Course guide
270961 - AML - Advanced Machine Learning

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.
Degree: MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).
Academic year: 2022 ECTS Credits: 6.0 Languages: English

LECTURER
Coordinating lecturer: LUIS ANTONIO BELANCHE MUÑOZ
Others: Primer quadrimestre: LUIS ANTONIO BELANCHE MUÑOZ - 10

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
CE3. Apply data integration methods to solve data science problems in heterogeneous data environments
CE5. Model, design, and implement complex data systems, including data visualization
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

Generical:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats

Transversal:
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.

TEACHING METHODOLOGY
The course delves into the most important machine learning paradigms with a solid foundation in probability, statistics and math. The theory is introduced in lectures where the teacher exposes the concepts. These concepts are put into practice in the laboratory classes, in which the student learns to develop machine learning solutions to real problems of a certain complexity.

Students have to work and deliver a term project.
LEARNING OBJECTIVES OF THE SUBJECT

1. Advanced machine learning methods
2. Bayesian statistics
3. Linear models and generalized nonparametric linear models for regression
4. Optimization of neural networks and support vector machines
5. Data cleaning

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours large group</td>
<td>54,0</td>
<td>36.00</td>
</tr>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h

CONTENTS

Introduction to Bayesian machine learning

Description:
Introduction to Bayesian thinking for machine learning. Learning by solving a regularized problem. Illustrative example.

Learning in functional spaces

Description:

Fundamental kernel functions in R^d.

Description:
Description and demonstration of fundamental kernel functions in R^d. Polynomial and Gaussian kernels. General properties of kernel functions.

The support vector machine for classification, regression and novelty detection

Description:
The support vector machine (SVM) is the flagship in kernel methods. Its versions for classification, regression and novelty detection are fully explained and demonstrated.

Kernel functions for different data types

Description:
Some kernel functions for different data types are presented and demonstrated, such as text, trees, graphs, categorical variables, and many others.
<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other kernel-based learning algorithms</td>
<td>Additional kernel-based learning methods are explained, such as kernel PCA and kernel FDA. These are illustrated in several application examples.</td>
</tr>
<tr>
<td>Autoencoders and deep stacking networks</td>
<td>Autoencoders and deep stacking networks: restricted Boltzmann machines and deep belief networks</td>
</tr>
<tr>
<td>Convolutional neural networks and their applications</td>
<td>Successful applications of deep learning in diverse areas of signal and information processing and of applied artificial intelligence.</td>
</tr>
<tr>
<td>Advanced techniques in deep networks and kernel methods</td>
<td>Other methods are briefly introduced, such as the RVM and GPs. Nyström acceleration and Random Fourier features. Deep recurrent networks, deep kernel learning and maybe others.</td>
</tr>
</tbody>
</table>
ACTIVITIES

Theoretical lectures

Specific objectives:
1, 2, 3, 4

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE3. Apply data integration methods to solve data science problems in heterogeneous data environments
CE5. Model, design, and implement complex data systems, including data visualization
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
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.

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.
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.

Full-or-part-time: 60h
Theory classes: 40h
Self study: 20h
Practice lectures

Specific objectives:
1, 4, 5

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE3. Apply data integration methods to solve data science problems in heterogeneous data environments
CE5. Model, design, and implement complex data systems, including data visualization
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
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.

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.
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.

Full-or-part-time: 32h
Laboratory classes: 16h
Guided activities: 2h
Self study: 14h
Final exam

Specific objectives:
1, 2, 3, 4, 5

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE3. Apply data integration methods to solve data science problems in heterogeneous data environments
CE5. Model, design, and implement complex data systems, including data visualization
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
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.

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.
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.

Full-or-part-time: 19h
Guided activities: 3h
Self study: 16h
Term project

Specific objectives:
1, 2, 3, 4, 5

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE3. Apply data integration methods to solve data science problems in heterogeneous data environments
CE5. Model, design, and implement complex data systems, including data visualization
CE11. Analyze and extract knowledge from unstructured information using natural language processing techniques, text and image mining
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE8. Extract information from structured and unstructured data by considering their multivariate nature.
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.

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.
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.

Full-or-part-time: 29h
Guided activities: 4h
Self study: 25h

GRADING SYSTEM

The course is graded as follows:

F = Grade of the final exam
L = Grade of the practical work
S = Grade for the combined soft skills (CB 10 and CB 6)

Final grade = 40% F + 50% L + 10% S

BIBLIOGRAPHY

Basic: