EL2805 Förstärkande inlärning 7,5 hp

Reinforcement Learning

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

Fastställande

Kursplanen gäller från och med HT 2023 enligt skolchefsbeslut: J-2023-0479. Beslutsdatum: 2023-04-14

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

Efter godkänd kurs ska studenten kunna

- noggrant formulera stokastiska reglerproblem som Markovbeslutsprocessproblem (MDP), klassificera motsvarande problem och utvärdera deras spårbarhet
- ange principen om optimalitet i ändlig tid och oändlig tidshorisont för MDP, och lösa MDP med hjälp av dynamisk programmering
- härleda lösningar till MDP genom att använda värde- och policyiterationer
- lösa reglerproblem för system vars dynamik måste läras med Q-learning och SARSA-algoritmer
- förklara skillnaden mellan on-policy- och off-policy-algoritmer
- utveckla och implementera RL-algoritmer med funktionsapproximation (till exempel djupa RL-algoritmer där Q-funktionen approximeras av utgången från ett neuralt nätverk)
- lösa banditoptimeringsproblem.

Kursinnehåll

Kursen ger en djupgående behandling av de moderna teoretiska verktygen som används för att utforma och analysera förstärkande inlärningsalgoritmer (RL-algoritmer). Den innehåller en introduktion till RL och dess klassiska algoritmer som Q-learning och SARSA, och 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 de senaste framgångshistorierna för RL, t.ex. djupa RL-algoritmer.

Markovkedjor, Markovbeslutsprocessproblem (MDP), dynamisk programmering, värde- och policyiterationer, utformning av approximativa regulatorer för MDP, stokastisk linjär kvadratisk reglering, Multi-Armed Bandit-problemet, RL-algoritmer (Q-learning, Q-learning med funktionsapproximation).

Examination

- HEM1 - Hemuppgift 1, 1,0 hp, betygsskala: P, F
- HEM2 - Hemuppgift 2, 1,0 hp, betygsskala: P, F
- LAB1 - Lab 1, 1,0 hp, betygsskala: P, F
- LAB2 - Lab 2, 1,0 hp, betygsskala: P, F
- TENA - Skriftlig tentamen, 3,5 hp, betygsskala: A, B, C, D, E, FX, 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.

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.