

Reinforcement learning

ECTS : 6

Volume horaire : 24

Description du contenu de l'enseignement :

Outline:

- Introduction to reinforcement learning (RL) and Markov decision processes (MDP)
- Algorithms for MDPs: dynamic programming, value iteration, policy iteration, linear programming
- Multi armed bandits: exploration vs. exploitation, stochastic bandits, adversarial bandits
- Value-based RL: general ideas and basic algorithms (TD-learning, SARSA, Q-learning)
- Function approximation: linear function approximation, deep neural networks and DQN
- Policy-based and actor-critic RL: policy gradient, natural policy gradient, actor-critic algorithms
- Bandits and RL: regret bounds and UCRL algorithm
- Decentralized MDPs and multi-agent RL: complexity results, simple algorithms (e.g. independent learners) and their applications (e.g. EV charging, wind farm control)

Organization of lectures in two parts (with 15min break):

- First part: lecture covering main notions
- Second part: more advanced material and/or exercises

Compétence à acquérir :

Reinforcement Learning (RL) refers to scenarios where the learning algorithm operates in closed-loop, simultaneously using past data to adjust its decisions and taking actions that will influence future observations. RL algorithms combine ideas from control, machine learning, statistics, and operations research. A common thread of all RL algorithms is the need to balance exploration (trying new things) and exploitation (choosing the most successful actions so far). This course will introduce the main models (multi-armed bandits and Markov decision processes) and key ideas for algorithm design (e.g. model-based vs. model-free RL, value based vs. policy based algorithms, on-policy vs. off-policy learning, function approximation).

Mode de contrôle des connaissances :

Homework assignments and project

Bibliographie, lectures recommandées :

Books MDPs:

- Martin L. Puterman. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons. 2014.
- Dimitri P. Bertsekas. Dynamic Programming and Optimal Control, Vol I, 4th Edition, Athena Scientific. 2017.
- Dimitri P. Bertsekas. Dynamic Programming and Optimal Control: Approximate Dynamic Programming , Vol II, 4th Edition, Athena Scientific. 2012.

Books RL:

- Richard S. Sutton and Andrew G. Barto. [Reinforcement Learning: An Introduction](#). Second Edition. MIT Press, Cambridge, MA, 2018.
- Csaba Szepesvari. [Algorithms for Reinforcement Learning](#). Morgan & Claypool Publishers. 2009.
- Sean Meyn. [Control Systems and Reinforcement Learning](#). Cambridge University Press. 2022.

Bandit algorithms:

- Tor Lattimore and Csaba Szepesvari. [Bandit algorithms](#). Cambridge University Press. 2020.

Implementation:

- [Gymnasium](#)
- [Stable Baselines](#) (reliable implementations of RL algorithms)
- [PettingZoo](#)
- [SustainGym](#) (A suite of environments designed to test the performance of RL algorithms on realistic sustainability tasks)

