Course guide
200644 - APE - Statistical Learning

Unit in charge: School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.
1004 - UB - (ENG)Universitat de Barcelona.

Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Optional subject).

Academic year: 2022  ECTS Credits: 5.0  Languages: Spanish, English

LECTURER

Coordinating lecturer: PEDRO FRANCISCO DELICADO USEROS

Others: Segon quadrimestre:
PEDRO FRANCISCO DELICADO USEROS - A
FERRAN REVERTER COMES - A
ESTEBAN VEGAS LOZANO - A

L'idioma d'impartició d'aquesta assignatura canvia depenent del professor
El idioma de impartición de esta asignatura cambia dependiendo del profesor que la imparta
The teaching language of this subject depends on the professor who teaches it

PRIOR SKILLS

Familiarity with the foundations of calculus in one and more variables. Intermediate studies in probability and inference. Skills using the R environment for statistical computing and programming. Any good online R course may help.

REQUIREMENTS

"Fundamentos de Inferencia Estadística" o "Inferencia Estadística Avanzada"
"Statistical Software: R and SAS"

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
MESIO-CE2. CE-2. Ability to master the proper terminology in a field that is necessary to apply statistical or operations research models and methods to solve real problems.
MESIO-CE3. CE-3. Ability to formulate, analyze and validate models applicable to practical problems. Ability to select the method and / or statistical or operations research technique more appropriate to apply this model to the situation or problem.
MESIO-CE6. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.
MESIO-CE8. CE-8. Ability to discuss the validity, scope and relevance of these solutions and be able to present and defend their conclusions.
MESIO-CE4. CE-4. Ability to use different inference procedures to answer questions, identifying the properties of different estimation methods and their advantages and disadvantages, tailored to a specific situation and a specific context.
Transversal:
CT1a. ENTREPRENEURSHIP AND INNOVATION: Being aware of and understanding how companies are organised and the principles that govern their activity, and being able to understand employment regulations and the relationships between planning, industrial and commercial strategies, quality and profit.

CT3. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.

TEACHING METHODOLOGY
Learning is organized into theoretical-practical sessions with the instructors. All the sessions combine a 50% of expository classes and other 50% of guided practice and workshops.
In the expository part of the sessions, the theoretical aspects are presented and discussed, accompanied by practical examples using slides that will be provided previously to the students.
The fundamental work environment of the practical sessions will be R, of which an intermediate knowledge is presumed (use of the environment and basic programming).

Autonomous learning will consist of the study and resolution of theoretical and practical problems that the student should turn in throughout the course.
Specifically, the planned activities are:
- Study of the learning materials, before and/or after each session with the instructors.
- Detailed analysis of diverse data sets. It will be attempted that each data set serves as a basis for a case study in several methods.
- The completion of theoretical and practical exercises on the studied methods. The practical exercises will require completion of programming tasks in R.

LEARNING OBJECTIVES OF THE SUBJECT
To know the structure of supervised and unsupervised learning problems.
To be able to fit a multiple linear regression model, and also a glm, using penalized version of the standard ordinary least squares (OLS) and maximum likelihood estimators.
To know the essential common characteristics of non-parametric regression estimators (bias-variance trade-off, smoothing parameter choice, effective number of parameters, etc.) and the details of three of them: local polynomial regression, spline smoothing, generalized additive models (GAM).
To know the principal Tree-based Methods and be able to apply these methods in real data sets.
To understand the fundamentals of the of Artificial Neural Networks (including deep-learning models and convolutional neural networks), and to acquire the necessary abilities to apply them.
To know the principal cross-validation procedures for assessing model accuracy.

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours large group</td>
<td>30,0</td>
<td>24.00</td>
</tr>
<tr>
<td>Self study</td>
<td>80,0</td>
<td>64.00</td>
</tr>
<tr>
<td>Hours small group</td>
<td>15,0</td>
<td>12.00</td>
</tr>
</tbody>
</table>

**Total learning time:** 125 h
## CONTENTS

### Introduction to statistical learning

**Description:**
1. Supervised and unsupervised learning.

**Full-or-part-time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m

### Penalized regression estimators: Ridge regression and Lasso

**Description:**
1. Ridge regression.
2. Cross-validation.
3. Lasso estimator in the multiple linear regression model. Cyclical coordinate optimization.
4. Lasso estimator in the GLM.

**Full-or-part-time:** 6h  
Theory classes: 4h  
Laboratory classes: 2h

### Non-parametric regression. Generalized Additive Models

**Description:**
1. Introduction to nonparametric modeling.
2. Local polynomial regression. The bias-variance trade-off. Linear smoothers. Choosing the smoothing parameter.

**Full-or-part-time:** 15h  
Theory classes: 10h  
Laboratory classes: 5h

### Tree-based Methods

**Description:**

**Full-or-part-time:** 10h 30m  
Theory classes: 7h  
Laboratory classes: 3h 30m
Artificial Neural Networks

Description:
1. Feed-Forward Network Functions.
2. Network Training.
3. Error Backpropagation.
5. Convolutional Neural Networks.

Full-or-part-time: 12h
Theory classes: 8h
Laboratory classes: 4h

GRADING SYSTEM

It is based on two parts:
1) Practical exercises done through the course: 50%
2) Final exam: 50%

BIBLIOGRAPHY

Basic:

Complementary:

RESOURCES

Other resources:
ATENEA