

Degree	Type	Year
2503852 Applied Statistics	OB	3

Contact

Name: Rosario Delgado De la Torre

Email: rosario.delgado@uab.cat

Teachers

Rosario Delgado De la Torre

Teaching groups languages

You can view this information at the [end](#) of this 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 Estadística 1, i
- Aprenentatge Automàtic 1.

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

Objectives and Contextualisation

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 validate the model 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

1. CM09 (Competence) Assess the suitability of the models with the correct use and interpretation of indicators and graphs.
2. CM09 (Competence) Assess the suitability of the models with the correct use and interpretation of indicators and graphs.
3. CM10 (Competence) Modify the existing software if required by the statistic model, or create new software, if necessary.
4. KM12 (Knowledge) Provide the experimental hypotheses of modelling, considering the technical and ethical implications involved.
5. KM12 (Knowledge) Provide the experimental hypotheses of modelling, considering the technical and ethical implications involved.
6. SM12 (Skill) Interpret the results obtained to formulate conclusions about the experimental hypotheses.
7. SM13 (Skill) Compare the degree of adjustment between diverse statistical models.
8. SM14 (Skill) Use graphs to visualise the fit and suitability of the model.

Content

1. Introduction to Bayesian Networks (BNs).
 Definition.
 Inference with BNs.
 Learning BNs (both structure and parameters).
2. The BNs as classifiers.
 The classification task within Supervised Machine Learning.
 The MAP criterion.
 Types of BN (Naive Bayes, Augmented Naive, TAN).
 Classification type: binary, multi-class, multi-label.
3. Validation and behavioral metrics.
 Cross-validation.
 Metrics for the binary and multi-class case.
 Metrics for the case of ordinal classification.
4. Other aspects.
 Multi-dimensional classification.
 Ensemble of classifiers.
 "Concept drive" and dynamic BNs.
 Gaussian and hybrid BNs.
 Multi-instance classification.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Practices (deliveries, controls)	12	0.48	
Problems	14	0.56	

Theory	26	1.04
Type: Supervised		
Tutorials	10	0.4
Type: Autonomous		
Practical work with computer tools	30	1.2
Study and think problems	40	1.6

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	Weighting	Hours	ECTS	Learning Outcomes
Exam	60%	3	0.12	CM09, SM12, SM13
PAC1	20%	6	0.24	CM09, CM10, KM12, SM12, SM13, SM14
PAC2	20%	9	0.36	CM09, SM13

The final grade for this subject is obtained as the weighted average of the grades of:

- PAC1 (20%)
- PAC2 (20%)
- Exam (60%)

The PAC1 and PAC2 continuous assessment tests consist of a delivery of problems/practical exercises/work with R, which will be specified throughout the course, and in their development in face-to-face classes throughout the semester.

Only those notes that are at least 3.5 out of 10 will be taken into account in the calculation of the weighted average (those that do not comply will weight 0).

To pass the subject it is necessary that this average is at least 5.0 out of 10. In case of not passing the subject in the first call, the student can present himself for recovery. The retake exam represents 100% of the final grade for those students who take the retake, which can only be students who have not passed the subject on the first call (the retake exam does not serve to improve the grade for students who have already passed).

The student who has presented the PAC1 or PAC2 deliveries, or has presented the exam or the recovery exam will be considered evaluable. Otherwise, it will be recorded in the minutes as Not Assessable.

For the eventual assignment of Honors, the marks of the second call will not be taken into account.

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.

Language list

Name	Group	Language	Semester	Turn
(PLAB) Practical laboratories	1	Catalan	second semester	afternoon
(TE) Theory	1	Catalan	second semester	afternoon