

COURSE OUTLINE

(1) GENERAL

SCHOOL	SCHOOL OF SCIENCE						
ACADEMIC UNIT	INFORMATICS AND TELECOMMUNICATIONS						
LEVEL OF STUDIES	UNDERGRADUATE						
COURSE CODE	ΕΠ25	SEMESTER		7 th			
COURSE TITLE	Reinforcement Learning and Stochastic Games						
INDEPENDENT TEACHING ACTIVITIES <i>if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits</i>		WEEKLY TEACHING HOURS		CREDITS			
Lectures		4		6			
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (4).							
COURSE TYPE <i>general background, special background, specialised general, knowledge, skills development</i>	Elective (ΠΜ)						
	Track	S1	S2	S3	S4	S5	S6
PREREQUISITE COURSES:							
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	GREEK/ENGLISH						
IS THE COURSE OFFERED TO ERASMUS STUDENTS	YES						
COURSE WEBSITE (URL)	https://eclass.uoa.gr/courses/DI729/						

(2) LEARNING OUTCOMES

Learning outcomes <i>The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.</i> <i>Consult Appendix A</i> • Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area • Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B • Guidelines for writing Learning Outcomes

- Acquisition of knowledge and skills in an important area of machine learning with strong connections to optimal control theory, stochastic optimization, function approximation (neural networks), and game theory, which has achieved remarkable successes in recent years.
- Upon completion of the course, students will be able to:
 - correctly formulate reinforcement learning problems using the appropriate formalism for each case;
 - implement from scratch optimal control and stochastic optimization algorithms for multi-armed bandits and Markov decision processes based on the Bellman equations;
 - understand and distinguish between deep reinforcement learning algorithms, and apply them both theoretically and in programming environments;
 - design experimental methodologies for applying reinforcement learning algorithms through numerical/computational simulations and evaluate their results.

General Competences

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology

Project planning and management

Respect for difference and multiculturalism

Adapting to new situations

Respect for the natural environment

Decision-making

Showing social, professional and ethical responsibility and sensitivity to gender issues

Working independently

Criticism and self-criticism

Team work

Production of free, creative and inductive thinking

Working in an international environment

Others...

Working in an interdisciplinary environment

Production of new research ideas

Decision-making

Working independently

Adapting to new situations

Working in an international environment

Working in an interdisciplinary environment

Production of new research ideas

Criticism and self-criticism

Production of free, creative and inductive thinking

(3) SYLLABUS

This course addresses the following question: How can an agent learn to act towards achieving a certain goal in a dynamic, uncertain, and partially observable environment, possibly in the presence of other acting agents? This question and its specific variants (single agent, knowledge of the environment, stochasticity, learning from prior knowledge, common objectives, and coordination among agents) are of central importance in a broad range of modern applications, including computer vision, robotics, autonomous driving, natural language processing, games (board games, card games, and video games), and wireless communications.

The course focuses on the theory, methods, and algorithms of reinforcement learning dealing with the above question. Considering that the interaction between agents and their environment is continuous, the former observe the state of the environment through relevant measurements, and indirectly obtain information about

the actions of other agents in the overall system. Some information is shared among all agents, while other information remains private. Based on the available observations, each agent selects and executes an action. The state of the environment evolves over time, generally in an uncertain (probabilistic) manner, depending on its previous evolution and the actions of the agents. Each agent makes decisions sequentially based on learned rules; consequently, learning itself is sequential. This interaction cycle is repeated continuously, allowing agents to learn how to act so as to optimize not merely their immediate reward, but their overall long-term return.

The course 's content:

- Multi-armed bandit problems, with or without context, which simplify the decision-making process.
- Markov decision processes (MDPs) with known, possibly stochastic, system models; the perspective of dynamic programming and the Bellman equations will be studied, extending the formulation to partially observable MDPs (POMDPs).
- Algorithms for learning optimal policies (Q-learning, Monte Carlo/REINFORCE) when the environment model is unknown.
- Deep reinforcement learning algorithms that learn approximately optimal decisions using artificial neural networks.
- Extensions of deep reinforcement learning algorithms to world models and multi-agent environments involving cooperation or competition.
- Programming environments for deep reinforcement learning algorithms in Python/PyTorch using the Gymnasium and Stable-Baselines libraries.

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY <i>Face-to-face, Distance learning, etc.</i>	In Class (Face to Face)	
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY <i>Use of ICT in teaching, laboratory education, communication with students</i>	The learning process will be supported by the e-class platform. More specifically: Course Description, Material Supply, Announcements, Calendar, Assignment and Submission of Exercises, Discussion on Exercises, Questionnaires, External Links) E-mail communication	
TEACHING METHODS <i>The manner and methods of teaching are described in detail.</i> <i>Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.</i> <i>The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS</i>	Activity	Semester workload
	Lectures	39
	Seminars	13
	One (1) personalized programming assignment	78
	Independent study	20
	Course total	150
STUDENT PERFORMANCE EVALUATION <i>Description of the evaluation procedure</i> <i>Language of evaluation, methods of evaluation, summative or conclusive, multiple</i>	Students are assessed through one (1) compulsory individual assignment and one (1) final oral examination based on the assignment and the overall material of the course. The assignment is divided into sub-tasks covering the entire course material and includes programming implementations of algorithms and problems, as well as theoretical questions related to the implemented material. It is accompanied by an extensive written report	

choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other

Specifically-defined evaluation criteria are given, and if and where they are accessible to students.

presenting and discussing the results.

The oral examination is based on the assignment report and aims to assess the student's understanding of the implemented material through further theoretical questions.

Assessment methods	Number	Percentage
Personalized assignment	1	40%
Final oral examination	1	60%

(5) ATTACHED BIBLIOGRAPHY

1. "Reinforcement Learning: An Introduction," R. S. Sutton and A. G. Barto, MIT Press, Cambridge, MA, 2018.
2. "Dynamic Programming and Optimal Control," D. P. Bertsekas, Athena Scientific, 4th edition, 2007.
3. "Reinforcement Learning: State-of-the-Art," M. Wiering and M. Otterlo, Springer Berlin, Heidelberg, 2012.