Course guides
270222 - AA2 - Machine Learning 2

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.

Degree: BACHELOR'S DEGREE IN DATA SCIENCE AND ENGINEERING (Syllabus 2017). (Compulsory subject).

Academic year: 2021 ECTS Credits: 6.0 Languages: Catalan

LECTURER

Coordinating lecturer: LUIS ANTONIO BELANCHE MUÑOZ

Others: Primer quadrimestre:
LUIS ANTONIO BELANCHE MUÑOZ - 11
XAVIER GIRÓ NIETO - 11

PRIOR SKILLS

Programming in Python.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

General:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CG3. Work in multidisciplinary teams and projects related to the processing and exploitation of complex data, interacting fluently with engineers and professionals from other disciplines.
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
Transversal:
CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

TEACHING METHODOLOGY

The theory classes introduce all the knowledge, techniques, concepts and results necessary to reach a well-founded level. 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.

In the problem classes we will delve deeper to understand the theory by problem solving or by extending the concepts already seen.

In the laboratory classes, code is provided in various computing environments that allow solving a problem completely with the technique or techniques corresponding to the current topic. This laboratory also serves as a guide for the corresponding part of the practical works developed by the students.

LEARNING OBJECTIVES OF THE SUBJECT

1. Organize the flow of solution of a machine learning problem, analyzing the possible options and elegant the mismatches to the problem
2. Decide, defend and criticise a solution for a machine learning problem, arguing the strong and weak points of appropriation.
3. Know and know how to apply advanced techniques to solve non-supervised learning problems, especially clustering.
4. Know how to apply deep feed-forward multilayer neuronal network techniques to solve complex supervised learning problems.
5. Know and apply advanced techniques of learning methods based on kernel functions, for the resolution of learning problems, both supervised and unsupervised.
6. To know and to know how to apply the advanced techniques for the resolution of learning problems by reinforcement, and its relation with techniques of deep learning.
7. Know how to identify problems involving signal processing, such as data in the form of audio, image or video, or a combination of you, and solve them with advanced computer learning techniques.
8. Know how to identify problems involving human or non-human speech processing, such as data in the form of audio or text, or a combination thereof, and solve them with advanced computational learning techniques.

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours small group</td>
<td>30,0</td>
<td>20.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>30,0</td>
<td>20.00</td>
</tr>
<tr>
<td>Self study</td>
<td>90,0</td>
<td>60.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h

CONTENTS

**Introduction to Machine learning**

**Description:**
Description and approach of the problems attacked by machine learning. Identification of modern areas and problems (data, scalability, heterogeneity, etc). Introduction to modern and advanced techniques covered in progress. Advanced application examples.
<table>
<thead>
<tr>
<th>Introduction to kernel based methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-standard kernel functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Non-standard kernel functions: refresh and expand the kernel function definition. Learning in RKHS. Kernel functions for text analysis, graphs, biomedical data (-OMIC).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Advanced kernel methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Advanced kernel methods: kPCA, KFDA, Relevance Vector Machines. Advanced applications.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neural layers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>- Perceptron and multilayer perceptron.</td>
</tr>
<tr>
<td>- Convolutional layers.</td>
</tr>
<tr>
<td>- Recurrent layers.</td>
</tr>
<tr>
<td>- Residual layers</td>
</tr>
<tr>
<td>- Mechanisms of attention</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training of deep neural networks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>- Backpropagation.</td>
</tr>
<tr>
<td>- Loss functions.</td>
</tr>
<tr>
<td>- Optimizers.</td>
</tr>
<tr>
<td>- Methodology.</td>
</tr>
<tr>
<td>- Data augmentation.</td>
</tr>
<tr>
<td>- Batch normalization</td>
</tr>
</tbody>
</table>
ACTIVITIES

Development of subject 1

Specific objectives:
1

Related competencies:
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 3h
Theory classes: 1h
Self study: 2h

Development of subject 2

Specific objectives:
3

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.

Full-or-part-time: 18h
Theory classes: 4h
Laboratory classes: 4h
Self study: 10h
Development of subject 3

Specific objectives:
5

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 18h
Theory classes: 4h
Laboratory classes: 4h
Self study: 10h
Development of subject 4

Specific objectives:
2, 5, 7, 8

Related competencies:
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG3. Work in multidisciplinary teams and projects related to the processing and exploitation of complex data, interacting fluently with engineers and professionals from other disciplines.
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 27h
Theory classes: 6h
Laboratory classes: 6h
Self study: 15h
Development of subject 5

Specific objectives:
2, 4, 7, 8

Related competencies:
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG3. Work in multidisciplinary teams and projects related to the processing and exploitation of complex data, interacting fluently with engineers and professionals from other disciplines.
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 27h
Theory classes: 6h
Laboratory classes: 6h
Self study: 15h
Development of subject 6

Specific objectives:
1, 2, 4, 7, 8

Related competencies:
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG3. Work in multidisciplinary teams and projects related to the processing and exploitation of complex data, interacting fluently with engineers and professionals from other disciplines.
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 27h
Theory classes: 6h
Laboratory classes: 6h
Self study: 15h
Development of subject 7

Specific objectives:
1, 2, 3, 6, 7, 8

Related competencies:
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG3. Work in multidisciplinary teams and projects related to the processing and exploitation of complex data, interacting fluently with engineers and professionals from other disciplines.
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 15h
Theory classes: 3h
Laboratory classes: 4h
Self study: 8h
## Practicum evaluation

**Specific objectives:**

1, 2

**Related competencies:**

- **CG1.** To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
- **CG2.** Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
- **CE4.** Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
- **CE1.** Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
- **CE6.** Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
- **CE2.** To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
- **CE9.** Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.
- **CE3.** Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
- **CE8.** Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
- **CE5.** Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
- **CT6.** Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

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

Guided activities: 2h
Practicum report

Specific objectives:
1, 2, 3, 4, 5, 6, 7, 8

Related competencies:
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG3. Work in multidisciplinary teams and projects related to the processing and exploitation of complex data, interacting fluently with engineers and professionals from other disciplines.
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE4. Use current computer systems, including high performance systems, for the process of large volumes of data from the knowledge of its structure, operation and particularities.
CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.
CE2. To be able to program solutions to engineering problems: Design efficient algorithmic solutions to a given computational problem, implement them in the form of a robust, structured and maintainable program, and check the validity of the solution.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 3h
Guided activities: 3h

GRADING SYSTEM

The subject is evaluated through a partial exam, a final exam and practical work in which real problems are addressed, writing the corresponding reports.

The partial exam will correspond to the kernel-based methods part and will discard this part of the syllabus.

The final grade is calculated as:

Grade = 0,4 * Jobs + 0,6 * (Final + Partial) / 2

For the students who attend the re-evaluation, the re-evaluation exam note will include the two parts of the syllabus and will substitute 0.6 * (Final + Partial) / 2.
BIBLIOGRAPHY

Basic: