ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

93000948 - Reinforcement Learning

DEGREE PROGRAMME

09AT - Master Universitario en Teoria de la Señal y Comunicaciones

ACADEMIC YEAR & SEMESTER

2020/21 - Semester 2

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	93000948 - Reinforcement Learning
No of credits	3 ECTS
Type	Optional
Academic year ot the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	09AT - Master Universitario en Teoria de la Señal y Comunicaciones
Centre	09 - Escuela Tecnica Superior de Ingenieros de Telecomunicacion
Academic year	2020-21

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Santiago Zazo Bello (Subject coordinator)	C-326	santiago.zazo@upm.es	Sin horario. Arrange the meeting by email
Julian Cabrera Quesada	C-320	julian.cabrera@upm.es	Sin horario. Arrange the meeting by email

Juan Parras Moral	C-303	j.parras@upm.es	Sin horario. Arrange the meeting by email

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

- Fundamentos De OptimizaciÓn

- TÉcnicas De OptimizaciÓn Para AnÁlisis De Datos Masivos

3.2. Other recommended learning outcomes

- Statistical Signal Processing

4. Skills and learning outcomes *

4.1. Skills to be learned

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB08 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB09 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un

modo que habrá de ser en gran medida autodirigido o autónomo

CE02 - Evaluar y sintetizar los resultados de un trabajo en equipo en proyectos relacionados con la teoría de la señal y las comunicaciones, en un entorno internacional.

CE03 - Valorar y contrastar la utilización de las diferentes técnicas disponibles para la resolución de problemas reales dentro del área de teoría de la señal y comunicaciones.

CT01 - Capacidad para comprender los contenidos de clases magistrales, conferencias y seminarios en lengua inglesa

CT03 - Capacidad para adoptar soluciones creativas que satisfagan adecuadamente las diferentes necesidades planteadas

CT04 - Capacidad para trabajar de forma efectiva como individuo, organizando y planificando su propio trabajo, de forma independiente o como miembro de un equipo

4.2. Learning outcomes

RA1 - Capacidad para desarrollar técnicas de tratamiento de señal específicas para datos masivos y diseñar aplicaciones sobre señales como: imágenes, señales de video, voz, audio y las procedentes de sensores de diversanaturaleza

RA12 - Capability to construct probabilistic models from experimental data using inference tools.

RA14 - Capability to model real phenomena using probability theory.

RA15 - Capability to relate the foundations of statistical inference with standard machine learning schemes.

RA4 - Formular problemas relacionados con la ingeniería como problemas de optimización en forma estándar

RA2 - Capacidad para planificar, diseñar y realizar aplicaciones que integren técnicas de tratamiento de señal, análisis estadístico y aprendizaje automático sobre datos masivos.

RA25 - Handle with ease the bases of linear algebra and calculus necessary to formulate problems optimization.

RA32 - Capability for planning, design and implement applications, incorporating signal processing, statistical analysis and machine learning

RA34 - Capability to develop and evaluate machine-learning techniques and to design big data learning systems

RA17 - Capacidad para aplicar conocimientos de modelado estadístico, técnicas de optimización y modelos de series temporales en el análisis de datos y como base para el desarrollo de algoritmos de aprendizaje automático

RA26 - Ability of oral and written communication

RA18 - Knowledge of tools for description, analysis and modeling of discrete-time random processes

RA7 - Capacidad para desarrollar y evaluar técnicas de aprendizaje automático y diseñar sistemas de aprendizaje para datos masivos

RA5 - Saber resolver problemas de optimización básicos como los de programación lineal o cuadrática

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

This subject describes the problem where an agent has to make optimum decisions just by interacting with the environment that provides a reward. Starting from MDP (Markov Decision Processes) as a proper model of the problem we will emphasize the case where this model is unknown and has to be inferred. Finally, different solutions suitable for large scale problems are developed where value and policy functions are approximated by linear and non linear architectures. We will distinguish three main blocks

1. Fundamentals of Reinforcement Learning
2. Planning and learning in small scale problems
3. Learning in large and continuous spaces

5.2. Syllabus

1. Introduction

2. Multi-Armed Bandits

3. Markov Decision Processes

4. Planning by Dynamic Programming

5. Model-Free methods.

6. Linear Approximation

7. Non-Linear Approximation. Deep RL

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1				
2				
3				
4				
5				
6				
7				
8	**1. Introduction 2. Multiarrmed bandits** Duration: 03:00 Lecture	**Personal work related to chapter 2** Duration: 01:00 Laboratory assignments		
9	**3. Markov decision processes** Duration: 03:00 Lecture	**Personal work related to chapter 3** Duration: 01:00 Laboratory assignments		
10	**4. Planning by dynamic programming** Duration: 03:00 Lecture	**Personal work related to chapter 4** Duration: 01:00 Laboratory assignments		
11	**5. Model free methods** Duration: 03:00 Lecture	**Personal work related to chapter 5** Duration: 01:00 Laboratory assignments		
12	**6. Linear approximation** Duration: 02:00 Lecture			**Mid term exam corresponding to chapters 1-4. The student will have to solve an optimization problem equivalent to the laboratory activities** Problem-solving test Continuous assessment Presential Duration: 02:00
13	**7. Non-linear approximation. Deep RL** Duration: 02:00 Lecture	**Personal work related to chapter 6** Duration: 02:00 Laboratory assignments		
14	**7. Non-linear approximation. Deep RL** Duration: 01:00 Lecture	**Personal work related to chapter 7** Duration: 01:00 Laboratory assignments		
15				
16				
				Final exam.Second option for the first part. The student will have to solve an optimization problem equivalent to the laboratory activities covering all the chapters. Problem-solving test Final examination Presential Duration: 02:00 **Second exam. The student will have to solve an optimization problem equivalent**

17				**to the laboratory activities** Problem-solving test Continuous assessment and final examination Presential Duration: 02:00 **Deliver the final report with solved exercises** Individual work Continuous assessment and final examination Not Presential Duration: 00:00

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
12	Mid term exam corresponding to chapters 1-4. The student will have to solve an optimization problem equivalent to the laboratory activities	Problem-solving test	Face-to-face	02:00	30%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03 CB10
17	Second exam. The student will have to solve an optimization problem equivalent to the laboratory activities	Problem-solving test	Face-to-face	02:00	30%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03 CB10
17	Deliver the final report with solved exercises	Individual work	No Presential	00:00	40%	3 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03 CB10

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills

17	Final exam.Second option for the first part. The student will have to solve an optimization problem equivalent to the laboratory activities covering all the chapters.	Problem-solving test	Face-to-face	02:00	30%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03 CB10
17	Second exam. The student will have to solve an optimization problem equivalent to the laboratory activities	Problem-solving test	Face-to-face	02:00	30%	3.5 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03 CB10
17	Deliver the final report with solved exercises	Individual work	No Presential	00:00	40%	3 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03 CB10

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Final exam. The student will have to solve an optimization problem equivalent to the laboratory activities It will cover all the chapters. The student has to provide all the reports corrsponding to the exercises of the whole course	Written test	Face-to-face	03:00	100%	0 / 10	CB08 CB09 CT01 CB07 CT03 CB06 CE02 CT04 CE03

						CB10

7.2. Assessment criteria

Students will be qualified through continuous evaluation by default. According to the Normativa de Evaluación del Aprendizaje de la Universidad Politécnica de Madrid, students willing to renounce to continuous evaluation must complete the Moodle task entitled "Renounce to continuous evaluation" before the first intermediate exam (deadline will be announced in Moodle).

Evaluation will assess if students have acquired all the competences of the subject. Thus, evaluation through final assessment will be carried out considering all the evaluation techniques used in continuous evaluation (EX, ET, TG, etc.), and will be celebrated in the exam period approved by Junta de Escuela for the current academic semester and year. Evaluation activities that assess learning outcomes that cannot be evaluated through a single exam can be carried out along the semester.

Extraordinary examination will be carried out exclusively by the final assessment method.

The evaluation procedure for the continuous assessment will be as follows:

One mid term exam including the first 4 chapters counting 30% of the final mark has to be completed. The student will have to solve a certain number of theoretical / practical issues similar as those contents of the practices and lectures. A minimum mark (3.5) is required.

One second term exam including the last chapters from 5 to 7 counting 30% of the final mark has to be completed. The student will have to solve a certain number of theoretical / practical issues similar as those contents of the practices and lectures. A minimum mark (3.5) is required.

The second exam will include solving in Matlab a short problem with the computer. The mark will be between 0 and 1 and will be considered as an indicator to check if the student is able to solve problems individually.

Report including all requested exercises. The mark between 0-4 will be weighted by the indicator described in the previous paragraph

The evaluation procedure for the final and re-sit examination will be as follows:

A final exam counting 60% of the final mark has to be completed. The student will have to solve a certain number of theoretical / practical issues similar as those contents of the practices and lectures.

The final exam will include solving in Matlab a short problem with the computer. The mark will be between 0 and 1 and will be considered as will be an indicator to check if the student is able to solve problems individually.

Report including all requested exercises. The mark between 0-4 will be weighted by the indicator described in the previous paragraph

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Reinforcement learning	Bibliography	Notes describing all the contents of the course
Matlab code of case studies	Others	We provide a Matlab code solving all the case studies proposed in the course
Course slides	Bibliography	Slides to be presented by the instructor to support the explanations
Reinforcement learning. An introduction. R. Sutton, A. Barto.The MIT Press. Draft second edition, 2015	Bibliography	Main reference of chapters 1 - 5
Dynamic Programming and Optimal Control. D. Bertsekas.Third edition. Vol. 2. Athena Scientific Pub.	Bibliography	Important reference to chapters 3 and 4
Reinforcement learning and dynamic programming using function approximators. L. Busoniu et al. CRC Press 2010	Bibliography	Important reference chapters 6 and 7

9. Other information

9.1. Other information about the subject

This subject shows the fundamental ideas of reinforcement learning that could be used to model different ODS objectives in dynamic forms ,as biological models (ODS 3), climate changing (ODS 13) or ecosystems (ODSs 14 y 15). It could also be applied to the efficient use of resources as water (ODS 6) or energy (ODS 7)).

In more general terms, we teach applied mathematics used exhaustively in engineering, in particular will affect telecommunications infraestructures (ODS 9).

This course also contributes to subobjetives 4.4: to increase the number of persons with professional competences and techniques to access to employment and entrepeunership and 4.7, to guarantee that all students acquire solid practical and theoretical knowledge required to promote sustainable development