EL2805 Reinforcement Learning 7.5 credits

Förstärkande inlärning

This is a translation of the Swedish, legally binding, course syllabus.

If the course is discontinued, students may request to be examined during the following two academic years

Establishment

The official course syllabus is valid from the fall semester 2023 in accordance with the decision by the Head of School: J-2023-0479. Date of decision: 14/04/2023

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Electrical Engineering

Specific prerequisites

For non-program students: 120 higher education credits and documented knowledge in English B or an equivalent discipline.

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

After passing the course, the student should be able to

- carefully formulate stochastic control problems as Markov decision-making process problems (MDP), classify equivalent problems and evaluate their traceability
- state the principle about optimality in finite time and infinite time horizon for MDP and solve MDP by means of dynamic programming
- derive solutions to MDP by using value- and policy iterations
- solve control problems for systems whose dynamics must be learnt with Q learning and SARSA algorithms
- explain the difference between on-policy and off-policy algorithms
- develop and implement RL algorithms with function approximation (for example deep RL algorithms where the Q function is approximated by the output of a neural network)
- solve bandit optimisation problems.

Course contents

The course gives an in-depth treatment of the modern theoretical tools that are used to design and analyse reinforcement learning algorithms (RL algorithms). It contains an introduction to RL and to its classical algorithms like Q-learning and SARSA, and present furthermore a justification behind the design of the latest algorithms, such as the striking optimal trade-off between exploration and exploitation. The course also covers algorithms that are used in the latest success histories for RL, e.g., deep RL algorithms.

Markov chains, Markov decision process (MDP), dynamic programming and value- and policy iterations, design of approximate controllers for MDP, stochastic linear quadratic control, the Multi-Armed Bandit problem, RL algorithms (Q-learning, Q-learning with function approximation).

Examination

- HEM1 - Homework 1, 1.0 credits, grading scale: P, F
- HEM2 - Homework 2, 1.0 credits, grading scale: P, F
- LAB1 - Lab 1, 1.0 credits, grading scale: P, F
- LAB2 - Lab 2, 1.0 credits, grading scale: P, F
- TENA - Written exam, 3.5 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.