

Spring 2020

Exercises

An Introduction to Reinforcement Learning

Author
Kurath Samuel

May 4, 2020

Contents

1	Reinforcement Learning	2
1.1	Übung 1	2
1.1.1	Episode	2
1.1.2	Reward	3
1.1.3	Discounted Reward	3
1.1.4	State-Value-Function	3

1 Reinforcement Learning

Ausgangslage Die Übungen zu Reinforcement Learning basieren auf dem Markov Decision Process zu **Kurs absolvieren**, welcher in Abbildung 1.1) illustriert ist. Gestartet wird im Zustand *Attend Cours*. *Fail Cours* und *Success Cours* sind beides Terminalzustände.

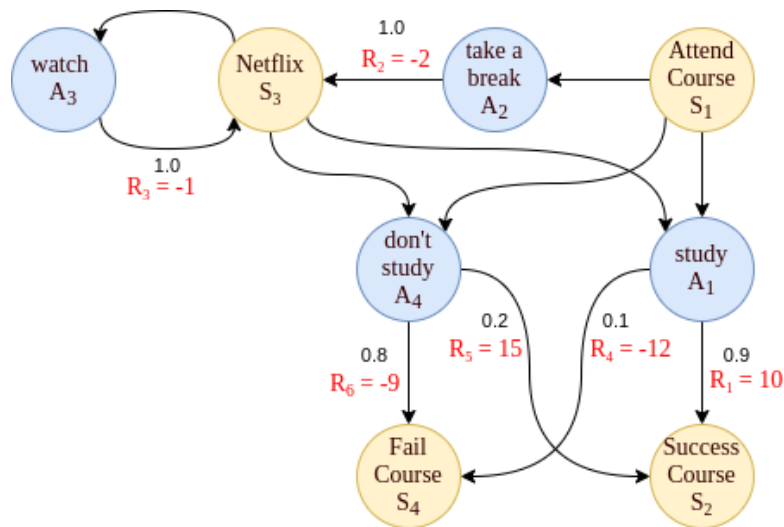


Figure 1.1: Markov Decision Process - Kurs absolvieren

1.1 Übung 1

1.1.1 Episode

Welcher der folgenden **Episoden** sind valide in Bezug auf **Kurs absolvieren**?

- a) S_1, A_1, R_1, S_2
- b) $S_1, A_2, R_2, S_3, A_3, R_3, S_3, A_4, R_6, S_4$
- c) $S_1, A_2, R_2, S_3, A_3, R_3, S_3, R_{15}, S_2$
- d) S_1, A_4, R_4, S_4
- e) $S_1, A_2, R_2, S_3, A_3, R_3, S_3, A_3, R_3, S_3$

1.1.2 Reward

Welchen **Reward** erhalten wir in der Episode (1.1)?

$$S_1, A_2, R_2, S_3, A_3, R_3, S_3, A_3, R_3, S_3, A_4, R_5, S_2 \quad (1.1)$$

1.1.3 Discounted Reward

Angenommen wir gehen von einem Discount Factor von $\gamma = 0.8$ aus und befinden uns im Startzustand S_1 ($t = 0$). Was für ein **Discounted Reward** resultiert in Bezug der Episode (1.1)?

1.1.4 State-Value-Function

a) Bestimme den Erwartungswert im Zustand S_1 unter Bezug der Policy (1.2) aka. "Musterstudent" und $\gamma = 0.8$.

$$\pi : \left\{ S_1 \rightarrow A_1 \right. \quad (1.2)$$

b) Nun ändern wir unser Verhalten vom "Musterstudent" zum "Lazy guy". Dies bedeutet, dass wir die Policy (1.3) wählen, die restlichen Parameter bleiben gleich. (Zustand S_1 , $\gamma = 0.8$) Wie verändert sich der Erwartungswert?

$$\pi : \left\{ \begin{array}{l} S_1 \rightarrow A_2 \\ S_3 \rightarrow A_4 \end{array} \right. \quad (1.3)$$