EL2805 Förstärkande inlärning 7,5 hp

Reinforcement Learning

Fastställande

Kursplan för EL2805 gäller från och med VT19

Betygsskala

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

Utbildningsnivå

Avancerad nivå

Huvudområden

Elektroteknik

Särskild behörighet

För fristående kursstuderande: 120 hp samt dokumenterade kunskaper i engelska B eller motsvarande.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Kursen ger en djupgående behandling av de moderna teoretiska verktygen som används för att utforma och analysera RL-algoritmer. Den innehåller en introduktion till RL och dess klassiska algoritmer som Q-learning och SARSA, men presenterar vidare motiveringen bakom utformningen av de senaste algoritmerna, såsom de slående optimala avvägningarna mellan prospektering och exploatering. Kursen täcker även algoritmer som används i senaste RL-framgångshistorier, t.ex djupa RL-algoritmer.

Efter kursen ska du kunna:

- Noggrant formulera stokastiska reglerproblem som Markov Beslut Process (MDP) problem, klassificera motsvarande problem, och utvärdera deras spårbarhet
- Ange principen om optimalitet i ändlig tid och oändlig tidshorisont MDP, och lösa MDP mha. dynamisk programmering
- Härleda lösningar till MDP genom att använda värde och policy iterationer
- Lösa reglerproblem för system vars dynamik måste läras Control stochastic systems with unknown dynamics using Q-learning or SARSA algorithms
- Förstå skillnaden mellan on-policy och off-policy RL problem
- Utveckla och implementera RL-algoritmer med funktion approximation (t.ex djupa RL-algoritmer - där Q-funktionen approximeras av utgången från ett neuralt nätverk)
- Lösa banditoptimeringsproblem
- Föreslå RL-algoritmer som slår på en bättre utforskning av exploateringsutnyttjande än Q-inlärningsbaserade algoritmer

Kursinnehåll

Markov kedjor, Markov beslut Process (MDP), dynamisk programmering och värde / policy iteration metoder, utformning av approximativa regulatorer för MDP, stokastisk linjär kvadratisk reglering, Multi-Armed Bandit problem,RL algorithmer (Q-learning, Q-learning med funktion approximation, UCRL).

Kursupplägg

Föreläsningar, övningar, datorlaborationer, läxor.

Kurslitteratur

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 - Hemuppgift 1, 1,0 hp, betygsskala: P, F
- TEN1 - Tentamen, 3,5 hp, betygsskala: P, F
- HEM2 - Hemuppgift 2, 1,0 hp, betygsskala: P, F
- LAB2 - Lab 2, 1,0 hp, betygsskala: P, F
- LAB1 - Lab 1, 1,0 hp, betygsskala: P, F

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

HEM2 - Hemuppgift 2, 1,0, betygsskala: P, F

LAB1 - Lab 1, 1,5, betygsskala: P, F

LAB2 - Lab 2, 1,5, betygsskala: P, F

TEN1 - Tentamen, 3,5, betygsskala: A, B, C, D, E, FX, F

Övriga krav för slutbetyg

H1: Läxor (Homework), 1, grade scale: P/F

LAB1: Datorlaborationer 1, 1.5, grade scale: P/F

LAB2: Datorlaborationer 2, 1.5, grade scale: P/F

TEN1: Skriftlig tentamen, 3.5, grade scale: A, B, C, D, E, FX, F

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.