series techniques to study models associated with practical problems.
6. Choose the best description of a system on the basis of its particular characteristics
7. Identify the parameters that determine how a system works.
8. Implement the proposed solutions reliably and efficiently.
9. Isolate the main difficulty in a complex problem from other, less important issues.
10. Recognise problems that require the use of temporary-series techniques to build models associated with practical problems.
11. Use specific software to solve optimisation problems..

Content

Part 1: Introduction to R (6h)

Part 2: Visualization of large-scale datasets with R (6h)

Part 3: Data Simulation, Boostrapping and Permutation testing (12h)

1) Permutation tests.

2) Jackknife.

3) Parametric Bootstrap.

4) Non-parametric Bootstrap.

Part 4: Bayesian Networks (14h)

1) Block 1: Basics.

2) Block 2: Causal networks and Inference in Bayesian networks.

3) Block 3: Learning Bayesian network parameters.

Methodology

In this course lectures, in which the determining factor is the teacher's explanation, are the basis of the learnig process. It is also very important the participation of the students, combined with practical sessions in which it is the student him/herself who must use the knowledge to solve problems.

Activities

Title	Hours	ECTS	Learning outcomes
Type: Directed			
Exercises	16	0.64	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Lectures	38	1.52	4, 5, 6, 7, 8, 10
Projects + Assigments	18	0.72	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Type: Supervised			
Practical sessions	20	0.8	4, 5, 6, 7, 8, 10

Evaluation

The evaluation of the course consists in a continuous assessment.

There are 3 assessments during the course, weighted as 20%, 40%, 45% corresponding to each part.

Each professor will explain his or her own type of assessment.

Part 1 assessment: Daily homework + final project (individual simple real data analysis with R).

Part 2 assessment: Daily homework + final project

Part 3 assessment: Daily homework + delivery of some exercises.

Part 4 assessment: Daily homework + delivery of some exercises + exam.

Evaluation activities

Title	Weighting	Hours	ECTS	Learning outcomes
Daily homework	50	38	1.52	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Projects + Exam	50	20	0.8	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Bibliography

- Resampling methods: a practical guide to data Analysis. Phillip I. Good, 2006.
- The jackknife, the bootstrap and other resampling plans. Bradley Efron, 1982.
- Bootstrap methods and their application. A.C. Davison, D.V. Hinkley, 1997.
- "Learning Bayesian Networks" by R. E. Neapolitan, Prentice Hall Series in Artificial Intelligence, 2004.
- "Probabilistic Methods for Bioinformatics with an Introduction to Bayesian Networks" by R. E. Neapolitan, Elsevier, 2009.