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

Course syllabus for EL2805 valid from Autumn 2018

Grading scale

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

Education cycle

Second cycle

Main field of study

Electrical Engineering

Specific prerequisites

For single course students: 120 credits and documented proficiency in English B or equivalent.

Language of instruction

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

Intended learning outcomes

The course provides an in-depth treatment of the modern theoretical tools used to devise and analyse RL algorithms. It includes an introduction to RL and to its classical algorithms such as Q-learning, and SARSA, but further presents the rationale behind the design of more recent algorithms, such as those striking optimal trade-off between exploration and exploitation. The course also covers algorithms used in recent RL success stories, i.e., deep RL algorithms. After the course, you should be able to:

- Model the problem of controlling discrete-time stochastic systems with known dynamics, and formulate this problem as a Markov Decision Process (MDP)
- Solve MDPs using Bellman optimality principle and Dynamical Programming
- Derive solutions of MDPs using the value and policy iteration methods
- Control stochastic systems with unknown dynamics using Q-learning or SARSA algorithms
- Distinguish between on-policy and off-policy RL problems
- Develop and implement RL algorithms with function approximation (e.g. deep RL algorithms – in which the Q function is approximated by the output of a neural network)
- Understand solutions to solve bandit optimization problems
- Design RL algorithms striking a better exploration-exploitation trade-off than Q-learning based algorithms

Course contents

Markov chains, Markov Decision Process (MDP), Dynamic Programming and value / policy iteration methods, Multi-Armed Bandit problems, RL algorithms (Q-learning, Q-learning with function approximation, UCRL).

Disposition

Lectures, exercises, computer laboratory, homework

Course literature

Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley.

Bertsekas, Dynamic Programming and Optimal Control, vol. 1, Athena Scientific.

Bubeck and Cesa-Bianchi, Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems, Now publisher, Foundations and trends in machine learning, 2012

Sutton and Barto, Introduction to Reinforcement Learning, MIT Press, Cambridge, MA, USA, 1st edition, 1998

Szepesvari. Algorithms for Reinforcement Learning, Synthesis Lectures on Articial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2010

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
- TEN1 - Exam, 3.5 credits, grading scale: P, 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.

HEM1 - Homework 1, 1.0, grade scale: P, F

LAB1 - Lab 1, 1.5, grade scale: P, F

LAB2 - Lab 2, 1.5, grade scale: P, F

TEN1 - Exam, 3.5, grade scale: A, B, C, D, E, FX, F

Other requirements for final grade

H1: Homework, 1, grade scale: P/F

LAB1: Computer lab 1, 1, grade scale: P/F

LAB2: Computer lab 2, 1, grade scale: P/F

TEN1: Written exam, grade scale: A, B, C, D, E, FX, F

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.