Course guides
270652 - MVA - Multivariate Analysis

Unit in charge: Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.

Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).

Academic year: 2021  ECTS Credits: 6.0  Languages: English

LECTURER

Coordinating lecturer: DANIEL FERNÁNDEZ MARTÍÑEZ - ARTURO PALOMINO GAYETE

Others:
Primer quadrimestre:
DANTE CONTI - 11, 12
ARTURO PALOMINO GAYETE - 11, 12

Segon quadrimestre:
DANIEL FERNÁNDEZ MARTÍÑEZ - 10
BELCHIN ADRIYANOV KOSTOV - 10

PRIOR SKILLS

The course implies having previously done a basic course in statistics, programming and mathematics; in particular having acquired the following concepts:
- Average, covariance and correlation matrix.
- Hypothesis Test
- Matrix algebra, eigenvalues and eigenvectors.
- Programming algorithms.
- Multiple linear-regression

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEC1. Ability to apply scientific methodologies in the study and analysis of phenomena and systems in any field of Information Technology as well as in the conception, design and implementation of innovative and original computing solutions.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.

Generical:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
TEACHING METHODOLOGY

The course aims to give the statistical foundations for data mining. Learning is done through a combination of theoretical explanation and its application to a real case. The lectures will develop the necessary scientific knowledge, while lab classes will be its application to solving problems of data mining. The implementation of practices fosters generic skills related to teamwork and presentation of results and serve to integrate different knowledge of the subject. The software used will be primarily R.

LEARNING OBJECTIVES OF THE SUBJECT

1. Multivariate description of data
2. Data visualisation
3. Multivariate inference
4. Classification of new individuals

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes</td>
<td>25,5</td>
<td>17.00</td>
</tr>
<tr>
<td>Laboratory classes</td>
<td>25,5</td>
<td>17.00</td>
</tr>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
<tr>
<td>Guided activities</td>
<td>3,0</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h

CONTENTS

Introduction to Multivariate Data Analysis

Description:

Principal Component Analysis

Description:
Analysis of individuals. Analysis of variables. Visual representation of the information. Dimensionality reduction. Supplementary information

Singular Value Decomposition. Biplots

Description:
Method for exploring and visualizing rows and columns of a table through single value decomposition

Factor Analysis

Description:
Dimension reduction method.
### Multidimensional Scaling

**Description:**
This method deals with data relating to distances between elements. Usually uses data from distances or similarities. The method reveals a common structure of all the elements and the specificity of each of them, evidencing what makes them close or distant.

### Hierarchical and Partitioning Clustering

**Description:**
Two approaches to clustering methods used to classify observations, within a data set, into multiple groups based on their similarity.

### Automatic profiling methods

**Description:**
Profiling methods help to understand the common characteristics of clusters.

### Multivariate normal distribution

**Description:**
Particularities of the normal distribution in the general case of multivariate approaches, where the points are distributed in several dimensions.

### Discriminant Analysis

**Description:**
Discriminant Analysis (DA) and Naïve Bayes (NB) are classification methods. DA classifies observations into non-overlapping groups, based on scores on one or more quantitative predictor variables. NB is a simple learning algorithm that utilises Bayes rule together with a strong assumption that the attributes are conditionally independent, given the class.

### Classification and Regression Trees

**Description:**
This method can predict or classify. Explains how the values of a result variable can be predicted or classified based on other values. It has a very useful graphic structure.

### Association rules

**Description:**
Find common patterns, associations, correlations, or causal structures between sets of items or objects in transaction databases, relational databases, and other information repositories.
<table>
<thead>
<tr>
<th>ACTIVITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction to the course + Multivariate Data Analysis</strong></td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
</tr>
<tr>
<td><strong>Full-or-part-time:</strong></td>
</tr>
<tr>
<td>Theory classes:</td>
</tr>
<tr>
<td>Self study:</td>
</tr>
</tbody>
</table>

| **Principal Component Analysis** |
| **Specific objectives:** | 1, 2 |
| **Full-or-part-time:** | 9h |
| Theory classes: | 2h |
| Laboratory classes: | 2h |
| Self study: | 5h |

| **Singular value decomposition** |
| **Specific objectives:** | 1, 2 |
| **Full-or-part-time:** | 9h |
| Theory classes: | 2h |
| Laboratory classes: | 2h |
| Self study: | 5h |

| **Automatic profiling methods** |
| **Specific objectives:** | 1, 2 |
| **Full-or-part-time:** | 9h |
| Theory classes: | 2h |
| Laboratory classes: | 2h |
| Self study: | 5h |

<p>| <strong>Factor Analysis</strong> |
| <strong>Specific objectives:</strong> | 1, 2 |
| <strong>Full-or-part-time:</strong> | 9h |
| Theory classes: | 2h |
| Laboratory classes: | 2h |
| Self study: | 5h |</p>
<table>
<thead>
<tr>
<th>Module</th>
<th>Specific objectives</th>
<th>Full-or-part-time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor Analysis</td>
<td>2, 4</td>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self study: 5h</td>
</tr>
<tr>
<td>Multidimensional Scaling</td>
<td>1, 2</td>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self study: 5h</td>
</tr>
<tr>
<td>Discriminant Analysis</td>
<td>3, 4</td>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self study: 5h</td>
</tr>
<tr>
<td>Classification and Regression Trees</td>
<td>2, 3, 4</td>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self study: 5h</td>
</tr>
<tr>
<td>Hierarchical and Partitioning Clustering</td>
<td>2, 4</td>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self study: 5h</td>
</tr>
</tbody>
</table>
### Multivariate normal distribution

**Specific objectives:**
- 2, 4

**Full-or-part-time:** 9h
- Theory classes: 2h
- Laboratory classes: 2h
- Self study: 5h

### Association rules

**Specific objectives:**
- 4

**Full-or-part-time:** 9h
- Theory classes: 2h
- Laboratory classes: 2h
- Self study: 5h

### Final Practical Work

**Full-or-part-time:** 14h 54m
- Guided activities: 1h 54m
- Self study: 13h

### Quiz

**Full-or-part-time:** 13h 06m
- Self study: 13h 06m

### Summary and Practice. 1st part

**Specific objectives:**
- 1, 2, 3, 4

**Full-or-part-time:** 7h
- Laboratory classes: 2h
- Self study: 5h

### Summary and Practice. 2nd part

**Specific objectives:**
- 1, 2, 3, 4

**Full-or-part-time:** 7h
- Laboratory classes: 2h
- Self study: 5h
**Practice doubts**

**Specific objectives:**
1, 2, 3, 4

**Full-or-part-time:** 2h
Theory classes: 2h

---

**GRADING SYSTEM**

The course evaluation will be based on the marks obtained in practical exercises conducted during the course, a theory grade, and the grade obtained in the final practice.

Each practice will lead to the drafting of the relevant report writing and may be made jointly, up to a maximum of four students per group.

The exercises conducted throughout the course aim to consolidate the learning of multivariate techniques.

The final practice is that students show their maturity to solve a real problem using multivariate visualisation techniques, clustering interpretation, and prediction. Students will choose between different alternatives to solve the problem. This practice will be presented and publicly defended, in which the student must answer any questions about the theoretical models and methods used in the solution. Practices are conducted using the software R.

The written tests will evaluate the assimilation of the basic concepts of the subject. There will be three tests during the curse, in theory class. While the presentation of the practice will be done during the examination period.

The in-class exercises are weighted 20%, theory 40%, and final practice 40%.

---

**BIBLIOGRAPHY**

**Basic:**

**Complementary:**

---

**RESOURCES**

**Hyperlink:**
- [https://cran.r-project.org/](https://cran.r-project.org/)
- [https://rstudio.com/](https://rstudio.com/)