Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Matthias Schubert
Sebastian Schmoll, Sabrina Friedl

Munich, 14.01.2019

Deep Learning and Artificial Intelligence

WS 2018/19

Exercise 11: Modelfree Reinforcement Learning

Exercise 11-1 RL Theory Questions

(a) In Reinforcement Learning, what do the terms 'prediction' and 'control' refer to?

(b) What are the differences between model-free and model-based methods? What do they have in common? Which category do Dynamic Programming (DP), Temporal-Difference-Learning (TD) and Monte Carlo (MC) -methods belong to?

(c) Write down the Bellman equations for optimal state values (utilities), denoted $u_*(s)$, and optimal state-action values, denoted $q_*(s, a)$. How are $u_*(s)$ and $q_*(s, a)$ related to each other?

(d) What does on- and off-policy learning mean? Name an example of each.

(e) Why is Q-learning an off-policy method? What is the difference to Sarsa?

(f) When do we need exploration and why? Why is there a trade-off between exploitation and exploration?

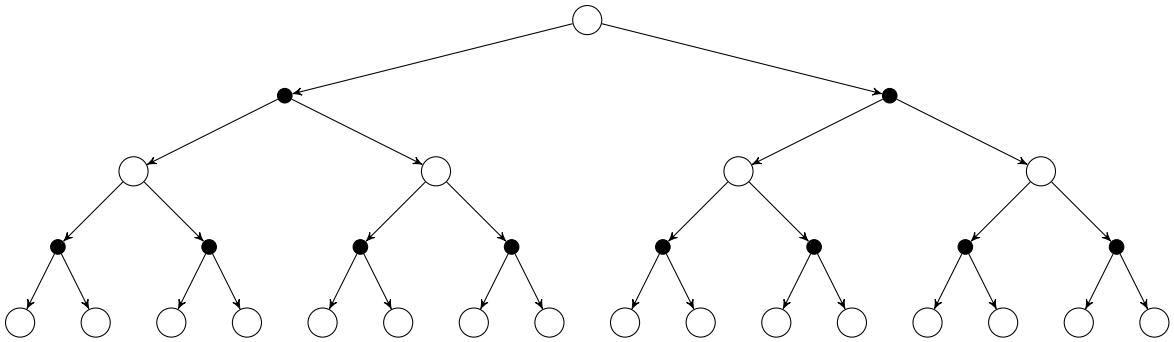
(g) How is TD(λ) related to Monte Carlo control and one-step TD? Remember from the lecture that the λ-Return G_t^λ combines all n-step returns $G_t^{(n)}$ as follows:

$$G_t^\lambda = (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}G_t^{(n)}$$

Exercise 11-2 Backup Strategies

In the lecture you learned three different backup strategies for the Bellman equation (Dynamic Programming (DP), Monte Carlo (MC) and Temporal Difference Learning (TD)).

(a) Assume we have an MDP with exactly two steps (two times an action is performed). The figure below visualizes this MDP. An unfilled circle represents a state and a filled circle represents an action respectively.
For each of the different backup strategies, mark the pathes that are used. For DP, assume a fixed policy $\pi(S_t)$ in the first step (instead of max). For MC and TD, just choose arbitrary actions. Also write down the formulas for these updates.



(b) How do the three backup strategies compare regarding variance, efficiency, necessity of a model and bias?

Exercise 11-3 Model-free RL in Python

On the lecture web-page you can find a Jupyter notebook file with a programming exercise for model-free reinforcement learning. Please follow the instructions in the notebook.