Course guide
270953 - 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 DATA SCIENCE (Syllabus 2021). (Compulsory subject).
Academic year: 2022
ECTS Credits: 6.0
Languages: English

LECTURER

Coordinating lecturer: CARINA GIBERT OLIVERAS - SERGI RAMIREZ MITJANS

Others: Primer quadrimestre:
CARINA GIBERT OLIVERAS - 11
SERGI RAMIREZ MITJANS - 11
DAVID RODRIGUEZ SEGADO - 12

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:
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE5. Model, design, and implement complex data systems, including data visualization
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science

Generic:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CG3. Define, design and implement complex systems that cover all phases in data science projects
Transversal:
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Basic:
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

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>Hours small group</td>
<td>24,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Guided activities</td>
<td>6,0</td>
<td>4.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>24,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular Value Decomposition. Biplots</strong></td>
<td>Method for exploring and visualizing rows and columns of a table through single value decomposition</td>
</tr>
<tr>
<td><strong>Factor Analysis</strong></td>
<td>Dimension reduction method.</td>
</tr>
<tr>
<td><strong>Multidimensional Scaling</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Hierarchical and Partitioning Clustering</strong></td>
<td>Two approaches to clustering methods used to classify observations, within a data set, into multiple groups based on their similarity.</td>
</tr>
<tr>
<td><strong>Automatic profiling methods</strong></td>
<td>Profiling methods help to understand the common characteristics of clusters.</td>
</tr>
<tr>
<td><strong>Multivariate normal distribution</strong></td>
<td>Particularities of the normal distribution in the general case of multivariate approaches, where the points are distributed in several dimensions.</td>
</tr>
<tr>
<td><strong>Discriminant Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Description:</strong></td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Classification and Regression Trees</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Association rules</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Find common patterns, associations, correlations, or causal structures between sets of items or objects in transaction databases, relational databases, and other information repositories.</td>
</tr>
</tbody>
</table>
ACTIVITIES

**Introduction to the course + Multivariate Data Analysis**

**Specific objectives:**
1, 2

**Related competencies:**
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

**Full-or-part-time:** 7h
Theory classes: 2h
Self study: 5h
Principal Component Analysis

Specific objectives:
1, 2

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

Full-or-part-time: 9h
Theory classes: 2h
Laboratory classes: 2h
Self study: 5h
Singular value decomposition

Specific objectives:
1, 2

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

Full-or-part-time: 9h
Theory classes: 2h
Laboratory classes: 2h
Self study: 5h
Automatic profiling methods

Specific objectives:
1, 2

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

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

Specific objectives:
1, 2

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

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

Specific objectives:
2, 4

Related competencies:
CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

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

Specific objectives:
1, 2

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

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

**Specific objectives:**
3, 4

**Related competencies:**
CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

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

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

**Related competencies :**

- CG3. Define, design and implement complex systems that cover all phases in data science projects
- CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
- CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
- CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
- CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
- CE5. Model, design, and implement complex data systems, including data visualization
- CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
- CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

**CT1. ENTREPRENEURSHIP AND INNOVATION:** Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.

**CT4. INFORMATION LITERACY:** Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

**CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.**

**CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.**

**CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.**

**CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.**

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

Specific objectives:
2, 4

Related competencies:
CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

Full-or-part-time: 9h
Theory classes: 2h
Laboratory classes: 2h
Self study: 5h
Multivariate normal distribution

Specific objectives:

2, 4

Related competencies:

CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE5. Model, design, and implement complex data systems, including data visualization
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

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

Specific objectives:
4

Related competencies:
CG3. Define, design and implement complex systems that cover all phases in data science projects
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

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

Related competencies:
CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats.
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation.
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem.
CE5. Model, design, and implement complex data systems, including data visualization.
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

Full-or-part-time: 7h
Laboratory classes: 2h
Self study: 5h
Summary and Practice. 2nd part

Specific objectives:
1, 2, 3, 4

Related competencies:
CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE5. Model, design, and implement complex data systems, including data visualization
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

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

Specific objectives:
1, 2, 3, 4

Related competencies:
CG3. Define, design and implement complex systems that cover all phases in data science projects
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE5. Model, design, and implement complex data systems, including data visualization
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

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:

RESOURCES

Hyperlink:
- https://cran.r-project.org/- https://rstudio.com/