200644 - APE - Statistical Learning

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 1004 - UB - (ENG)Universitat de Barcelona
Academic year: 2017
Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Optional)
ECTS credits: 5
Teaching languages: Spanish

Teaching staff

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

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, like http://www.ub.edu/stat/docencia/EADB/Curso%20basico%20de%20R.htm.

Requirements

"Fundamentos de Inferencia Estadística" o "Inferencia Estadística Avanzada"
"Computação en Estadística y en Optimización"

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
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 essentials of Support Vector Machines, and to acquire the necessary abilities for applying these methods in diverse situations with a practical interest.
To understand the fundamentals of the of Artificial 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>Total learning time: 125h</th>
<th>Hours large group:</th>
<th>30h</th>
<th>24.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours small group:</td>
<td>15h</td>
<td>12.00%</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>80h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
## Content

### Introduction to statistical learning

**Learning time:** 1h 30m  
Theory classes: 1h 30m

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

### Penalized regression estimators: Ridge regression and Lasso

**Learning time:** 6h  
Theory classes: 6h

**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.  

### Generalized Additive Models

**Learning time:** 13h 30m  
Theory classes: 13h 30m

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

### Tree-based Methods

**Learning time:** 6h  
Theory classes: 6h

**Description:**  
200644 - APE - Statistical Learning

<table>
<thead>
<tr>
<th>Artificial Neural Networks</th>
<th>Learning time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Theory classes: 9h</td>
</tr>
<tr>
<td>1. Feed-Forward Network Functions.</td>
<td></td>
</tr>
<tr>
<td>2. Network Training.</td>
<td></td>
</tr>
<tr>
<td>3. Error Backpropagation.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Support Vector Machine</th>
<th>Learning time: 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td>1. Maximum Margin Classifier.</td>
<td></td>
</tr>
</tbody>
</table>

**Qualification system**

It is based on two parts:
1) Practical exercises done through the course: 50%
2) Final exam: 50%
200644 - APE - Statistical Learning

Bibliography

Basic:


Complementary:


Others resources:
ATENEA