Mathematics of Reinforcement Learning
Winter 2024

Reinforcement Learning (RL) is a rapidly growing field with significant importance in modern machine learning and artificial intelligence. Its impact spans diverse applications, from game-playing AI to robotics, autonomous systems, and healthcare. At its core, RL focuses on how agents can learn optimal actions through trial and error to maximize cumulative rewards within an environment. Unlike supervised learning, where labeled data guides decisions, RL involves interacting with environments to discover policies that balance immediate and long-term rewards.

In this seminar on the "Mathematics of Reinforcement Learning", we delve into the mathematical foundation that underpins RL algorithms. We begin with a primer on Markov chains, which model the probabilistic transitions between states in a system, forming the backbone of RL environments. From there, we transition into Markov decision processes (MDPs), which introduce decision-making into these chains by incorporating actions and rewards, allowing us to formalize the problem of finding optimal policies for agents. By understanding the mathematics behind these processes, including dynamic programming and the Bellman equations, participants gain a solid grounding in the theory driving state-of-the-art RL algorithms. We then dive into the actual implementation of Reinforcement Learning.

Instructor: Prof. D. Nils Detering

kick-off meeting: 23.09.2024, 16.30 Uhr im Seminarraum 2522.01.81

Prerequirements: A basic course in stochastics, some knowledge of Markov Chains or stochastic processes.

Grading: No grades. Only pass/no pass. For passing this course you need to give a 90 minute presentation and participate in at least 80% of the presentations throughout the semester.

Language: Depending on the number of non-german speakers, this course might require you to do the presentation in english but this will be discussed during the kick off meeting. You are certainly allowed to do the presentation in english.

References:

- Kemeny, J. G., & Snell, J. L. (1976). *Finite Markov Chains*. Springer.
- Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society.
- Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.
- Bertsekas, D. P. (2017). *Dynamic Programming and Optimal Control*. Athena Scientific.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press. **(Standard textbook on RI)**

- Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards* (PhD thesis). King's College, University of Cambridge.
- Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- Silver, D., et al. (2014). Deterministic Policy Gradient Algorithms. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*.
- Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.
- Dobrow R. P. (2016), Introduction to Stochastic Processes with R, *John Wiley & Sons Inc.*, **(accessible reference for Markov Chains)**
- Sheldon M. Ross, (1992) Applied Probability Models with Optimization Applications, *Dover Books on Mathematics* **(reference for Markov Decision Processes)**

Seminar Presentation Preliminary Breakdown

Part 1: Markov Chains (Talks 1-5)

1. **Introduction to Markov Chains**: Basic concepts, transition matrices, Chapman-Kolmogorov equations.
   References: Dobrow R. P. (2016), Norris (1997)
2. **Classification of States and Long-term Behavior**: Recurrence, transience, periodicity, stationary distributions.
   References: Dobrow R. P. (2016), Norris (1997)
3. **Ergodic Theorem and Mixing Times**: Convergence properties, ergodic Markov chains, mixing times.
   References: Dobrow R. P. (2016), Norris (1997)
4. **Absorbing Markov Chains and First Passage Times**: Absorbing states, computation of first passage probabilities and times.
   References: Dobrow R. P. (2016), Norris (1997), Levin et al. (2009)
5. **(Optional, depending on number of participants) Applications of Markov Chains**: Examples in queueing theory, random walks, and population models.
   References: Dobrow R. P. (2016), Kemeny & Snell (1976), Norris (1997)

Part 2: Markov Decision Processes (Talks 6-9)

1. **Introduction to Markov Decision Processes**: Defining MDPs, actions, rewards, and policies.
   References: Sheldon M. Ross (1992), Puterman (1994), Bertsekas (2017)
2. **Dynamic Programming and Bellman Equations**: Policy evaluation, Bellman expectation and optimality equations.
   References: Sheldon M. Ross (1992), Puterman (1994), Bertsekas (2017)
3. **Policy Iteration and Value Iteration**: Algorithms for solving MDPs, convergence properties.
   References: Sheldon M. Ross (1992), Puterman (1994), Sutton & Barto (2018)
4. **Applications of MDPs**: Real-world applications in operations research, robotics, and finance.
   References: Sheldon M. Ross (1992), Puterman (1994), Bertsekas (2017)

Part 3: Reinforcement Learning (Talks 10-14)

1. **Introduction to Reinforcement Learning**: Exploration vs. exploitation, reward signal, RL as a generalization of MDPs.
   References: Sutton & Barto (2018)

2. **Q-learning and SARSA**: Model-free RL methods, temporal difference learning, convergence guarantees.
   References: Sutton & Barto (2018), Watkins (1989)

3. **Policy Gradient Methods**: Introduction to policy-based methods, REINFORCE algorithm, actor-critic methods.
   References: Sutton & Barto (2018), Silver et al. (2014)

4. **Deep Reinforcement Learning**: Combining deep learning with RL, success stories (e.g., AlphaGo, DQN).
   References: Sutton & Barto (2018), Mnih et al. (2015), Silver et al. (2016)

5. **Recent Advances in RL**: Off-policy learning, distributed RL, and meta-learning.
   References: Sutton & Barto (2018), Silver et al. (2016)