230329 - MLR - Machine Learning Through Reinforcement

Coordinating unit: 230 - ETSETB - Barcelona School of Telecommunications Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications
Academic year: 2019
Degree: BACHELOR'S DEGREE IN TELECOMMUNICATIONS TECHNOLOGIES AND SERVICES ENGINEERING (Syllabus 2015). (Teaching unit Optional)
BACHELOR'S DEGREE IN NETWORK ENGINEERING (Syllabus 2010). (Teaching unit Optional)
BACHELOR'S DEGREE IN TELECOMMUNICATIONS SYSTEMS ENGINEERING (Syllabus 2010). (Teaching unit Optional)
BACHELOR'S DEGREE IN ELECTRONIC SYSTEMS ENGINEERING (Syllabus 2009). (Teaching unit Optional)
BACHELOR'S DEGREE IN AUDIOVISUAL SYSTEMS ENGINEERING (Syllabus 2009). (Teaching unit Optional)
BACHELOR'S DEGREE IN TELECOMMUNICATIONS SCIENCE AND TECHNOLOGY (Syllabus 2010). (Teaching unit Optional)

ECTS credits: 2
Teaching languages: Catalan, English

Teaching staff
Coordinator: Vidal Manzano, Jose
Others: Cabrera Bean, Margarita Asuncion
Giró Nieto, Xavier

Prior skills
Algebra, Probability and stochastic processes, Signals and systems

Requirements
Algebra, Probability and stochastic processes, Signals and systems

Degree competences to which the subject contributes

General:
08 CRPE. ABILITY TO IDENTIFY, FORMULATE AND SOLVE ENGINEERING PROBLEMS. To plan and solve engineering problems in the ICT with initiative, making decisions and with creativity. To develop a method of analysis and problem solving in a systematic and creative way.

Transversal:
04 COE N2. EFFICIENT ORAL AND WRITTEN COMMUNICATION - Level 2. Using strategies for preparing and giving oral presentations. Writing texts and documents whose content is coherent, well structured and free of spelling and grammatical errors.
06 URI N1. EFFECTIVE USE OF INFORMATION RESOURCES - Level 1. Identifying information needs. Using collections, premises and services that are available for designing and executing simple searches that are suited to the topic.
06 URI N2. EFFECTIVE USE OF INFORMATION RESOURCES - Level 2. Designing and executing a good strategy for advanced searches using specialized information resources, once the various parts of an academic document have been identified and bibliographical references provided. Choosing suitable information based on its relevance and quality.
06 URI N3. EFFECTIVE USE OF INFORMATION RESOURCES - Level 3. Planning and using the information necessary for an academic assignment (a final thesis, for example) based on a critical appraisal of the information resources used.
Teaching methodology

Classroom lectures and labs

Learning objectives of the subject

Master the principles of learning for reinforcement as an artificial intelligence tool based on the interaction of the machine with its environment, and which is at the base of systems such as autonomous vehicles, software that plays chess or go, or the organization of complex communication systems. We will work on its practical implementation and the evaluation in specific cases.

Study load

<table>
<thead>
<tr>
<th>Total learning time: 50h</th>
<th>Hours large group: 20h</th>
<th>40.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self study: 30h</td>
<td>60.00%</td>
</tr>
</tbody>
</table>
| **1. Introduction to reinforcement learning** | **Learning time:** 2h 30m  
Theory classes: 2h 30m |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Describe with examples the fundamental concepts and the problems that can be solved.</td>
</tr>
</tbody>
</table>
| **2. Markov decision processes** | **Learning time:** 2h 30m  
Theory classes: 2h 30m |
| **Description:** | - The agent-environment interface  
- Goals and rewards  
- Markov Decision Processes  
- Value functions and optimality: Bellman equation |
| **3. Dynamic programming** | **Learning time:** 2h 30m  
Theory classes: 2h 30m |
| **Description:** | - Policy evaluation, improvement and iteration  
- Dynamic programming based on MDP |
| **4. Monte-Carlo methods** | **Learning time:** 2h 30m  
Theory classes: 2h 30m |
| **Description:** | - First-visit Monte-Carlo methods  
- Every-visit Monte-Carlo methods  
- Exploration and exploitation  
- On-policy and off-policy methods |
### 5. Temporal-difference and Q-learning

**Learning time:** 2h 30m  
**Theory classes:** 2h 30m

**Description:**  
- Model-free learning using time differences  
- Q-learning and discrete actions  
- Games

---

### 6. Policy gradient methods

**Learning time:** 2h 30m  
**Theory classes:** 2h 30m

**Description:**  
- Policy gradient

---

### 7. Deep reinforcement learning

**Learning time:** 2h 30m  
**Theory classes:** 2h 30m

**Description:**  
Modeling of Q functions with deep neural networks

---

### 8. Laboratory and applications

**Learning time:** 2h 30m  
**Theory classes:** 2h 30m

**Description:**  
Labs in Matlab and/or Python distributed through the lectures  
- Dynamic channel allocation  
- Blackjack  
- Job-shop scheduling

---

### Qualification system

Exam and evaluation of labs
Bibliography

Basic: