

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. Dr. Nils Detering

Organizational meeting: We will have an organizational meeting on Monday, March 23rd at 2.p.m. via Webex. To sign up for the seminar, please send an email to Sarah Wolter sek-fvm@hhu.de with your name and registration number (Matrikelnummer) ahead of the meeting. We will then send you the link to the Webex meeting. Please also let us know in case you cannot attend the organizational meeting but wish to attend the seminar.

Time and room: TBD. The time will be discussed during the organizational meeting.

Meeting with Instructor: Please fix a meeting with me at least one week before your talk. For this meeting you must have an outline of your talk ready and you should be familiar with the material that you need to cover during your talk. I will give you feedback and based on this feedback you will revise your presentation. You are welcome to contact me earlier in case of questions.

Prerequisites: A basic course in stochastics, some knowledge of Markov Chains or stochastic processes is an advantage but not required.

Grading: Graded if grades are needed for your study program. By default only pass/no pass. For passing this course you need to give a 90 minute presentation, participate in at least 80% of the presentations throughout the semester and hand in notes of your presentation.

Language: The language of instruction is english. The seminar is a great opportunity to practice your english presentation skills in a low stake environment.

References:

- L. Döring, (2025) Lecture Notes: The Mathematics of Reinforcement Learning https://www.wim.uni-mannheim.de/media/Lehrstuehle/wim/doering/RL/RL_VORLESUNG.pdf
- Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.

- Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press. **(Standard textbook on RL)**
- 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* **(book reference for Markov Decision Processes)**

Seminar Presentation Preliminary Breakdown. The precise breakdown will depend on the number of students attending.

Markov Decision Processes (Talks 1-8) cover roughly the following topics

- **Primer on Markov Chains, 1 lecture:** Basic concepts, transition matrices, Chapman-Kolmogorov equations.
References: Dobrow R. P. (2016), Norris (1997)
- **Introduction to Markov Decision Processes, 1-2 lectures:** Defining MDPs, actions, rewards, and policies.
References: L. Döring, (2025), Sheldon M. Ross (1992), Puterman (1994)
- **Dynamic Programming and Bellman Equations, 2 lectures:** Policy evaluation, Bellman expectation and optimality equations.
References: L. Döring, (2025), Sheldon M. Ross (1992), Puterman (1994)
- **Policy Iteration and Value Iteration, 1-2 lectures:** Algorithms for solving MDPs, convergence properties.
References: L. Döring, (2025), Sheldon M. Ross (1992), Puterman (1994)

Reinforcement Learning (Talks 9-15) cover roughly the following topics

- **Introduction to Reinforcement Learning, 1 lecture:** Exploration vs. exploitation.
References: L. Döring, (2025), Sutton & Barto (2018)
- **Introduction to Reinforcement Learning, 1-2 lectures:** Stochastic Approximation.
References: L. Döring, (2025), Sutton & Barto (2018)
- **Introduction to Reinforcement Learning, 1-2 lectures:** Q-learning and SARSA.
References: L. Döring, (2025), Sutton & Barto (2018)
- **Q-learning and SARSA, 1-2 lectures:** Model-free RL methods, temporal difference learning, convergence guarantees.
References: L. Döring, (2025), Sutton & Barto (2018)

- **Policy Gradient Methods, 1-2 lectures:** Introduction to policy-based methods, REINFORCE algorithm, actor-critic methods.
References: Sutton & Barto (2018)