UAB
Universitat Autònoma de Barcelona

Complex Data Modelling

2026/2027

Code: 104864
Credits: 6

Degree programme	Type	Course
Applied Statistics	OB	3

Contact lecturer

Name : Rosario Delgado De la Torre

Email : rosario.delgado@uab.cat

Teaching staff

Rosario Delgado De la Torre

Group languages

You can consult this information at the end of the document.

Prerequisites

It is assumed that the student taking this subject has acquired the skills of the subjects of

- Càlcul 1,
- Eines informàtiques per a l'Estadística i Introducció a la Programació,
- Introducció a la Probabilitat i Inferència Estdística 1, i
- Aprenentatge Automàtic 1.

You will need a good level and practice in programming with R.

Objectives

Learn what Bayesian Networks (BN) are and how they are used: BN are a probabilistic model used in Supervised Machine Learning that describe the probabilistic relationships between variables that affect a given phenomenon of interest (which can be a complex system) and can be used as classifiers.

Understand how Bayesian Networks are used to assess and quantify risks, among other applications.

Know different methodologies that will have to be applied, or not, when working with these models, in the pre-process phase of the database depending on its characteristics or in the construction phase of the predictive model.

Know different behavioral metrics to validatethemodel and understand its usefulness and adequacy, depending on the characteristics of the database.

Learn how to build R scripts that allow you to learn these models from a database and do their validation, using the relevant libraries. Apply it with real data.

Learning outcomes

- CM09 (Assess the suitability of the models with the correct use and interpretation of indicators and graphs.) Assess the suitability of the models with the correct use and interpretation of indicators and graphs.
- KM12 (Provide the experimental hypotheses of modelling, considering the technical and ethical implications involved.) Provide the experimental hypotheses of modelling, considering the technical and ethical implications involved.
- SM12 (Interpret the results obtained to formulate conclusions about the experimental hypotheses.) Interpret the results obtained to formulate conclusions about the experimental hypotheses.
- SM13 (Compare the degree of adjustment between diverse statistical models.) Compare the degree of adjustment between diverse statistical models.
- SM14 (Use graphs to visualise the fit and suitability of the model.) Use graphs to visualise the fit and suitability of the model.

Contents

2. Introduction to Bayesian Networks (BNs).
   Definition.
   Inference with BNs.
   Learning BNs (both structure and parameters).

4. BNs as classifiers.
   Classification versus regression.
   Bayesian classifiers and the MAP criterion.
   Optimality of Naive Bayes.
   Types of classification taks.

6. Validation and behavioral metrics.
   Cross-validation.
   Metrics for the binary case in general. The unbalanced case.
   Metrics for the multiclasse case.
   Metrics for the ordinal case.

8. Other aspects.
   Multi-dimensional classification.
   MAP criterion and ordinal classification.
   Ensemble of classifiers.
   Dinamic BNs.

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
Practical exercices with computer tools	30	1.2	
Theory	26	1.04	
Problems	14	0.56	
Tutorials	10	0.4	
Study and think problems	40	1.6	
Practices (deliveries, controls)	12	0.48	

The subject is structured around theoretical classes, problems and practices. The follow-up of the subject is face-to-face, but it will be necessary to extend the teacher's explanations with the student's autonomous study, with the support of the reference bibliography and the material provided by the teacher.

The problem class will focus on solving some of the proposed problems. In the practical classes we will work with R and his libraries. Student participation in problem and practice classes will be especially valued.

Annotation: within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continuous assessment activities

Title	Weight	Hours	ECTS	Learning outcomes
PAC1	25%	3	0.12	CM09, KM12, SM12, SM13, SM14
PAC2	15%	9	0.36	KM12, SM12, SM13
Problems Exam	50%	4	0.16	KM12, SM12, SM13
Practice Exam	10%	2	0.08	KM12, SM12, SM13, SM14

The final grade for this course is obtained as the weighted average of the following components:

- PAC1 (25%)
- PAC2 (15%)
- Problems Exam (50%)
- Practice Exam (10%)

The continuous assessment activities are:

- PAC1 (a problem-solving test), and
- PAC2 (problems/practical exercises in R, whose specific content will be defined throughout the course and which will be completed and submitted during the face-to-face classes over the semester).

Only grades of 3.5 out of 10 or higher will be considered in the calculation of the weighted average. Any grade below this threshold will contribute 0 to the weighted average.

To pass the course, this weighted average must be at least 5.0 out of 10.

Students who do not pass the course in the first assessment period may sit the resit examination.

For students taking the resit examination, the resit exam accounts for 85% of the final grade. Only students who have failed the course in the first assessment period are eligible to take the resit examination. The resit examination may not be used to improve the grade of students who have already passed the course.

The remaining 15% corresponds to the grade obtained in PAC2, which is not recoverable.

A student will be considered assessable if they have taken the PAC1 test, submitted at least one PAC2 assignment, or sat any of the examinations.

Otherwise, the student will be recorded as "Not Assessed".

For the possible award of Honours Distinction (Matrícula d'Honor), grades obtained in the resit examination period will not be taken into account.

USE OF ARTIFICIAL INTELLIGENCE (AI): In this course, the use of Artificial Intelligence (AI) technologies is permitted exclusively for support tasks, such as literature or information searches, text revision, or translation. Students must clearly identify any parts of their submitted work that have been generated using AI technologies, specify the tools used, and include a critical reflection on how these tools have influenced both the process and the final outcome of the activity. Lack of transparency regarding the use of AI will be considered a breach of academic integrity and may result in a partial or total penalty in the grade for the activity, or more severe disciplinary measures in serious cases.

Any irregularity committed in an assessment activity (academic fraud, plagiarism, or improper use of Artificial Intelligence, unless such use is expressly authorized in the course syllabus) that may lead to a significant alteration of the grade will result in that assessment activity being graded with a 0 (fail).

If the course syllabus establishes that obtaining a minimum grade in that assessment activity is an essential requirement to pass the course, or if multiple irregularities are detected in the assessment activities of the same course, the final grade for the course will be 0 (fail).

In addition, disciplinary proceedings may be initiated against any student who commits any of these irregularities.

Bibliography

- Norman Fenton and Martin Neil, "Risk Assessment and Decision Analysis with Bayesian Networks", CRC Press. A Chapman & Hall Book, 2013. (Disponible en línia)
- Radhakrishnan Nagarajan, Marco Scutari and Sophie Lèbre, "Bayesian Networks in R with applications in Systems Biology", Springer, 2013. (Disponible en línia)
- Oliver Pourret, Patrick Naïm and Bruce Marcot, "Bayesian Networks. A practical guide to Applications", Series: Statistics in Practice. Wiley, 2008. (Disponible en línia)
- Richard E. Neapolitan, "Learning Bayesian Networks", Prentice Hall Series in Artificial Intelligence, 2004.
- Adnan Darwiche, "Modeling and reasoning with Bayesian networks", Cambridge, 2009.
- Kevin B. Korb and Ann E. Nicholson, "Bayesian Artificial Intelligence" (2nd edition), Series: Computer Science and Data Analysis. CRC Press. A Chapman & Hall book, 2011. (Disponible en línia)
- Daphne Koller and Nir Friedman, "Probabilistic Graphical Models", The MIT Press Cambridge, Massachusetts London, England, 2009. http://mcb111.org/w06/KollerFriedman.pdf
- Radhakrishnan Nagarajan, Marco Scutari and Sophie Lèbre, "Bayesian Networks in R with applications in systems biology", Series: Use R! Springer, 2013. (Disponible en línia)
- Marco Scutari and Jean-Baptiste Denis, "Bayesian networks with examples in R", Series: Texts in Statistical Science. CRC Press. A Chapman & Hall Book, 2015.

Software

The R software will be used with some libraries that will be indicated in due course throughout the course. Preferably in the RStudio environment.

Course groups and languages

The information provided is provisional until November 30. After this

date, you will be able to consult the language of each group through this link. To access the information, you will need to enter the course CODE

Type of teaching	Group	Language	Semester	Shift
(TE) Theory	1	Catalan	second semester	afternoon
(PLAB) Practical laboratories	1	Catalan	second semester	afternoon