Deep Reinforcement Learning – Winter 2018/19

Home | Lectures | Assignments | Exam Questions | Related Courses

The lecture content, including references to study materials.

The main study material is the Reinforcement Learning: An Introduction; second edition by Richard S. Sutton and Andrew G. Barto (http://incompleteideas.net/book/the-book-2nd.html) (reffered to as RLB). It is available online (https://drive.google.com/open?id=1opPSz5AZ_kVa1uWOdOiveNiBFiEOHjkG) and also as a hardcopy since October 15, 2018.

References to study materials cover **all theory required** at the exam, and sometimes even more – the references in *italics* cover topics **not required** for the exam.

1. Introduction to Reinforcement Learning

Oct 08 | Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?01)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-01.pdf)

multiarmed_bandits

- *History of RL [Chapter 1 of RLB]*
- Multi-armed bandits [Chapter 2 of RLB]

2. Markov Decision Process, Optimal Solutions, Monte Carlo Methods

Oct 15 | Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?02)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-02.pdf)

policy_iteration | monte_carlo

- Markov Decision Process [Sections 3-3.3 of RLB]
- Policies and Value Functions [Sections 3.5-3.6 of RLB]
- Value Iteration [Sections 4 and 4.4 of RLB]

  - Proof of convergence only in slides
- Policy Iteration [Sections 4.1-4.3 of RLB]
- Generalized Policy Iteration [Section 4.6 or RLB]
- Monte Carlo Methods [Sections 5-5.4 of RLB]

3. Temporal Difference Methods, Off-Policy Methods

Oct 22 Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?03)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-03.pdf)

q_learning importance_sampling lunar_lander

- Model-free and model-based methods, using state-value or action-value functions [Chapter 8 before Section 8.1, and Section 6.8 of RLB]
- Temporal-difference methods [Sections 6-6.3 of RLB]
- Sarsa [Section 6.4 of RLB]
- Q-learning [Section 6.5 of RLB]
- Off-policy Monte Carlo Methods [Sections 5.5-5.7 of RLB]
- Expected Sarsa [Section 6.6 of RLB]

4. N-step Methods, Function Approximation

Nov 05 Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?04)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-04.pdf)

q_learning_tiles

- Double Q-learning [Section 6.7 of RLB]
- N-step TD policy evaluation [Section 7.1 of RLB]
- Off-policy n-step Sarsa [Section 7.3 of RLB]
- Tree backup algorithm [Section 7.5 of RLB]
- Function approximation [Sections 9-9.3 of RLB]
- Tile coding [Section 9.5.4 of RLB]

5. Function Approximation, Deep Q Network

Nov 12 Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?05)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-05.pdf)

q_network

- Linear function approximation [Section 9.4 of RLB, without the Proof of Convergence if Linear TD(0)]
- Semi-Gradient TD methods [Sections 9.3, 10-10.2 of RLB]

- Off-policy function approximation TD divergence [Sections 11.2-11.3 of RLB]
- Deep Q Network [Volodymyr Mnih et al.: Human-level control through deep reinforcement learning (https://storage.googleapis.com/deepmind-media/dqn/DQNNaturePaper.pdf)]

6. Rainbow

Nov 19 | Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?06)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-06.pdf)

car_racing | reinforce

- Double Deep Q Network (DDQN) [Hado van Hasselt et al.: Deep Reinforcement Learning with Double Q-learning (https://arxiv.org/abs/1509.06461)]
- Prioritized Experience Replay [Tom Schaul et al.: Prioritized Experience Replay (https://arxiv.org/abs/1511.05952)]
- Dueling Deep Q Network [Ziyu Wang et al.: Dueling Network Architectures for Deep Reinforcement Learning (https://arxiv.org/abs/1511.06581)]
- Noisy Nets [Meire Fortunato et al.: Noisy Networks for Exploration (https://arxiv.org/abs/1706.10295)]
- Distributional Reinforcement Learning [Marc G. Bellemare et al.: A Distributional Perspective on Reinforcement Learning (https://arxiv.org/abs/1707.06887)]
- Rainbow [Matteo Hessel et al.: Rainbow: Combining Improvements in Deep Reinforcement Learning (https://arxiv.org/abs/1710.02298)]

7. Policy Gradient Methods

Nov 26 | Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?07)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-07.pdf)

reinforce_with_baseline | cart_pole_pixels

- Policy Gradient Methods [Sections 13-13.1 of RLB]
- Policy Gradient Theorem [Section 13.2 of RLB]
- REINFORCE algorithm [Section 13.3 of RLB]
- REINFORCE with baseline algorithm [Section 13.4 of RLB]
- Actor-Critic methods [Section 13.5 of RLB, without the eligibility traces variant]

8. Advantage Actor-Critic, Continuous Action Space

Dec 03 Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?08)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-08.pdf)

paac paac_continuous

- A3C and asynchronous RL [Volodymyr Mnih et al.: Asynchronous Methods for Deep Reinforcement Learning (https://arxiv.org/abs/1602.01783)]
- PAAC [Alfredo V. Clemente et al.: Efficient Parallel Methods for Deep Reinforcement Learning (https://arxiv.org/abs/1705.04862)]
- Gradient methods with continuous actions [Section 13.7 of RLB]

9. Deterministic Policy Gradient, Advanced RL Algorithms

Dec 10 Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?09)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-09.pdf)

ddpg walker

- Deterministic policy gradient theorem (DPG) [David Silver et al.: Deterministic Policy Gradient Algorithms (http://proceedings.mlr.press/v32/silver14.pdf)]
- Deep deterministic policy gradient (DDPG) [Timothy P. Lillicrap et al.: Continuous Control with Deep Reinforcement Learning (https://arxiv.org/abs/1509.02971)]
- *Natural policy gradient (NPG) [Sham Kakade: A Natural Policy Gradient (https://papers.nips.cc/paper/2073-a-natural-policy-gradient.pdf)]*
- *Truncated natural policy gradient (TNPG), Trust Region Policy Optimalization (TRPO) [John Schulman et al.: Trust Region Policy Optimization (https://arxiv.org/abs/1502.05477)]*
- *Proximal policy optimization (PPO) [John Schulman et al.: Proximal Policy Optimization Algorithms (https://arxiv.org/abs/1707.06347)]*
- *Soft actor-critic (SAC) [Tuomas Haarnoja et al.: Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor (https://arxiv.org/abs/1801.01290)]*

10. TD3, Monte Carlo Tree Search

Dec 17 Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?10)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-10.pdf)

walker_hardcore az_quiz az_quiz_randomized

- Twin delayed deep deterministic policy gradient (TD3) [Scott Fujimoto et al.: Addressing Function Approximation Error in Actor-Critic Methods

(https://arxiv.org/abs/1802.09477)]
- AlphaZero [David Silver et al.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play (https://deepmind.com/documents/260/alphazero_preprint.pdf)]

11. V-trace, PopArt Normalization, Partially Observable MDPs

Jan 07 | Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides/?11)

PDF Slides (https://ufal.mff.cuni.cz/~straka/courses/npfl122/1819/slides.pdf/npfl122-11.pdf)

vtrace | memory_game

- The V-trace algorithm of IMPALA [Lasse Espeholt et al.: IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures (https://arxiv.org/abs/1802.01561)]
- PopArt reward normalization [Matteo Hessel et al.: Multi-task Deep Reinforcement Learning with PopArt (https://arxiv.org/abs/1809.04474)]
- *MERLIN model [Greg Wayne et al.:Unsupervised Predictive Memory in a Goal-Directed Agent (https://arxiv.org/abs/1803.10760)]*
- *FTW agent for multiplayer CTF [Max Jaderberg et al.: Human-level performance in first-person multiplayer games with population-based deep reinforcement learning (https://arxiv.org/abs/1807.01281)]*

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Malostranské náměstí 25
118 00 Praha
Czech Republic

+420 951 554 278 (phone)
ufal@ufal.mff.cuni.cz (mailto:ufal@ufal.mff.cuni.cz)