Grundlagen adaptiver Wissenssysteme

Übungen zur Vorlesung

Prof. Dr. Thomas Gabel
Frankfurt University of Applied Sciences
Faculty of Computer Science and Engineering
tgabel@fb2.fra-uas.de

Aufgabenblatt 3

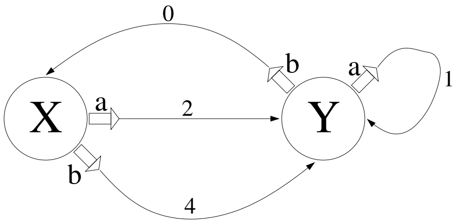
1. **Aufgabenblatt 3 – Übung 7**
2. Aufgabenblatt 3 – Übung 8
3. Aufgabenblatt 3 – Übung 9
4. Aufgabenblatt 3 – Übung 10

Aufgabe 7: SKP-Probleme vs. Diskontierung

Betrachten Sie den in der Abbildung dargestellten MDP. Alle Transitionen sind deterministisch; ein Diskontierungsfaktor von $\gamma < 1$ wird verwendet.

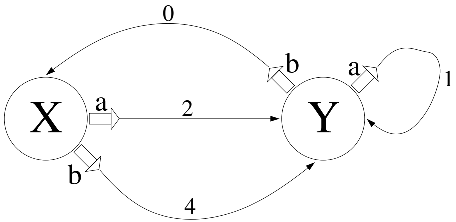
Aufgabe 7: SKP-Probleme vs. Diskontierung

Betrachten Sie den in der Abbildung dargestellten MDP. Alle Transitionen sind deterministisch; ein Diskontierungsfaktor von $\gamma < 1$ wird verwendet.



Aufgabe 7: SKP-Probleme vs. Diskontierung

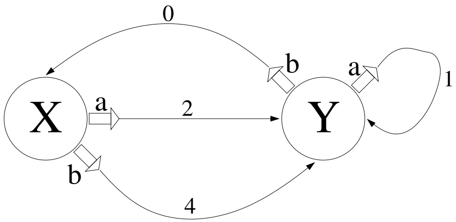
Betrachten Sie den in der Abbildung dargestellten MDP. Alle Transitionen sind deterministisch; ein Diskontierungsfaktor von $\gamma < 1$ wird verwendet.



(a) Begründen Sie, warum es sinnvoll ist, für den dargestellten MDP Diskontierung zu verwenden.

Aufgabe 7: SKP-Probleme vs. Diskontierung

Betrachten Sie den in der Abbildung dargestellten MDP. Alle Transitionen sind deterministisch; ein Diskontierungsfaktor von $\gamma < 1$ wird verwendet.



(a) Begründen Sie, warum es sinnvoll ist, für den dargestellten MDP Diskontierung zu verwenden.

- Im betrachteten MDP existiert kein absorbierender Terminalzustand.
- Verwendet man keine Diskontierung, so sind die Pfadkosten in beiden Zuständen unbeschränkt.

Aufgabe 7: SKP-Probleme vs. Diskontierung

(b) Gibt es eine Klasse von Problemstellungen, für die es problematisch wäre, diese als SKP-Probleme zu modellieren?

Aufgabe 7: SKP-Probleme vs. Diskontierung

(b) Gibt es eine Klasse von Problemstellungen, für die es problematisch wäre, diese als SKP-Probleme zu modellieren?

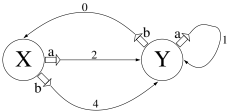
- Nicht-episodische Aufgaben sind schwierig als SKP zu formulieren, da sie nicht terminieren.
- Also Probleme, bei denen kein absorbierender Terminalzustand gegeben ist, in dem keine weiteren Kosten mehr anfallen.
- In diesem Fall ist die Verwendung von Diskontierungsraten sinnvoll.
- Beispiele:

Aufgabe 7: SKP-Probleme vs. Diskontierung

(b) Gibt es eine Klasse von Problemstellungen, für die es problematisch wäre, diese als SKP-Probleme zu modellieren?

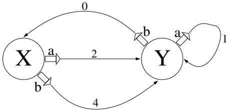
- Nicht-episodische Aufgaben sind schwierig als SKP zu formulieren, da sie nicht terminieren.
- Also Probleme, bei denen kein absorbierender Terminalzustand gegeben ist, in dem keine weiteren Kosten mehr anfallen.
- In diesem Fall ist die Verwendung von Diskontierungsraten sinnvoll.
- Beispiele:
  - Regelungsaufgaben (z.B. Heizungsventil)
  - Das (unendlich lange) Balancieren eines inversen Pendels (vgl. Demo-Video aus der Vorlesung).

Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^{\pi}(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

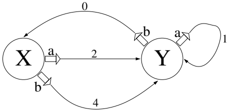
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^\pi(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

- Da es zwei Zustände gibt, in denen jeweils zwei Aktionen ausgeführt werden können, ergeben sich $2^2 = 4$ mögliche Strategien:

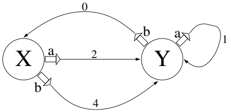
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^{\pi}(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

- Da es zwei Zustände gibt, in denen jeweils zwei Aktionen ausgeführt werden können, ergeben sich $2^2 = 4$ mögliche Strategien:
  - $\pi_1(X) = a$, $\pi_1(Y) = a$
  - $\pi_2(X) = a$, $\pi_2(Y) = b$
  - $\pi_3(X) = b$, $\pi_3(Y) = a$
  - $\pi_4(X) = b$, $\pi_4(Y) = b$

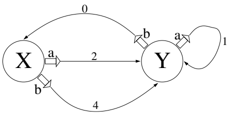
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^\pi(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

- Ermittlung der Pfadkosten für alle Zustände und alle Strategien

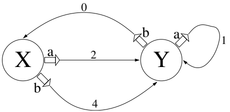
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^\pi(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

- Ermittlung der Pfadkosten für alle Zustände und alle Strategien $\rightarrow$ Tafel

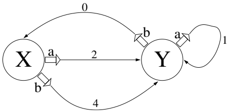
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^{\pi}(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

- Die optimale Strategie ist π_2 mit $\pi_2(X) = a$ und $\pi_2(Y) = b$.

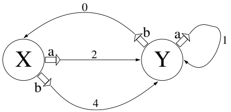
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^\pi(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

- Die optimale Strategie ist π_2 mit $\pi_2(X) = a$ und $\pi_2(Y) = b$.
- Erläuterung:

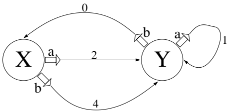
Aufgabe 7: SKP-Probleme vs. Diskontierung



(c) Wie viele Strategien gibt es in diesem MDP? Geben Sie für alle möglichen Strategien π und für alle Zustände i die erwarteten Pfadkosten $V^{\pi}(i)$ als Funktion von γ an und ermitteln Sie die optimale Strategie.

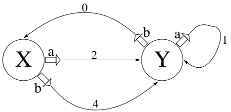
- Die optimale Strategie ist π_2 mit $\pi_2(X) = a$ und $\pi_2(Y) = b$.
- Erläuterung: $\rightarrow$ Tafel

Aufgabe 7: SKP-Probleme vs. Diskontierung



(d) Diskontierte Probleme können als Spezialfall von SKP-Problemen angesehen werden. Konstruieren Sie einen MDP mit Terminalzustand, der zu dem in der Abbildung dargestellten MDP äquivalent ist.

Aufgabe 7: SKP-Probleme vs. Diskontierung



(d) Diskontierte Probleme können als Spezialfall von SKP-Problemen angesehen werden. Konstruieren Sie einen MDP mit Terminalzustand, der zu dem in der Abbildung dargestellten MDP äquivalent ist.

- → Tafel

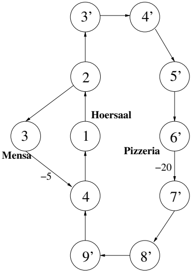
Aufgabenblatt 3

1. Aufgabenblatt 3 – Übung 7
2. **Aufgabenblatt 3 – Übung 8**
3. Aufgabenblatt 3 – Übung 9
4. Aufgabenblatt 3 – Übung 10

Aufgabe 8: Mensa oder Pizzeria

Betrachten Sie den 11-Zustands-MDP, dessen Übergangsgraph in der Abbildung gegeben ist. Alle Transitionen sind deterministisch.

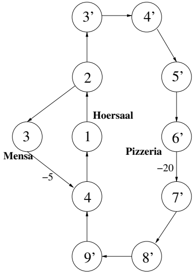
Der Agent erfährt Kosten in Höhe von -5, wenn er die Mensa verlässt und zum Hörsaal zurückkehrt und Kosten von -20, wenn er von der Pizzeria zurückkehrt. Alle anderen Übergange sind kostenfrei.



Aufgabe 8: Mensa oder Pizzeria

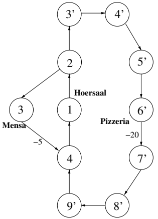
Betrachten Sie den 11-Zustands-MDP, dessen Übergangsgraph in der Abbildung gegeben ist. Alle Transitionen sind deterministisch.

Der Agent erfährt Kosten in Höhe von -5, wenn er die Mensa verlässt und zum Hörsaal zurückkehrt und Kosten von -20, wenn er von der Pizzeria zurückkehrt. Alle anderen Übergange sind kostenfrei.



(a) Wie viele verschiedene deterministische Strategien gibt es in diesem MDP?

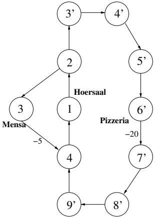
Aufgabe 8: Mensa oder Pizzeria



(a) Wie viele verschiedene deterministische Strategien gibt es in diesem MDP?

- Es gibt 2 deterministische Strategien.

Aufgabe 8: Mensa oder Pizzeria



(a) Wie viele verschiedene deterministische Strategien gibt es in diesem MDP?

- Es gibt 2 deterministische Strategien.
- Einziger Zustand mit "Wahlfreiheit" des Agenten ist Zustand 2. Hier erfolgt die Entscheidung für den linken oder rechten Pfad durch den MDP.
- Die beiden korrespondierenden Strategien werden im Folgenden mit π_M und π_P bezeichnet.

Aufgabe 8: Mensa oder Pizzeria

(b) Geben Sie für jede Strategie einen Ausdruck für die im Zustand 1 zu erwartenden Kosten an (Diskontierung mit $\gamma < 1$ wird vorausgesetzt).

Aufgabe 8: Mensa oder Pizzeria

(b) Geben Sie für jede Strategie einen Ausdruck für die im Zustand 1 zu erwartenden Kosten an (Diskontierung mit $\gamma < 1$ wird vorausgesetzt).

- Zur Beantwortung der Frage ist die Strategie π für $\pi = \pi_M$ und π_P zu bewerten. **Wir beginnen mit π_M.**

Aufgabe 8: Mensa oder Pizzeria

(b) Geben Sie für jede Strategie einen Ausdruck für die im Zustand 1 zu erwartenden Kosten an (Diskontierung mit $\gamma < 1$ wird vorausgesetzt).

- Zur Beantwortung der Frage ist die Strategie π für $\pi = \pi_M$ und π_P zu bewerten. **Wir beginnen mit π_M.**

$$V_{k+1}^{\pi}(i) = \sum_{j=0}^{n} p_{ij}(\pi(i)) \left(c(i, \pi(i)) + \gamma V_k^{\pi}(j)\right)$$

- Wir starten mit $V_0^{\pi} \equiv 0$ und weil $c(i, a) = 0$ für alle $i \neq 3$ erhalten wir nach einfacher Anwendung des Updates auf V_0^{π}:

Aufgabe 8: Mensa oder Pizzeria

(b) Geben Sie für jede Strategie einen Ausdruck für die im Zustand 1 zu erwartenden Kosten an (Diskontierung mit $\gamma < 1$ wird vorausgesetzt).

- Zur Beantwortung der Frage ist die Strategie π für $\pi = \pi_M$ und π_P zu bewerten. **Wir beginnen mit π_M.**

$$V_{k+1}^{\pi}(i) = \sum_{j=0}^{n} p_{ij}(\pi(i))\,(c(i,\pi(i)) + \gamma V_k^{\pi}(j))$$

- Wir starten mit $V_0^{\pi} \equiv 0$ und weil $c(i,a) = 0$ für alle $i \neq 3$ erhalten wir nach einfacher Anwendung des Updates auf V_0^{π}:

$$V_1^{\pi}(i) = \begin{cases} -5 & \text{für } i = 3 \\ 0 & \text{sonst} \end{cases}$$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi_M}(i)$	0	0	0	0
$V_1^{\pi_M}(i)$	0	0	-5	0

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi_M}(i)$	0	0	0	0
$V_1^{\pi_M}(i)$	0	0	-5	0
$V_2^{\pi_M}(i)$	0	-5γ	-5	0

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi_M}(i)$	0	0	0	0
$V_1^{\pi_M}(i)$	0	0	-5	0
$V_2^{\pi_M}(i)$	0	-5γ	-5	0
$V_3^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	0

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi_M}(i)$	0	0	0	0
$V_1^{\pi_M}(i)$	0	0	-5	0
$V_2^{\pi_M}(i)$	0	-5γ	-5	0
$V_3^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi M}(i)$	0	0	0	0
$V_1^{\pi M}(i)$	0	0	-5	0
$V_2^{\pi M}(i)$	0	-5γ	-5	0
$V_3^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi_M}(i)$	0	0	0	0
$V_1^{\pi_M}(i)$	0	0	-5	0
$V_2^{\pi_M}(i)$	0	-5γ	-5	0
$V_3^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi_M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_6^{\pi_M}(i)$	$-5\gamma^2$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi_M}(i)$	0	0	0	0
$V_1^{\pi_M}(i)$	0	0	-5	0
$V_2^{\pi_M}(i)$	0	-5γ	-5	0
$V_3^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi_M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi_M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_6^{\pi_M}(i)$	$-5\gamma^2$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_7^{\pi_M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi M}(i)$	0	0	0	0
$V_1^{\pi M}(i)$	0	0	-5	0
$V_2^{\pi M}(i)$	0	-5γ	-5	0
$V_3^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_6^{\pi M}(i)$	$-5\gamma^2$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_7^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_8^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3(1+\gamma^4)$
...	...	...	...	...

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi M}(i)$	0	0	0	0
$V_1^{\pi M}(i)$	0	0	-5	0
$V_2^{\pi M}(i)$	0	-5γ	-5	0
$V_3^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_6^{\pi M}(i)$	$-5\gamma^2$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_7^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_8^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3(1+\gamma^4)$
...	...	...	...	...

- Fortführung für $i = 3$ (für $k \to \infty$): $-5(1+\gamma^4(1+\gamma^4(1+\gamma^4(\ldots))))$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi M}(i)$	0	0	0	0
$V_1^{\pi M}(i)$	0	0	-5	0
$V_2^{\pi M}(i)$	0	-5γ	-5	0
$V_3^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_6^{\pi M}(i)$	$-5\gamma^2$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_7^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_8^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3(1+\gamma^4)$
...	...	...	...	...

- Fortführung für $i = 3$ (für $k \to \infty$): $-5(1+\gamma^4(1+\gamma^4(1+\gamma^4(\ldots))))$
- Fortführung für $i = 1$ (für $k \to \infty$): $-5\gamma^2(1+\gamma^4(1+\gamma^4(1+\gamma^4(\ldots))))$

Aufgabe 8: Mensa oder Pizzeria

Zustand i	1	2	3	4
$V_0^{\pi M}(i)$	0	0	0	0
$V_1^{\pi M}(i)$	0	0	-5	0
$V_2^{\pi M}(i)$	0	-5γ	-5	0
$V_3^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	0
$V_4^{\pi M}(i)$	$-5\gamma^2$	-5γ	-5	$-5\gamma^3$
$V_5^{\pi M}(i)$	$-5\gamma^2$	-5γ	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_6^{\pi M}(i)$	$-5\gamma^2$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_7^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3$
$V_8^{\pi M}(i)$	$-5\gamma^2(1+\gamma^4)$	$-5\gamma(1+\gamma^4)$	$-5(1+\gamma^4)$	$-5\gamma^3(1+\gamma^4)$
...	...	...	...	...

- Fortführung für $i = 3$ (für $k \to \infty$): $-5(1+\gamma^4(1+\gamma^4(1+\gamma^4(\dots))))$
- Fortführung für $i = 1$ (für $k \to \infty$): $-5\gamma^2(1+\gamma^4(1+\gamma^4(1+\gamma^4(\dots))))$
- Bekannt: $\lim_{k\to\infty} V_k^\pi(i) = V^\pi(i)$ für $i = 1, \dots, n$

Aufgabe 8: Mensa oder Pizzeria

- Damit ergibt sich für $i = 1$ (Hörsaal):

Aufgabe 8: Mensa oder Pizzeria

- Damit ergibt sich für $i = 1$ (Hörsaal):

$$\begin{aligned} V^{\pi_M}(1) = -5\gamma^2(1+\gamma^4(1+\gamma^4(1+\gamma^4(\ldots)))) &= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \gamma^{4k} \\ &= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \left(\gamma^4\right)^k \\ &= -5\gamma^2 \cdot \frac{1}{1-\gamma^4} \end{aligned}$$

Aufgabe 8: Mensa oder Pizzeria

- Damit ergibt sich für $i = 1$ (Hörsaal):

$$
\begin{aligned}
V^{\pi_M}(1) = -5\gamma^2(1+\gamma^4(1+\gamma^4(1+\gamma^4(\ldots)))) &= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \gamma^{4k} \\
&= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \left(\gamma^4\right)^k \\
&= -5\gamma^2 \cdot \frac{1}{1-\gamma^4}
\end{aligned}
$$

- In Analogie erhalten wir für die Strategie π_P:

Aufgabe 8: Mensa oder Pizzeria

- Damit ergibt sich für $i = 1$ (Hörsaal):

$$V^{\pi_M}(1) = -5\gamma^2(1+\gamma^4(1+\gamma^4(1+\gamma^4(\ldots)))) = -5\gamma^2 \cdot \sum_{k=0}^{\infty} \gamma^{4k}$$

$$= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \left(\gamma^4\right)^k$$

$$= -5\gamma^2 \cdot \frac{1}{1-\gamma^4}$$

- In Analogie erhalten wir für die Strategie π_P:

$$V^{\pi_P}(1) = -20\gamma^5 \cdot \sum_{k=0}^{\infty} \gamma^{10k}$$

Aufgabe 8: Mensa oder Pizzeria

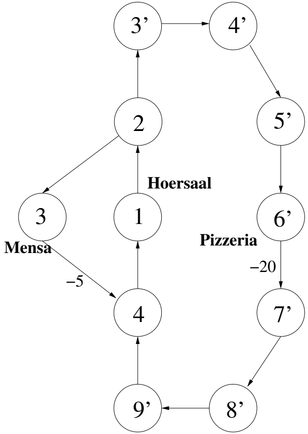
- Damit ergibt sich für $i = 1$ (Hörsaal):

$$\begin{aligned} V^{\pi_M}(1) = -5\gamma^2(1+\gamma^4(1+\gamma^4(1+\gamma^4(\dots)))) &= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \gamma^{4k} \\ &= -5\gamma^2 \cdot \sum_{k=0}^{\infty} \left(\gamma^4\right)^k \\ &= -5\gamma^2 \cdot \frac{1}{1-\gamma^4} \end{aligned}$$

- In Analogie erhalten wir für die Strategie π_P:

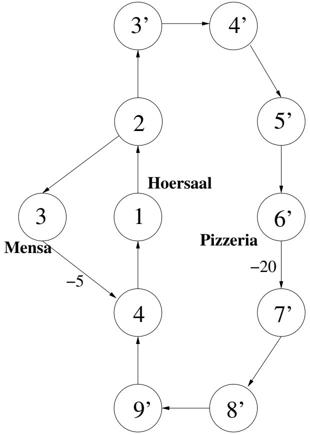
$$V^{\pi_P}(1) = -20\gamma^5 \cdot \sum_{k=0}^{\infty} \gamma^{10k} = -20\gamma^5 \cdot \sum_{k=0}^{\infty} \left(\gamma^{10}\right)^k = -20\gamma^5 \cdot \frac{1}{1-\gamma^{10}}$$

Aufgabe 8: Mensa oder Pizzeria



(c) Für welche Werte von γ wird die optimale Strategie den Agenten in die Mensa führen?
Für welche Werte in die Pizzeria?

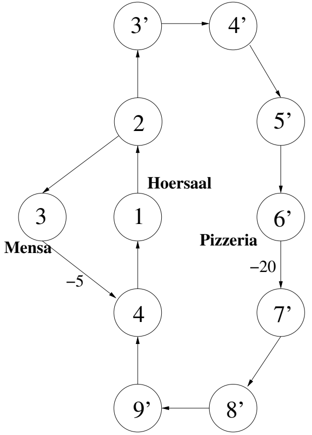
Aufgabe 8: Mensa oder Pizzeria



(c) Für welche Werte von γ wird die optimale Strategie den Agenten in die Mensa führen?
Für welche Werte in die Pizzeria?

- Für $V^{\pi_M}(2) < V^{\pi_P}(2)$ wird die Mensa bevorzugt.
- Es ist $V^{\pi_M}(2) = -5\gamma \cdot \frac{1}{1-\gamma^4}$
- Es ist $V^{\pi_P}(2) = -20\gamma^4 \cdot \frac{1}{1-\gamma^{10}}$

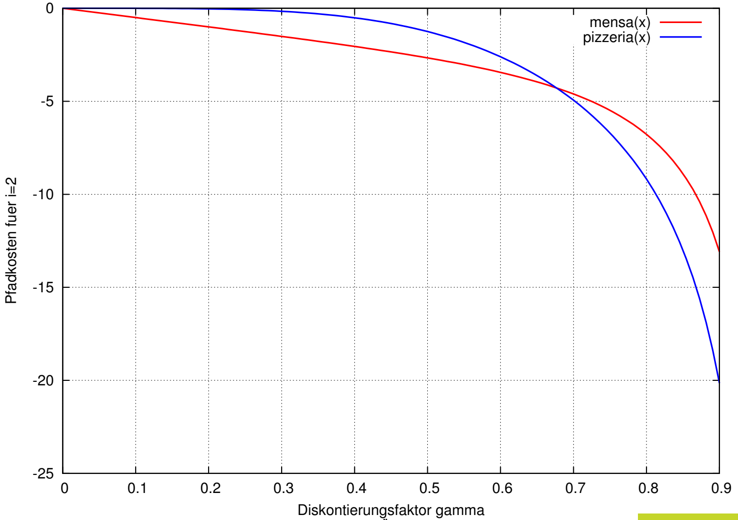
Aufgabe 8: Mensa oder Pizzeria



(c) Für welche Werte von γ wird die optimale Strategie den Agenten in die Mensa führen?
Für welche Werte in die Pizzeria?

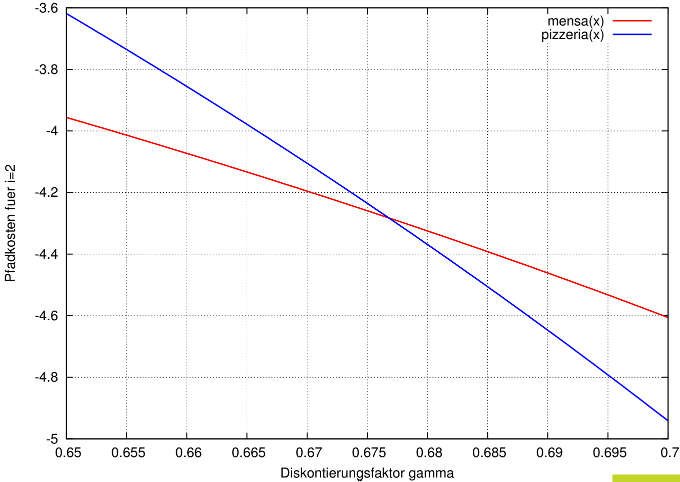
- Für $V^{\pi_M}(2) < V^{\pi_P}(2)$ wird die Mensa bevorzugt.
- Es ist $V^{\pi_M}(2) = -5\gamma \cdot \frac{1}{1-\gamma^4}$
- Es ist $V^{\pi_P}(2) = -20\gamma^4 \cdot \frac{1}{1-\gamma^{10}}$
- Näherungslösung: nächste Folie

Aufgabe 8: Mensa oder Pizzeria



Aufgabe 8: Mensa oder Pizzeria

- Für $\gamma >\approx 0.677$ wird die Pizzeria bevorzugt.



Aufgabenblatt 3

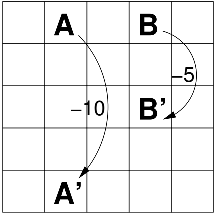
1. Aufgabenblatt 3 – Übung 7
2. Aufgabenblatt 3 – Übung 8
3. **Aufgabenblatt 3 – Übung 9**
4. Aufgabenblatt 3 – Übung 10

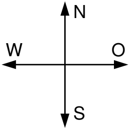
Aufgabe 9: Bellman-Gleichung

In der in der Abbildung dargestellten Gitterwelt sind die Zellen des Gitters die Zustände. In jeder Zelle sind vier Aktionen möglich (Norden, Süden, Osten, Westen), die den Agenten deterministisch in die jeweilige Nachbarzelle des Gitters bewegen. Aktionen, bei denen der Agent das Gitter verlassen würde, führen zu keinem Zellenwechsel, verursachen aber direkte Kosten in Höhe von 1. Alle anderen Aktionen sind mit keinen Kosten verbunden, mit Ausnahme der Zustände A und B. Jede in A ausgeführte Aktion bewegt den Agenten nach A' unter Kosten von -10, jede in B ausgeführte Aktion bewegt den Agenten nach B' unter Kosten von -5.

Aufgabe 9: Bellman-Gleichung

In der in der Abbildung dargestellten Gitterwelt sind die Zellen des Gitters die Zustände. In jeder Zelle sind vier Aktionen möglich (Norden, Süden, Osten, Westen), die den Agenten deterministisch in die jeweilige Nachbarzelle des Gitters bewegen. Aktionen, bei denen der Agent das Gitter verlassen würde, führen zu keinem Zellenwechsel, verursachen aber direkte Kosten in Höhe von 1. Alle anderen Aktionen sind mit keinen Kosten verbunden, mit Ausnahme der Zustände A und B. Jede in A ausgeführte Aktion bewegt den Agenten nach A' unter Kosten von -10, jede in B ausgeführte Aktion bewegt den Agenten nach B' unter Kosten von -5.

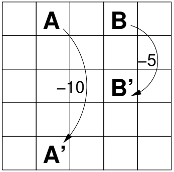


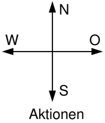


Aktionen

–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

Aufgabe 9: Bellman-Gleichung

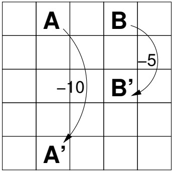


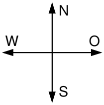


–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(a) Wir nehmen an, dass der Agent in allen Zuständen alle Aktionen mit gleicher Wahrscheinlichkeit auswählt. Im rechten Teil der Abbildung ist die zugehörige Kostenfunktion V^π für einen Diskontierungsfaktor von $\gamma = 0.9$ eingetragen. Ermitteln Sie die fehlenden Einträge.

Aufgabe 9: Bellman-Gleichung





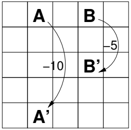
Aktionen

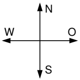
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(a) Wir nehmen an, dass der Agent in allen Zuständen alle Aktionen mit gleicher Wahrscheinlichkeit auswählt. Im rechten Teil der Abbildung ist die zugehörige Kostenfunktion V^π für einen Diskontierungsfaktor von $\gamma = 0.9$ eingetragen. Ermitteln Sie die fehlenden Einträge.

- π ist eine reine Zufalsstrategie, daher gilt $\pi(i, a) = 0.25$ für alle i und alle $a \in \{N, S, W, O\}$

Aufgabe 9: Bellman-Gleichung



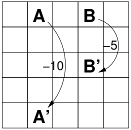


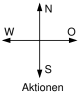
Aktionen

−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Sei $i = s_{3,2}$:

Aufgabe 9: Bellman-Gleichung



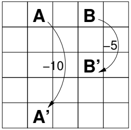


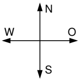
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Sei $i = s_{3,2}$:

$$V^{\pi}(i) \quad =$$

Aufgabe 9: Bellman-Gleichung





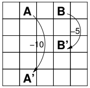
Aktionen

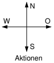
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Sei $i = s_{3,2}$:

$$\begin{aligned}
V^\pi(i) &= \sum_{a \in A(i)} \pi(i,a) \sum_{j=1}^{n} (c(i,a,j) + \gamma V^\pi(j)) \\
&= 0.25 \cdot \left(0 + \gamma V^\pi(s_{2,2})\right) + 0.25 \cdot \left(0 + \gamma V^\pi(s_{4,2})\right) \\
&\quad + 0.25 \cdot \left(0 + \gamma V^\pi(s_{3,1})\right) + 0.25 \cdot \left(0 + \gamma V^\pi(s_{3,3})\right) \\
&= 0.25 \cdot \gamma(-3.0 + 0.4 - 0.1 - 0.7) \\
&= -0.765
\end{aligned}$$

Aufgabe 9: Bellman-Gleichung

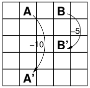


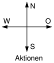


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Sei $i = s_{5,5}$:

Aufgabe 9: Bellman-Gleichung



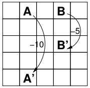


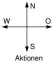
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Sei $i = s_{5,5}$:

$$V^{\pi}(i) = \sum_{a \in A(i)} \pi(i,a) \sum_{j=1}^{n} (c(i,a,j) + \gamma V^{\pi}(j))$$

Aufgabe 9: Bellman-Gleichung



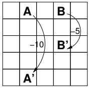


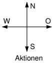
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Sei $i = s_{5,5}$:

$$\begin{aligned}
V^\pi(i) &= \sum_{a \in A(i)} \pi(i,a) \sum_{j=1}^{n} (c(i,a,j) + \gamma V^\pi(j)) \\
&= 0.25 \cdot (0 + \gamma V^\pi(s_{4,5})) + 0.25 \cdot (1 + \gamma V^\pi(s_{5,5})) \\
&\quad +0.25 \cdot (0 + \gamma V^\pi(s_{5,4})) + 0.25 \cdot (1 + \gamma V^\pi(s_{5,5})) \\
&= 0.25 \cdot (\gamma 1.2 + (1 + \gamma V^\pi(i)) + \gamma 1.4 + (1 + \gamma V^\pi(i))) \\
&= 0.25 \cdot (\gamma 2.6 + 2 + 2\gamma V^\pi(i))
\end{aligned}$$

Aufgabe 9: Bellman-Gleichung





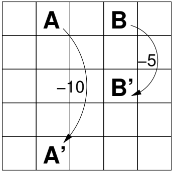
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

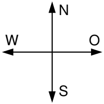
- Sei $i = s_{5,5}$:

$$
\begin{aligned}
V^\pi(i) &= \sum_{a \in A(i)} \pi(i,a) \sum_{j=1}^{n} (c(i,a,j) + \gamma V^\pi(j)) \\
&= 0.25 \cdot (0 + \gamma V^\pi(s_{4,5})) + 0.25 \cdot (1 + \gamma V^\pi(s_{5,5})) \\
&\quad + 0.25 \cdot (0 + \gamma V^\pi(s_{5,4})) + 0.25 \cdot (1 + \gamma V^\pi(s_{5,5})) \\
&= 0.25 \cdot (\gamma 1.2 + (1 + \gamma V^\pi(i)) + \gamma 1.4 + (1 + \gamma V^\pi(i))) \\
&= 0.25 \cdot (\gamma 2.6 + 2 + 2\gamma V^\pi(i))
\end{aligned}
$$

- Damit $4V^\pi(i) = 4.34 + 2\gamma V^\pi(i)$ und $V^\pi(i) = 1.973$

Aufgabe 9: Bellman-Gleichung



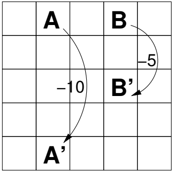


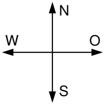
Aktionen

–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(b) Zeigen Sie beispielhaft für den Zustand $s_{3,3}$ in der Mitte des Gitters mit $V^{\pi}(s_{3,3}) = -0.7$, dass die Bellman-Gleichung bezüglich aller seiner Nachbarzustände erfüllt ist.

Aufgabe 9: Bellman-Gleichung





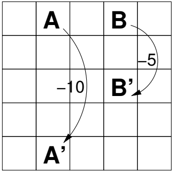
Aktionen

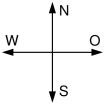
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(b) Zeigen Sie beispielhaft für den Zustand $s_{3,3}$ in der Mitte des Gitters mit $V^\pi(s_{3,3}) = -0.7$, dass die Bellman-Gleichung bezüglich aller seiner Nachbarzustände erfüllt ist.

- Alle Nachbarzustände werden mit gleicher Wahrscheinlichkeit erreicht (wegen reiner Zufallsauswahl der Aktionen unter π und wegen des Determinismus des MDP).

Aufgabe 9: Bellman-Gleichung



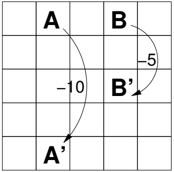


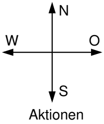
Aktionen

–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für jede mögliche Aktion a und jeden möglichen Folgezustand j von $s_{3,3}$ muss gelten
  $V(s_{3,3}) = \sum_{a \in A} \pi(s_{3,3}, a) \sum_j p_{s_{3,3},j}(a) \left(c(s_{3,3}, a, j) + \gamma V^{\pi}(j)\right)$

Aufgabe 9: Bellman-Gleichung



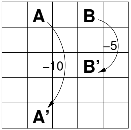


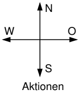
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für jede mögliche Aktion a und jeden möglichen Folgezustand j von $s_{3,3}$ muss gelten
  $V(s_{3,3}) = \sum_{a \in A} \pi(s_{3,3}, a) \sum_j p_{s_{3,3},j}(a) \left(c(s_{3,3}, a, j) + \gamma V^\pi(j)\right)$
- Da für keinen der möglichen Übergänge direkte Kosten anfallen, gilt:

$$\begin{aligned} V^\pi(s_{3,3}) &= 0.25 \cdot \gamma\,(-2.3 + 0.4 - 0.765 - 0.4) \\ &= -0.689 \approx -0.7 \end{aligned}$$

Aufgabe 9: Bellman-Gleichung

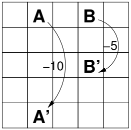


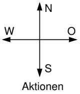


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(c) Erläutern Sie, warum unter der betrachteten Zufallsstrategie π im Zustand B die zu erwartenden Pfadkosten niedriger sind als die direkten Kosten.

Aufgabe 9: Bellman-Gleichung



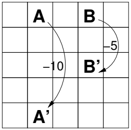


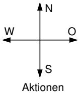
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(c) Erläutern Sie, warum unter der betrachteten Zufallsstrategie π im Zustand B die zu erwartenden Pfadkosten niedriger sind als die direkten Kosten.

- Betrachten wir zunächst den Zustand B’: Dieser hat deshalb negative erwartete Kosten, weil das Risiko von dort aus in eine der nahen Wände zu rennen (Kosten von 1), mehr als kompensiert wird durch die Chance, – zufällig – nach A oder B zu gelangen.

Aufgabe 9: Bellman-Gleichung



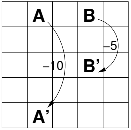


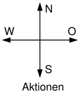
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(c) Erläutern Sie, warum unter der betrachteten Zufallsstrategie π im Zustand B die zu erwartenden Pfadkosten niedriger sind als die direkten Kosten.

- Betrachten wir zunächst den Zustand B': Dieser hat deshalb negative erwartete Kosten, weil das Risiko von dort aus in eine der nahen Wände zu rennen (Kosten von 1), mehr als kompensiert wird durch die Chance, – zufällig – nach A oder B zu gelangen.
- Im Gegensatz dazu gilt dies für den Zustand A' nicht: Hier ist die Gefahr unter der Zufallsstrategie in eine Wand zu laufen ungleich höher, weswegen die zu erwartenden Kosten von $V^\pi(A')$ größer als null sind.

Aufgabe 9: Bellman-Gleichung

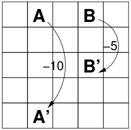


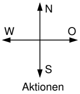


–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Betrachten wir nun Zustand B: Dieser hat niedrigere erwartete (-5.3) als direkte Kosten (-5), da der Agent im Zustand B stets nach B' versetzt wird, der ebenfalls mit negativen erwarteten Kosten (-0.4) aufwartet.

Aufgabe 9: Bellman-Gleichung

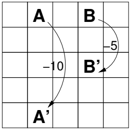


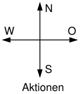


–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Betrachten wir nun Zustand B: Dieser hat niedrigere erwartete (-5.3) als direkte Kosten (-5), da der Agent im Zustand B stets nach B' versetzt wird, der ebenfalls mit negativen erwarteten Kosten (-0.4) aufwartet.
- Schließlich ist noch zu bemerken, dass der Zustand A insgesamt der "beste" Zustand ist, in dem sich der Agent unter π befinden kann, wobei jedoch – im Gegensatz zu Zustand B – seine erwarteten Kosten (-8.8) nicht so niedrig sind wie seine direkten Kosten (-10).

Aufgabe 9: Bellman-Gleichung

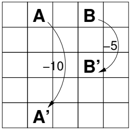


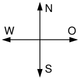


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(d) Ermitteln Sie die zu erwartenden Pfadkosten in den Zuständen A und B unter der optimalen Strategie, d.h. $V^\star(A)$ und $V^\star(B)$. Zeichnen Sie ferner die optimale Strategie in das Gitter ein.

Aufgabe 9: Bellman-Gleichung





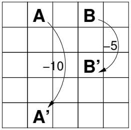
Aktionen

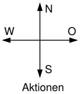
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(d) Ermitteln Sie die zu erwartenden Pfadkosten in den Zuständen A und B unter der optimalen Strategie, d.h. $V^\star(A)$ und $V^\star(B)$. Zeichnen Sie ferner die optimale Strategie in das Gitter ein.

- Frage: Wohin wird die optimale Strategie den Agenten "tendenziell" bewegen?

Aufgabe 9: Bellman-Gleichung



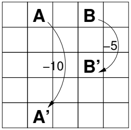


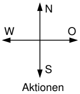
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(d) Ermitteln Sie die zu erwartenden Pfadkosten in den Zuständen A und B unter der optimalen Strategie, d.h. $V^\star(A)$ und $V^\star(B)$. Zeichnen Sie ferner die optimale Strategie in das Gitter ein.

- Frage: Wohin wird die optimale Strategie den Agenten "tendenziell" bewegen?
- Die optimale Strategie wird den Agenten "tendenziell" zum Zustand A bewegen, von wo aus hohe negative Kosten (hohe Belohnungen) zu erwarten sind.

Aufgabe 9: Bellman-Gleichung



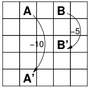


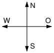
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(d) Ermitteln Sie die zu erwartenden Pfadkosten in den Zuständen A und B unter der optimalen Strategie, d.h. $V^{\star}(A)$ und $V^{\star}(B)$. Zeichnen Sie ferner die optimale Strategie in das Gitter ein.

- Frage: Wohin wird die optimale Strategie den Agenten “tendenziell” bewegen?
- Die optimale Strategie wird den Agenten “tendenziell” zum Zustand A bewegen, von wo aus hohe negative Kosten (hohe Belohnungen) zu erwarten sind.
- Insbesondere wird eine optimale Strategie den Agenten von A’ aus stets nach Norden in Richtung A bewegen.

Aufgabe 9: Bellman-Gleichung



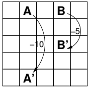


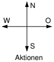
Aktionen

−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für $A = s_{1,2}$ gilt:

Aufgabe 9: Bellman-Gleichung



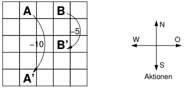


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für $A = s_{1,2}$ gilt:

$$V^{\pi}(A) \quad =$$

Aufgabe 9: Bellman-Gleichung

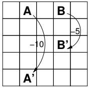


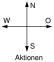
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für $A = s_{1,2}$ gilt:

$$V^\pi(A) \quad = \quad -10 + \gamma V^\pi(A') \qquad (\text{Anm.: } A' = s_{5,2})$$

Aufgabe 9: Bellman-Gleichung



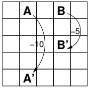


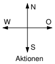
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für $A = s_{1,2}$ gilt:

$$\begin{aligned} V^\pi(A) &= -10 + \gamma V^\pi(A') \qquad (\text{Anm.: } A' = s_{5,2}) \\ &= -10 + \gamma(0 + \gamma V^\pi(s_{4,2})) \\ &= -10 + \gamma(0 + \gamma(0 + \gamma V^\pi(s_{3,2}))) \end{aligned}$$

Aufgabe 9: Bellman-Gleichung



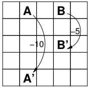


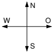
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- Für $A = s_{1,2}$ gilt:

$$
\begin{aligned}
V^\pi(A) &= -10 + \gamma V^\pi(A') \qquad (\text{Anm.: } A' = s_{5,2}) \\
&= -10 + \gamma(0 + \gamma V^\pi(s_{4,2})) \\
&= -10 + \gamma(0 + \gamma(0 + \gamma V^\pi(s_{3,2}))) \\
&= -10 + \gamma(0 + \gamma(0 + \gamma(0 + \gamma V^\pi(s_{2,2})))) \\
&= -10 + \gamma(0 + \gamma(0 + \gamma(0 + \gamma(0 + \gamma V^\pi(A)))))
\end{aligned}
$$

Aufgabe 9: Bellman-Gleichung





Aktionen

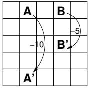
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

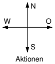
- Für $A = s_{1,2}$ gilt:

$$\begin{aligned}
V^\pi(A) &= -10 + \gamma V^\pi(A') \quad \text{(Anm.: } A' = s_{5,2}\text{)} \\
&= -10 + \gamma(0 + \gamma V^\pi(s_{4,2})) \\
&= -10 + \gamma(0 + \gamma(0 + \gamma V^\pi(s_{3,2}))) \\
&= -10 + \gamma(0 + \gamma(0 + \gamma(0 + \gamma V^\pi(s_{2,2})))) \\
&= -10 + \gamma(0 + \gamma(0 + \gamma(0 + \gamma(0 + \gamma V^\pi(A)))))
\end{aligned}$$

- Also $V^\pi(A) = -10 + \gamma^5 V^\pi(A)$ und damit $V^\pi(A) = \frac{-10}{1-\gamma^5} = -24.4$.

Aufgabe 9: Bellman-Gleichung

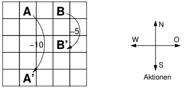




−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- In Analogie ergäbe sich für B
  $V^{\pi}(B) =$

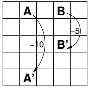
Aufgabe 9: Bellman-Gleichung



−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- In Analogie ergäbe sich für B
  $V^\pi(B) = -5 + \gamma^3 V^\pi(B)$ und damit $V^\pi(B) = \frac{-5}{1-\gamma^3} = -18.5$.

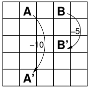
Aufgabe 9: Bellman-Gleichung



−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- In Analogie ergäbe sich für B
  $V^\pi(B) = -5 + \gamma^3 V^\pi(B)$ und damit $V^\pi(B) = \frac{-5}{1-\gamma^3} = -18.5$.
- Allerdings würde dies voraussetzen, dass sich der Agent unter der optimalen Strategie von B’ aus stets nach Norden bewegt. Frage: Aber wäre dies eine optimale Strageie?

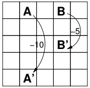
Aufgabe 9: Bellman-Gleichung

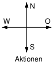


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- In Analogie ergäbe sich für B
  $V^\pi(B) = -5 + \gamma^3 V^\pi(B)$ und damit $V^\pi(B) = \frac{-5}{1-\gamma^3} = -18.5$.
- Allerdings würde dies voraussetzen, dass sich der Agent unter der optimalen Strategie von B' aus stets nach Norden bewegt.
  Frage: Aber wäre dies eine optimale Strageie?
- Nein, denn die langfristig niedrigeren Kosten ergeben sich in A; die optimale Strategie würde beispielsweise den Agenten im Zustand $s_{2,4}$ nach Westen bewegen.
  $\rightarrow$ siehe Tafelbild

Aufgabe 9: Bellman-Gleichung



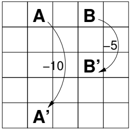


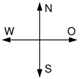
−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

- In Analogie ergäbe sich für B
  $V^\pi(B) = -5 + \gamma^3 V^\pi(B)$ und damit $V^\pi(B) = \frac{-5}{1-\gamma^3} = -18.5$.
- Allerdings würde dies voraussetzen, dass sich der Agent unter der optimalen Strategie von B' aus stets nach Norden bewegt.
  Frage: Aber wäre dies eine optimale Strageie?
- Nein, denn die langfristig niedrigeren Kosten ergeben sich in A; die optimale Strategie würde beispielsweise den Agenten im Zustand $s_{2,4}$ nach Westen bewegen.
  $\rightarrow$ siehe Tafelbild
- Zeichnung der optimalen Strategie $\rightarrow$ Tafel

Aufgabe 9: Bellman-Gleichung

Im Beispiel der gegebenen Gitterwelt werden Kosten größer null vergeben, wenn der Agent in eine Wand hineinläuft, Kosten kleiner null für das Erreichen der Zielpunkte und Nullkosten in allen anderen Situationen.



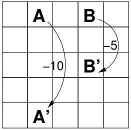


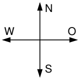
Aktionen

–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

Aufgabe 9: Bellman-Gleichung

Im Beispiel der gegebenen Gitterwelt werden Kosten größer null vergeben, wenn der Agent in eine Wand hineinläuft, Kosten kleiner null für das Erreichen der Zielpunkte und Nullkosten in allen anderen Situationen.





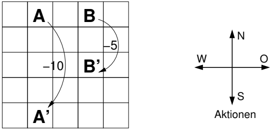
Aktionen

−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(e) Sind die Vorzeichen der direkten Kosten von Bedeutung oder aber nur die Abstände zwischen ihnen?

Aufgabe 9: Bellman-Gleichung

Im Beispiel der gegebenen Gitterwelt werden Kosten größer null vergeben, wenn der Agent in eine Wand hineinläuft, Kosten kleiner null für das Erreichen der Zielpunkte und Nullkosten in allen anderen Situationen.



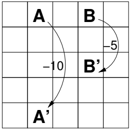
–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

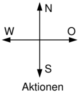
(e) Sind die Vorzeichen der direkten Kosten von Bedeutung oder aber nur die Abstände zwischen ihnen?

- Zunächst einmal ist es sinnvoll, zwischen SKP-Problemen und diskontierten Problemen zu unterscheiden.

Aufgabe 9: Bellman-Gleichung

Im Beispiel der gegebenen Gitterwelt werden Kosten größer null vergeben, wenn der Agent in eine Wand hineinläuft, Kosten kleiner null für das Erreichen der Zielpunkte und Nullkosten in allen anderen Situationen.



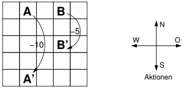


–3.3	–8.8	–4.4	–5.3	–1.5
–1.5	–3.0	–2.3	–1.9	–0.5
–0.1		–0.7	–0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(e) Sind die Vorzeichen der direkten Kosten von Bedeutung oder aber nur die Abstände zwischen ihnen?

- Zunächst einmal ist es sinnvoll, zwischen SKP-Problemen und diskontierten Problemen zu unterscheiden.
- Im Fall von SKP-Problemen ergibt es sich freilich, dass das Vorzeichen der Kosten in den Terminalzuständen eine Rolle spielt: Sind in einem absorbierenden Zustand die direkten Kosten ungleich null, so ergeben sich offensichtlich unendliche Pfadkosten.

Aufgabe 9: Bellman-Gleichung

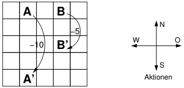


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(e) Sind die Vorzeichen der direkten Kosten von Bedeutung oder aber nur die Abstände zwischen ihnen?

- Die Fragestellung bezieht sich auf das diskontierte Beispiel aus der vorigen Teilaufgabe.

Aufgabe 9: Bellman-Gleichung

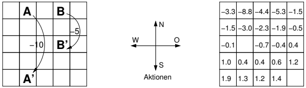


−3.3	−8.8	−4.4	−5.3	−1.5
−1.5	−3.0	−2.3	−1.9	−0.5
−0.1		−0.7	−0.4	0.4
1.0	0.4	0.4	0.6	1.2
1.9	1.3	1.2	1.4	

(e) Sind die Vorzeichen der direkten Kosten von Bedeutung oder aber nur die Abstände zwischen ihnen?

- Die Fragestellung bezieht sich auf das diskontierte Beispiel aus der vorigen Teilaufgabe.
- Für die Kostenfunktion einer Strategie spielen die Vorzeichen der direkten Kosten keine übergeordnete Rolle, ihre Abstände voneinander hingegen schon.

Aufgabe 9: Bellman-Gleichung



(e) Sind die Vorzeichen der direkten Kosten von Bedeutung oder aber nur die Abstände zwischen ihnen?

- Die Fragestellung bezieht sich auf das diskontierte Beispiel aus der vorigen Teilaufgabe.
- Für die Kostenfunktion einer Strategie spielen die Vorzeichen der direkten Kosten keine übergeordnete Rolle, ihre Abstände voneinander hingegen schon.
- Durch Verändern der relativen Beziehungen der direkten Kosten zueinander können jedoch selbstverständlich Änderungen in der Pfadkostenfunktion resultieren.

Aufgabenblatt 3

1. Aufgabenblatt 3 – Übung 7
2. Aufgabenblatt 3 – Übung 8
3. Aufgabenblatt 3 – Übung 9
4. **Aufgabenblatt 3 – Übung 10**

Spiel 10: Würfeln

Aus der Pro7-Sendung **“Schlag den Star”** ist das Spiel **“Würfeln”** bekannt.

Zwei Spieler würfeln gegeneinander, die jeweils erzielten Punkte werden summiert. Ein Spieler darf solange würfeln und weitere Punkte sammeln, wie er möchte. Aber erst, wenn er den Würfel an den anderen Spieler abgibt, werden ihm die bis dahin erwürfelten Punkte auf sein Konto gutgeschrieben.

Spiel 10: Würfeln

Aus der Pro7-Sendung **“Schlag den Star”** ist das Spiel **“Würfeln”** bekannt.

Zwei Spieler würfeln gegeneinander, die jeweils erzielten Punkte werden summiert. Ein Spieler darf solange würfeln und weitere Punkte sammeln, wie er möchte. Aber erst, wenn er den Würfel an den anderen Spieler abgibt, werden ihm die bis dahin erwürfelten Punkte auf sein Konto gutgeschrieben.

Außerdem gilt: Sobald eine 6 fällt, muss der Würfel ebenfalls an den anderen Spieler abgegeben werden, wobei in diesem Fall die bis dahin erwürfelten Punkte verfallen und nicht dem Konto des Spielers gutgeschrieben werden.

Spiel 10: Würfeln

Aus der Pro7-Sendung **“Schlag den Star”** ist das Spiel **“Würfeln”** bekannt.

Zwei Spieler würfeln gegeneinander, die jeweils erzielten Punkte werden summiert. Ein Spieler darf solange würfeln und weitere Punkte sammeln, wie er möchte. Aber erst, wenn er den Würfel an den anderen Spieler abgibt, werden ihm die bis dahin erwürfelten Punkte auf sein Konto gutgeschrieben.

Außerdem gilt: Sobald eine 6 fällt, muss der Würfel ebenfalls an den anderen Spieler abgegeben werden, wobei in diesem Fall die bis dahin erwürfelten Punkte verfallen und nicht dem Konto des Spielers gutgeschrieben werden.

Der würfelnde Spieler muss also in jedem Zeitschritt entscheiden, ob er den Würfel abgibt und die bislang erzielten Punkte einstreicht, oder ob er weiterwürfelt. Das Spiel gewinnt, wer als Erster 50 oder mehr Punkte auf seinem Konto hat.

Aufgabe 10: Würfeln

In dieser Aufgabe betrachten wir zunächst eine Vereinfachung der Aufgabenstellung, bei der nur ein einzelner Spieler beteiligt ist. Das Spiel endet zudem, sobald das erste Mal eine 6 fällt oder sobald der Spieler sich entscheidet aufzuhören und die bislang erwürfelten Punkte seinem Konto gutzuschreiben. Die 50-Punkte-Grenze entfällt also; stattdessen ist es das Ziel, möglichst viele Punkte auf seinem Konto zu haben.

Aufgabe 10: Würfeln

In dieser Aufgabe betrachten wir zunächst eine Vereinfachung der Aufgabenstellung, bei der nur ein einzelner Spieler beteiligt ist. Das Spiel endet zudem, sobald das erste Mal eine 6 fällt oder sobald der Spieler sich entscheidet aufzuhören und die bislang erwürfelten Punkte seinem Konto gutzuschreiben. Die 50-Punkte-Grenze entfällt also; stattdessen ist es das Ziel, möglichst viele Punkte auf seinem Konto zu haben.

(a) Modellieren Sie die vereinfachte Version des Spieles als SKP-Problem. Begründen Sie Ihre Definition der direkten Kosten.

- MDP $M = [T, S, A, f, c]$

Aufgabe 10: Würfeln

In dieser Aufgabe betrachten wir zunächst eine Vereinfachung der Aufgabenstellung, bei der nur ein einzelner Spieler beteiligt ist. Das Spiel endet zudem, sobald das erste Mal eine 6 fällt oder sobald der Spieler sich entscheidet aufzuhören und die bislang erwürfelten Punkte seinem Konto gutzuschreiben. Die 50-Punkte-Grenze entfällt also; stattdessen ist es das Ziel, möglichst viele Punkte auf seinem Konto zu haben.

(a) Modellieren Sie die vereinfachte Version des Spieles als SKP-Problem. Begründen Sie Ihre Definition der direkten Kosten.

- MDP $M = [T, S, A, f, c]$
- Zustandsraum S: Der Zustand des Spielers beschreibt die Anzahl der von ihm bisher erfwürfelten (und aufsummierten) Punkte.

Aufgabe 10: Würfeln

In dieser Aufgabe betrachten wir zunächst eine Vereinfachung der Aufgabenstellung, bei der nur ein einzelner Spieler beteiligt ist. Das Spiel endet zudem, sobald das erste Mal eine 6 fällt oder sobald der Spieler sich entscheidet aufzuhören und die bislang erwürfelten Punkte seinem Konto gutzuschreiben. Die 50-Punkte-Grenze entfällt also; stattdessen ist es das Ziel, möglichst viele Punkte auf seinem Konto zu haben.

(a) Modellieren Sie die vereinfachte Version des Spieles als SKP-Problem. Begründen Sie Ihre Definition der direkten Kosten.

- MDP $M = [T, S, A, f, c]$
- Zustandsraum S: Der Zustand des Spielers beschreibt die Anzahl der von ihm bisher erfwürfelten (und aufsummierten) Punkte.
- Damit ist $S = [0, i_0, i_1, i_2, i_3, \dots]$ unendlich. Der Index eines Zustdandes i_k drückt die Anzahl erwürfelter Punkte aus. Der Zustand “0” bezeichnet den Terminalzustand (Spiel beendet).

Aufgabe 10: Würfeln

- Die Menge der Aktionen

Aufgabe 10: Würfeln

- Die Menge der Aktionen umfasst nur zwei Elemente: $A = \{G, W\}$, wobei G (GIBAB) für das Abgeben des Würfelns und das damit verbundene Beenden des Spieles steht. Die Aktion W (WÜRFEL) besagt, dass der Spieler ein weiteres Mal würfelt.
- Die Übergangsfunktion

Aufgabe 10: Würfeln

- Die Menge der Aktionen umfasst nur zwei Elemente: $A = \{G, W\}$, wobei G (GIBAB) für das Abgeben des Würfelns und das damit verbundene Beenden des Spieles steht. Die Aktion W (WÜRFEL) besagt, dass der Spieler ein weiteres Mal würfelt.
- Die Übergangsfunktion ist stochastisch und überführt den Agenten in Folgezustände gemäß der folgenden Vorschrift:

$$p_{i_k,i_m}(G) = \begin{cases} 1.0 & \text{für } i_m = 0 \\ 0.0 & \text{sonst} \end{cases}$$

Aufgabe 10: Würfeln

- Die Menge der Aktionen umfasst nur zwei Elemente: $A = \{G, W\}$, wobei G (GIBAB) für das Abgeben des Würfelns und das damit verbundene Beenden des Spieles steht. Die Aktion W (WÜRFEL) besagt, dass der Spieler ein weiteres Mal würfelt.
- Die Übergangsfunktion ist stochastisch und überführt den Agenten in Folgezustände gemäß der folgenden Vorschrift:

$$p_{i_k,i_m}(G) = \begin{cases} 1.0 & \text{für } i_m = 0 \\ 0.0 & \text{sonst} \end{cases}$$

und

$$p_{i_k,i_m}(W) = \begin{cases} \frac{1}{6} & \text{für } i_m = 0 \\ \frac{1}{6} & \text{für } m - k \in \{1, \ldots, 5\} \\ 0.0 & \text{sonst} \end{cases}$$

Aufgabe 10: Würfeln

- Beispielsweise repräsentiert $p_{i_7,i_9}(a) = \frac{1}{6}$ die Wahrscheinlichkeit, eine 2 zu würfeln, wenn der Spieler aktuell 7 Punkte erwürfelt hat.

Aufgabe 10: Würfeln

- Beispielsweise repräsentiert $p_{i_7,i_9}(a) = \frac{1}{6}$ die Wahrscheinlichkeit, eine 2 zu würfeln, wenn der Spieler aktuell 7 Punkte erwürfelt hat.
- Die Kostenfunktion $c : S \times A$ muss zum Optimierungsziel korrespondieren, was im konkreten Fall bedeutet, so viele Punkte wie möglich zu erzielen.
- In der hier gegebenen Aufgabenstellung hängen die Kosten ganz entscheidend davon ab,

Aufgabe 10: Würfeln

- Beispielsweise repräsentiert $p_{i_7,i_9}(a) = \frac{1}{6}$ die Wahrscheinlichkeit, eine 2 zu würfeln, wenn der Spieler aktuell 7 Punkte erwürfelt hat.
- Die Kostenfunktion $c : S \times A$ muss zum Optimierungsziel korrespondieren, was im konkreten Fall bedeutet, so viele Punkte wie möglich zu erzielen.
- In der hier gegebenen Aufgabenstellung hängen die Kosten ganz entscheidend davon ab, unter Ausführung welcher Aktion der Agent in den Terminalzustand übergeht.
- Wir setzen daher

Aufgabe 10: Würfeln

- Beispielsweise repräsentiert $p_{i_7,i_9}(a) = \frac{1}{6}$ die Wahrscheinlichkeit, eine 2 zu würfeln, wenn der Spieler aktuell 7 Punkte erwürfelt hat.
- Die Kostenfunktion $c : S \times A$ muss zum Optimierungsziel korrespondieren, was im konkreten Fall bedeutet, so viele Punkte wie möglich zu erzielen.
- In der hier gegebenen Aufgabenstellung hängen die Kosten ganz entscheidend davon ab, unter Ausführung welcher Aktion der Agent in den Terminalzustand übergeht.
- Wir setzen daher

$$c(i_k, a, i_m) = \begin{cases} 0 & \text{falls } i_m \neq 0 \\ 0 & \text{falls } i_m = 0 \text{ und } a = W \\ -k & \text{sonst } (i_m = 0 \text{ und } a = G) \end{cases}$$

wobei klar ist, dass unter $a = G$ stets nach “0” übergegangen wird.

Aufgabe 10: Würfeln

In dieser Aufgabe betrachten wir zunächst eine Vereinfachung der Aufgabenstellung, bei der nur ein einzelner Spieler beteiligt ist. Das Spiel endet zudem, sobald das erste Mal eine 6 fällt oder sobald der Spieler sich entscheidet aufzuhören und die bislang erwürfelten Punkte seinem Konto gutzuschreiben. Die 50-Punkte-Grenze entfällt also; stattdessen ist es das Ziel, möglichst viele Punkte auf seinem Konto zu haben.

(b) Zeichnen Sie den Zustandsübergangsgraphen (ausschnittsweise) für das Problem.

Aufgabe 10: Würfeln

In dieser Aufgabe betrachten wir zunächst eine Vereinfachung der Aufgabenstellung, bei der nur ein einzelner Spieler beteiligt ist. Das Spiel endet zudem, sobald das erste Mal eine 6 fällt oder sobald der Spieler sich entscheidet aufzuhören und die bislang erwürfelten Punkte seinem Konto gutzuschreiben. Die 50-Punkte-Grenze entfällt also; stattdessen ist es das Ziel, möglichst viele Punkte auf seinem Konto zu haben.

(b) Zeichnen Sie den Zustandsübergangsgraphen (ausschnittsweise) für das Problem.

- $\rightarrow$ Tafel

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Spieler B verfolgt die Strategie, die Aktion GIBAB zu wählen, sobald er mindestens 18 Punkte erwürfelt hat. Spieler A hingegen gibt bereits ab, wenn er 10 Punkte erzielt hat. Bewerten Sie die beiden Strategien, indem Sie V^{π_A} und V^{π_B} ermitteln.

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Spieler B verfolgt die Strategie, die Aktion GIBAB zu wählen, sobald er mindestens 18 Punkte erwürfelt hat. Spieler A hingegen gibt bereits ab, wenn er 10 Punkte erzielt hat. Bewerten Sie die beiden Strategien, indem Sie V^{π_A} und V^{π_B} ermitteln.

- Frage: Ist Diskontierung angebracht?
- Antwort:

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Spieler B verfolgt die Strategie, die Aktion GIBAB zu wählen, sobald er mindestens 18 Punkte erwürfelt hat. Spieler A hingegen gibt bereits ab, wenn er 10 Punkte erzielt hat. Bewerten Sie die beiden Strategien, indem Sie V^{π_A} und V^{π_B} ermitteln.

- Frage: Ist Diskontierung angebracht?
- Antwort: Es handelt sich um ein SKP-Problem, bei dem alle Strategien erfüllend sind, Diskontierung ist nicht notwendig.

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Spieler B verfolgt die Strategie, die Aktion GIBAB zu wählen, sobald er mindestens 18 Punkte erwürfelt hat. Spieler A hingegen gibt bereits ab, wenn er 10 Punkte erzielt hat. Bewerten Sie die beiden Strategien, indem Sie V^{π_A} und V^{π_B} ermitteln.

- Frage: Ist Diskontierung angebracht?
- Antwort: Es handelt sich um ein SKP-Problem, bei dem alle Strategien erfüllend sind, Diskontierung ist nicht notwendig.
- Zur Erinnerung: Nach erfolgter Strategiebewertung gilt

$$V^\pi(i) = \sum_{j=0}^{n} p_{ij}(\pi(i)) \cdot (c(i, \pi(i), j) + \gamma V^\pi(j))$$

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Spieler B verfolgt die Strategie, die Aktion GIBAB zu wählen, sobald er mindestens 18 Punkte erwürfelt hat. Spieler A hingegen gibt bereits ab, wenn er 10 Punkte erzielt hat. Bewerten Sie die beiden Strategien, indem Sie V^{π_A} und V^{π_B} ermitteln.

- Wir betrachten zunächst die Strategie π_A, gemäß derer gilt $\pi_A(i_k) = W \Leftrightarrow k \leq 9$.
- Es ergibt sich unter π_A die folgende Markov-Kette.

Aufgabe 10: Würfeln

(c) Wir bezeichnen die Aktion des Spielers, die ihn den Würfel abgeben (und damit das Spiel beenden) lässt, im Folgenden als GIBAB.

Spieler B verfolgt die Strategie, die Aktion GIBAB zu wählen, sobald er mindestens 18 Punkte erwürfelt hat. Spieler A hingegen gibt bereits ab, wenn er 10 Punkte erzielt hat. Bewerten Sie die beiden Strategien, indem Sie V^{π_A} und V^{π_B} ermitteln.

- Wir betrachten zunächst die Strategie π_A, gemäß derer gilt $\pi_A(i_k) = W \Leftrightarrow k \leq 9$.
- Es ergibt sich unter π_A die folgende Markov-Kette. $\rightarrow$ Tafel

Aufgabe 10: Würfeln

- Wir betrachten weiter die Strategie π_A, wobei wir Strategiebewertung im Folgenden für eine besondere Reihenfolge der Zustände vornehmen, so dass sich $V^{\pi_A}(i)$ für $i \in \{0, i_0, \ldots, i_9\}$ direkt nach einem einzigen Berechnungsschritt ergibt.

Aufgabe 10: Würfeln

- Wir betrachten weiter die Strategie π_A, wobei wir Strategiebewertung im Folgenden für eine besondere Reihenfolge der Zustände vornehmen, so dass sich $V^{\pi_A}(i)$ für $i \in \{0, i_0, \dots, i_9\}$ direkt nach einem einzigen Berechnungsschritt ergibt.
- Für den Terminalzustand gilt natürlich:

Aufgabe 10: Würfeln

- Wir betrachten weiter die Strategie π_A, wobei wir Strategiebewertung im Folgenden für eine besondere Reihenfolge der Zustände vornehmen, so dass sich $V^{\pi_A}(i)$ für $i \in \{0, i_0, \dots, i_9\}$ direkt nach einem einzigen Berechnungsschritt ergibt.
- Für den Terminalzustand gilt natürlich: $V^{\pi_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 10$:

Aufgabe 10: Würfeln

- Wir betrachten weiter die Strategie π_A, wobei wir Strategiebewertung im Folgenden für eine besondere Reihenfolge der Zustände vornehmen, so dass sich $V^{\pi_A}(i)$ für $i \in \{0, i_0, \dots, i_9\}$ direkt nach einem einzigen Berechnungsschritt ergibt.
- Für den Terminalzustand gilt natürlich: $V^{\pi_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 10$: $V^{\pi_A}(i_k) = -k$.
- Wir erhalten für i_9:

Aufgabe 10: Würfeln

- Wir betrachten weiter die Strategie π_A, wobei wir Strategiebewertung im Folgenden für eine besondere Reihenfolge der Zustände vornehmen, so dass sich $V^{\pi_A}(i)$ für $i \in \{0, i_0, \ldots, i_9\}$ direkt nach einem einzigen Berechnungsschritt ergibt.
- Für den Terminalzustand gilt natürlich: $V^{\pi_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 10$: $V^{\pi_A}(i_k) = -k$.
- Wir erhalten für i_9:

$$
\begin{aligned}
V^{\pi_A}(i_9) &= \sum_{i_j} p_{i_9,i_j}(W) \cdot (0 + \gamma V^{\pi_A}(i_j)) \\
&= \frac{1}{6} \left(V^{\pi_A}(0) + V^{\pi_A}(i_{10}) + V^{\pi_A}(i_{11}) + V^{\pi_A}(i_{12}) + V^{\pi_A}(i_{13}) + V^{\pi_A}(i_{14}) \right) \\
&= \frac{1}{6} (0 - 10 - 11 - 12 - 13 - 14) = \frac{-60}{6} = -10
\end{aligned}
$$

Aufgabe 10: Würfeln

- Wir setzen analog fort für i_8:

Aufgabe 10: Würfeln

- Wir setzen analog fort für i_8:

$$
\begin{aligned}
V^{\pi A}(i_8) &= \sum_{i_j} p_{i_8,i_j}(W) \cdot (0 + \gamma V^{\pi A}(i_j)) \\
&= \frac{1}{6} (V^{\pi A}(0) + V^{\pi A}(i_9) + V^{\pi A}(i_{10}) + V^{\pi A}(i_{11}) + V^{\pi A}(i_{12}) + V^{\pi A}(i_{13})) \\
&= \frac{1}{6} \left(0 - \frac{60}{6} - 10 - 11 - 12 - 13\right) =\approx -9.33
\end{aligned}
$$

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Aufgabe 10: Würfeln

- Wir setzen analog fort für i_8:

$$
\begin{aligned}
V^{\pi_A}(i_8) &= \sum_{i_j} p_{i_8,i_j}(W) \cdot (0 + \gamma V^{\pi_A}(i_j)) \\
&= \frac{1}{6}\left(V^{\pi_A}(0) + V^{\pi_A}(i_9) + V^{\pi_A}(i_{10}) + V^{\pi_A}(i_{11}) + V^{\pi_A}(i_{12}) + V^{\pi_A}(i_{13})\right) \\
&= \frac{1}{6}\left(0 - \frac{60}{6} - 10 - 11 - 12 - 13\right) =\approx -9.33
\end{aligned}
$$

- Für alle weiteren Zustände ...
  - $V^{\pi_A}(i_7) = \frac{1}{6}(0 - 9.33 - 10 - 10 - 11 - 12) =\approx -8.72$
  - $V^{\pi_A}(i_6) = \frac{1}{6}(0 - 8.72 - 9.33 - 10 - 10 - 11) =\approx -8.18$
  - $V^{\pi_A}(i_5) = \frac{1}{6}(0 - 8.18 - 8.72 - 9.33 - 10 - 10) =\approx -7.71$
  - $V^{\pi_A}(i_4) = \frac{1}{6}(0 - 7.71 - 8.18 - 8.72 - 9.33 - 10) =\approx -7.32$

Aufgabe 10: Würfeln

- Für alle weiteren Zustände (Forts.) ...
  - $V^{\pi_A}(i_4) = \frac{1}{6}(0 - 7.71 - 8.18 - 8.72 - 9.33 - 10) =\approx -7.32$
  - $V^{\pi_A}(i_3) = \frac{1}{6}(0 - 7.32 - 7.71 - 8.18 - 8.72 - 9.33) =\approx -6.88$
  - $V^{\pi_A}(i_2) = \frac{1}{6}(0 - 6.88 - 7.32 - 7.71 - 8.18 - 8.72) =\approx -6.47$
  - $V^{\pi_A}(i_1) = \frac{1}{6}(0 - 6.47 - 6.88 - 7.32 - 7.71 - 8.18) =\approx -6.09$
  - $V^{\pi_A}(i_0) = \frac{1}{6}(0 - 6.09 - 6.47 - 6.88 - 7.32 - 7.71) =\approx -5.75$

Aufgabe 10: Würfeln

- Für alle weiteren Zustände (Forts.) ...
  - $V^{\pi_A}(i_4) = \frac{1}{6}(0 - 7.71 - 8.18 - 8.72 - 9.33 - 10) =\approx -7.32$
  - $V^{\pi_A}(i_3) = \frac{1}{6}(0 - 7.32 - 7.71 - 8.18 - 8.72 - 9.33) =\approx -6.88$
  - $V^{\pi_A}(i_2) = \frac{1}{6}(0 - 6.88 - 7.32 - 7.71 - 8.18 - 8.72) =\approx -6.47$
  - $V^{\pi_A}(i_1) = \frac{1}{6}(0 - 6.47 - 6.88 - 7.32 - 7.71 - 8.18) =\approx -6.09$
  - $V^{\pi_A}(i_0) = \frac{1}{6}(0 - 6.09 - 6.47 - 6.88 - 7.32 - 7.71) =\approx -5.75$
- In Analogie erhalten wir für π_B:
  - Für den Terminalzustand gilt natürlich: $V^{\pi_B}(0) = 0$.
  - Offensichtlich gilt für i_k mit $k \geq 18$: $V^{\pi_B}(i_k) = -k$.
  - Wir erhalten für i_{17}:

Aufgabe 10: Würfeln

- Für alle weiteren Zustände (Forts.) ...
  - $V^{\pi_A}(i_4) = \frac{1}{6}(0 - 7.71 - 8.18 - 8.72 - 9.33 - 10) =\approx -7.32$
  - $V^{\pi_A}(i_3) = \frac{1}{6}(0 - 7.32 - 7.71 - 8.18 - 8.72 - 9.33) =\approx -6.88$
  - $V^{\pi_A}(i_2) = \frac{1}{6}(0 - 6.88 - 7.32 - 7.71 - 8.18 - 8.72) =\approx -6.47$
  - $V^{\pi_A}(i_1) = \frac{1}{6}(0 - 6.47 - 6.88 - 7.32 - 7.71 - 8.18) =\approx -6.09$
  - $V^{\pi_A}(i_0) = \frac{1}{6}(0 - 6.09 - 6.47 - 6.88 - 7.32 - 7.71) =\approx -5.75$
- In Analogie erhalten wir für π_B:
  - Für den Terminalzustand gilt natürlich: $V^{\pi_B}(0) = 0$.
  - Offensichtlich gilt für i_k mit $k \geq 18$: $V^{\pi_B}(i_k) = -k$.
  - Wir erhalten für i_{17}:

$$\begin{aligned} V^{\pi_B}(i_{17}) &= \sum_{i_j} p_{i_{17},i_j}(W) \cdot (0 + \gamma V^{\pi_B}(i_j)) \\ &= \frac{1}{6}(0 - 18 - 19 - 20 - 21 - 22) = \frac{-100}{6} = -16.67 \end{aligned}$$

Aufgabe 10: Würfeln

- Für alle weiteren Zustände (Forts.) ...
  - $V^{\pi_B}(i_{16}) = \frac{1}{6}(0 - 16.67 - 18 - 19 - 20 - 21) = -15.78$
  - $V^{\pi_B}(i_{15}) = \frac{1}{6}(0 - 15.78 - 16.67 - 18 - 19 - 20) = -14.91$
  - $V^{\pi_B}(i_{14}) = \frac{1}{6}(0 - 14.91 - 15.78 - 16.67 - 18 - 19) = -14.06$
  - $V^{\pi_B}(i_{13}) = \frac{1}{6}(0 - 14.06 - 14.91 - 15.78 - 16.67 - 18) = -13.24$
  - $V^{\pi_B}(i_{12}) = \frac{1}{6}(0 - 13.24 - 14.06 - 14.91 - 15.78 - 16.67) = -12.44$
  - $V^{\pi_B}(i_{11}) = \frac{1}{6}(0 - 12.44 - 13.24 - 14.06 - 14.91 - 15.78) = -11.74$
  - $V^{\pi_B}(i_{10}) = \frac{1}{6}(0 - 11.74 - 12.44 - 13.24 - 14.06 - 14.91) = -11.07$
  - $V^{\pi_B}(i_9) = \frac{1}{6}(0 - 11.07 - 11.74 - 12.44 - 13.24 - 14.06) = -10.43$
  - $V^{\pi_B}(i_8) = \frac{1}{6}(0 - 10.43 - 11.07 - 11.74 - 12.44 - 13.24) = -9.82$

Aufgabe 10: Würfeln

- Für alle weiteren Zustände (Forts.) ...
  - $V^{\pi_B}(i_7) = \frac{1}{6}(0 - 9.82 - 10.43 - 11.07 - 11.74 - 12.44) = -9.25$
  - $V^{\pi_B}(i_6) = \frac{1}{6}(0 - 9.25 - 9.82 - 10.43 - 11.07 - 11.74) = -8.72$
  - $V^{\pi_B}(i_5) = \frac{1}{6}(0 - 8.72 - 9.25 - 9.82 - 10.43 - 11.07) = -8.22$
  - $V^{\pi_B}(i_4) = \frac{1}{6}(0 - 8.22 - 8.72 - 9.25 - 9.82 - 10.43) = -7.74$
  - $V^{\pi_B}(i_3) = \frac{1}{6}(0 - 7.74 - 8.22 - 8.72 - 9.25 - 9.82) = -7.29$
  - $V^{\pi_B}(i_2) = \frac{1}{6}(0 - 7.29 - 7.74 - 8.22 - 8.72 - 9.25) = -6.87$
  - $V^{\pi_B}(i_1) = \frac{1}{6}(0 - 6.87 - 7.29 - 7.74 - 8.22 - 8.72) = -6.47$
  - $V^{\pi_B}(i_0) = \frac{1}{6}(0 - 6.47 - 6.87 - 7.29 - 7.74 - 8.22) = -6.10$

Aufgabe 10: Würfeln

- Für alle weitere Zustände (Forts.) ...
  - $V^{\pi_B}(i_7) = \frac{1}{6}(0 - 9.82 - 10.43 - 11.07 - 11.74 - 12.44) = -9.25$
  - $V^{\pi_B}(i_6) = \frac{1}{6}(0 - 9.25 - 9.82 - 10.43 - 11.07 - 11.74) = -8.72$
  - $V^{\pi_B}(i_5) = \frac{1}{6}(0 - 8.72 - 9.25 - 9.82 - 10.43 - 11.07) = -8.22$
  - $V^{\pi_B}(i_4) = \frac{1}{6}(0 - 8.22 - 8.72 - 9.25 - 9.82 - 10.43) = -7.74$
  - $V^{\pi_B}(i_3) = \frac{1}{6}(0 - 7.74 - 8.22 - 8.72 - 9.25 - 9.82) = -7.29$
  - $V^{\pi_B}(i_2) = \frac{1}{6}(0 - 7.29 - 7.74 - 8.22 - 8.72 - 9.25) = -6.87$
  - $V^{\pi_B}(i_1) = \frac{1}{6}(0 - 6.87 - 7.29 - 7.74 - 8.22 - 8.72) = -6.47$
  - $V^{\pi_B}(i_0) = \frac{1}{6}(0 - 6.47 - 6.87 - 7.29 - 7.74 - 8.22) = -6.10$
- Damit gilt $V^{\pi_B}(i_0) < V^{\pi_A}(i_0) \approx -5.75$, weshalb π_B offenbar die erfolgsversprechendere Strategie ist.

Aufgabe 10: Würfeln

(d) Schließen Sie an Ihre in Teilaufgabe (c) durchgeführten Strategiebewertungen (policy evaluation) jeweils einen Strategieverbesserungsschritt (policy improvement) an. Ermitteln Sie die erwarteten Pfadkosten der so verbesserten Strategie. Sind die nach dem Strategieverbesserungsschritt erhaltenen Strategien optimal?

Aufgabe 10: Würfeln

(d) Schließen Sie an Ihre in Teilaufgabe (c) durchgeführten Strategiebewertungen (policy evaluation) jeweils einen Strategieverbesserungsschritt (policy improvement) an. Ermitteln Sie die erwarteten Pfadkosten der so verbesserten Strategie. Sind die nach dem Strategieverbesserungsschritt erhaltenen Strategien optimal?

- Wir nehmen einen Strategieverbesserungsschritt für π_A vor, d.h. wir ermitteln ein verbessertes π'_A durch gierige Auswertung von V^{π_A}:

Aufgabe 10: Würfeln

(d) Schließen Sie an Ihre in Teilaufgabe (c) durchgeführten Strategiebewertungen (policy evaluation) jeweils einen Strategieverbesserungsschritt (policy improvement) an. Ermitteln Sie die erwarteten Pfadkosten der so verbesserten Strategie. Sind die nach dem Strategieverbesserungsschritt erhaltenen Strategien optimal?

- Wir nehmen einen Strategieverbesserungsschritt für π_A vor, d.h. wir ermitteln ein verbessertes π'_A durch gierige Auswertung von V^{π_A}:

$$\pi'_A(i) = \arg\min_{a \in A(i)} \sum_{j \in S} p_{ij}(a)(c(i,a,j) + V^{\pi_A}(j))$$

- Wir haben $V^{\pi_A}(i_k) = -k$ für $k \geq 10$.

Aufgabe 10: Würfeln

- Für i_0 erhalten wir

Aufgabe 10: Würfeln

- Für i_0 erhalten wir

$$\begin{aligned}
\pi'_A(i_0) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & 0 \\ \text{für } a = W: & \frac{1}{6}(0 - 6.09 - 6.47 - 6.88 - 7.32 - 7.71) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & 0 \\ \text{für } a = W: & V^{\pi_A}(0) \end{cases} \\
&= W \quad (\text{denn: } V^{\pi_A}(i_0) < 0)
\end{aligned}$$

- Also keine Änderung für i_0, d.h. $\pi_A(i_0) = \pi'_A(i_0)$

Aufgabe 10: Würfeln

- Für i_0 erhalten wir

$$
\begin{aligned}
\pi'_A(i_0) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & 0 \\ \text{für } a = W: & \frac{1}{6}(0 - 6.09 - 6.47 - 6.88 - 7.32 - 7.71) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & 0 \\ \text{für } a = W: & V^{\pi_A}(0) \end{cases} \\
&= W \;\; (\text{denn: } V^{\pi_A}(i_0) < 0)
\end{aligned}
$$

- Also keine Änderung für i_0, d.h. $\pi_A(i_0) = \pi'_A(i_0)$
- Allgemein: Für die Zustände $i_k \in \{i_0, \ldots, i_9\}$ erhalten wir den Zusammenhang

Aufgabe 10: Würfeln

- Für i_0 erhalten wir

$$\begin{aligned}
\pi'_A(i_0) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & 0 \\ \text{für } a = W: & \frac{1}{6}(0 - 6.09 - 6.47 - 6.88 - 7.32 - 7.71) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & 0 \\ \text{für } a = W: & V^{\pi_A}(0) \end{cases} \\
&= W \;\; (\text{denn: } V^{\pi_A}(i_0) < 0)
\end{aligned}$$

- Also keine Änderung für i_0, d.h. $\pi_A(i_0) = \pi'_A(i_0)$
- Allgemein: Für die Zustände $i_k \in \{i_0, \ldots, i_9\}$ erhalten wir den Zusammenhang

$$\pi'_A(i_k) = \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -k \\ \text{für } a = W: & V^{\pi_A}(i_k) \end{cases}$$

.

Aufgabe 10: Würfeln

- Für jene Zustände gilt stets $-k < V^{\pi_A}(i_k)$, weshalb die Aktion W für $k \in \{0, \ldots, 9\}$ die zu wählende Aktion ist, d.h. $\pi'_A(i_k) = W$.
- Also keine Änderung für $i_k \in \{i_0, \ldots, i_9\}$, d.h. $\pi_A(i_k) = \pi'_A(i_k)$

Aufgabe 10: Würfeln

- Für jene Zustände gilt stets $-k < V^{\pi_A}(i_k)$, weshalb die Aktion W für $k \in \{0, \dots, 9\}$ die zu wählende Aktion ist, d.h. $\pi'_A(i_k) = W$.
- Also keine Änderung für $i_k \in \{i_0, \dots, i_9\}$, d.h. $\pi_A(i_k) = \pi'_A(i_k)$
- Betrachte nun i_k mit $k \geq 10$:

Aufgabe 10: Würfeln

- Für jene Zustände gilt stets $-k < V^{\pi_A}(i_k)$, weshalb die Aktion W für $k \in \{0, \ldots, 9\}$ die zu wählende Aktion ist, d.h. $\pi'_A(i_k) = W$.
- Also keine Änderung für $i_k \in \{i_0, \ldots, i_9\}$, d.h. $\pi_A(i_k) = \pi'_A(i_k)$
- Betrachte nun i_k mit $k \geq 10$:

$$
\begin{aligned}
\pi'_A(i_{10}) &= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -10 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}11\text{-}12\text{-}13\text{-}14\text{-}15) \end{cases} \\
&= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -10 \\ \text{für } a = W: & -10.83 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{10}
\end{aligned}
$$

Aufgabe 10: Würfeln

- Für jene Zustände gilt stets $-k < V^{\pi_A}(i_k)$, weshalb die Aktion W für $k \in \{0, \ldots, 9\}$ die zu wählende Aktion ist, d.h. $\pi'_A(i_k) = W$.
- Also keine Änderung für $i_k \in \{i_0, \ldots, i_9\}$, d.h. $\pi_A(i_k) = \pi'_A(i_k)$
- Betrachte nun i_k mit $k \geq 10$:

$$\begin{aligned}
\pi'_A(i_{10}) &= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -10 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}11\text{-}12\text{-}13\text{-}14\text{-}15) \end{cases} \\
&= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -10 \\ \text{für } a = W: & -10.83 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{10}
\end{aligned}$$

$$\begin{aligned}
\pi'_A(i_{11}) &= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -11 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}12\text{-}13\text{-}14\text{-}15\text{-}16)\text{=-}11.67 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{11}
\end{aligned}$$

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$\pi'_A(i_{12}) = \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -12 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}13\text{-}14\text{-}15\text{-}16\text{-}17) \end{cases}$$

$$= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } u = G: & -12 \\ \text{für } a = W: & -12.5 \end{cases}$$

$$= W \rightarrow \text{Änderung für } i_{12}$$

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$\begin{aligned}
\pi'_A(i_{12}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -12 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}13\text{-}14\text{-}15\text{-}16\text{-}17) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } u = G: & -12 \\ \text{für } a = W: & -12.5 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{12}
\end{aligned}$$

$$\begin{aligned}
\pi'_A(i_{13}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -13 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}14\text{-}15\text{-}16\text{-}17\text{-}18) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -13 \\ \text{für } a = W: & -13.33 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{13}
\end{aligned}$$

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$\begin{aligned}
\pi'_A(i_{14}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -14 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}15\text{-}16\text{-}17\text{-}18\text{-}19) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -14 \\ \text{für } a = W: & -14.17 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{14}
\end{aligned}$$

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$\begin{aligned}
\pi'_A(i_{14}) &= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G\text{:} & -14 \\ \text{für } a = W\text{:} & \frac{1}{6}(0\text{-}15\text{-}16\text{-}17\text{-}18\text{-}19) \end{cases} \\
&= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G\text{:} & -14 \\ \text{für } a = W\text{:} & -14.17 \end{cases} \\
&= W \rightarrow \text{Änderung für } i_{14}
\end{aligned}$$

$$\begin{aligned}
\pi'_A(i_{15}) &= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G\text{:} & -15 \\ \text{für } a = W\text{:} & \frac{1}{6}(0\text{-}16\text{-}17\text{-}18\text{-}19\text{-}20) \end{cases} \\
&= \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G\text{:} & -15 \\ \text{für } a = W\text{:} & -15 \end{cases} \\
&= W \text{ oder } G \rightarrow \text{evtl. Änderung für } i_{15}
\end{aligned}$$

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$
\begin{aligned}
\pi'_A(i_{16}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -16 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}17\text{-}18\text{-}19\text{-}20\text{-}21) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -16 \\ \text{für } a = W: & -15.67 \end{cases} \\
&= G \rightarrow \text{keine Änderung für } i_{16}
\end{aligned}
$$

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$\begin{aligned}
\pi'_A(i_{16}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -16 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}17\text{-}18\text{-}19\text{-}20\text{-}21) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -16 \\ \text{für } a = W: & -15.67 \end{cases} \\
&= G \rightarrow \text{keine Änderung für } i_{16}
\end{aligned}$$

- Ab dem Zustand i_{16} ist es ratsamer,

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$\begin{aligned}
\pi'_A(i_{16}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -16 \\ \text{für } a = W: & \frac{1}{6}(0\text{-}17\text{-}18\text{-}19\text{-}20\text{-}21) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G: & -16 \\ \text{für } a = W: & -15.67 \end{cases} \\
&= G \rightarrow \text{keine Änderung für } i_{16}
\end{aligned}$$

- Ab dem Zustand i_{16} ist es ratsamer, abzugeben.

Aufgabe 10: Würfeln

- Betrachte nun i_k mit $k \geq 10$:

$$
\begin{aligned}
\pi'_A(i_{16}) &= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G\text{:} & -16 \\ \text{für } a = W\text{:} & \frac{1}{6}(0\text{-}17\text{-}18\text{-}19\text{-}20\text{-}21) \end{cases} \\
&= \arg\min_{a\in\{W,G\}} \begin{cases} \text{für } a = G\text{:} & -16 \\ \text{für } a = W\text{:} & -15.67 \end{cases} \\
&= G \rightarrow \text{keine Änderung für } i_{16}
\end{aligned}
$$

- Ab dem Zustand i_{16} ist es ratsamer, abzugeben.
- Die Strategie π'_A lautet also:

$$
\pi'_A(i_k) = \begin{cases} W & \text{für } k < 15 \\ G & \text{sonst} \end{cases}
$$

Aufgabe 10: Würfeln

- Wir berechnen noch zusätzlich $V^{\pi'_A}$.

Aufgabe 10: Würfeln

- Wir berechnen noch zusätzlich $V^{\pi'_A}$.
- Für den Terminalzustand gilt natürlich: $V^{\pi'_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 15$: $V^{\pi'_A}(i_k) = -k$.
- Wir erhalten für i_{14}:

Aufgabe 10: Würfeln

- Wir berechnen noch zusätzlich $V^{\pi'_A}$.
- Für den Terminalzustand gilt natürlich: $V^{\pi'_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 15$: $V^{\pi'_A}(i_k) = -k$.
- Wir erhalten für i_{14}:

$$\begin{aligned} V^{\pi'_A}(i_{14}) &= \sum_{i_j} p_{i_{14},i_j}(W) \cdot (0 + \gamma V^{\pi'_A}(i_j)) \\ &= \frac{1}{6}(0 - 15 - 16 - 17 - 18 - 19) = \frac{-85}{6} = -14.17 \end{aligned}$$

Aufgabe 10: Würfeln

- Wir berechnen noch zusätzlich $V^{\pi'_A}$.
- Für den Terminalzustand gilt natürlich: $V^{\pi'_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 15$: $V^{\pi'_A}(i_k) = -k$.
- Wir erhalten für i_{14}:

$$\begin{aligned} V^{\pi'_A}(i_{14}) &= \sum_{i_j} p_{i_{14},i_j}(W) \cdot (0 + \gamma V^{\pi'_A}(i_j)) \\ &= \frac{1}{6}(0 - 15 - 16 - 17 - 18 - 19) = \frac{-85}{6} = -14.17 \end{aligned}$$

- Werten wir diese Wertfunktion gierig aus, so erhalten wir π''_A, wobei
  - für alle $k \leq 14$ gilt $V^{\pi'_A}(i_k) < k$, so dass für $k \leq 14$ die Aktion W niedrigere Kosten verheißt,

Aufgabe 10: Würfeln

- Wir berechnen noch zusätzlich $V^{\pi'_A}$.
- Für den Terminalzustand gilt natürlich: $V^{\pi'_A}(0) = 0$.
- Offensichtlich gilt für i_k mit $k \geq 15$: $V^{\pi'_A}(i_k) = -k$.
- Wir erhalten für i_{14}:

$$\begin{aligned} V^{\pi'_A}(i_{14}) &= \sum_{i_j} p_{i_{14},i_j}(W) \cdot (0 + \gamma V^{\pi'_A}(i_j)) \\ &= \frac{1}{6}(0 - 15 - 16 - 17 - 18 - 19) = \frac{-85}{6} = -14.17 \end{aligned}$$

- Werten wir diese Wertfunktion gierig aus, so erhalten wir π''_A, wobei
  - für alle $k \leq 14$ gilt $V^{\pi'_A}(i_k) < k$, so dass für $k \leq 14$ die Aktion W niedrigere Kosten verheißt,
  - in Analogie zur Berechnung von π' ab $k \geq 15$ gilt: $-k \leq \frac{1}{6}(0 - (k+1) - (k+2) - (k+3) - (k+4) - (k+5))$

Aufgabe 10: Würfeln

- Damit wird sich durch einen weiteren Strategieverbesserungsschritt

Aufgabe 10: Würfeln

- Damit wird sich durch einen weiteren Strategieverbesserungsschritt nichts mehr an π' ändern, d.h. $\pi' = \pi'' = \pi^*$.
- Mit π' haben wir bereits die optimale Strategie ermittelt.

Aufgabe 10: Würfeln

- Damit wird sich durch einen weiteren Strategieverbesserungsschritt nichts mehr an π' ändern, d.h. $\pi' = \pi'' = \pi^*$.
- Mit π' haben wir bereits die optimale Strategie ermittelt.
- Als Nächstes betrachten wir die Strategie π_B und die für sie ermittelte Kostenfunktion V^{π_B} und wenden einen Strategieverbesserungsschritt an.

Aufgabe 10: Würfeln

- Damit wird sich durch einen weiteren Strategieverbesserungsschritt nichts mehr an π' ändern, d.h. $\pi' = \pi'' = \pi^*$.
- Mit π' haben wir bereits die optimale Strategie ermittelt.
- Als Nächstes betrachten wir die Strategie π_B und die für sie ermittelte Kostenfunktion V^{π_B} und wenden einen Strategieverbesserungsschritt an.
- In Analogie zu unseren Überlegungen zur Strategie π_A erhalten wir auch hier den Zusammenhang

$$\pi'_B(i_k) = \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G: & -k \\ \text{für } a = W: & V^{\pi_B}(i_k) \end{cases}$$

für $k = 0, \ldots, 17$.

Aufgabe 10: Würfeln

- Damit wird sich durch einen weiteren Strategieverbesserungsschritt nichts mehr an π' ändern, d.h. $\pi' = \pi'' = \pi^*$.
- Mit π' haben wir bereits die optimale Strategie ermittelt.
- Als Nächstes betrachten wir die Strategie π_B und die für sie ermittelte Kostenfunktion V^{π_B} und wenden einen Strategieverbesserungsschritt an.
- In Analogie zu unseren Überlegungen zur Strategie π_A erhalten wir auch hier den Zusammenhang

$$\pi'_B(i_k) = \arg\min_{a \in \{W,G\}} \begin{cases} \text{für } a = G\text{:} & -k \\ \text{für } a = W\text{:} & V^{\pi_B}(i_k) \end{cases}$$

für $k = 0, \ldots, 17$.

- Daher:
  $V^{\pi_B}(i_{17}) = -16.67,\ V^{\pi_B}(i_{16}) = -15.78, \ldots$

Aufgabe 10: Würfeln

- Daher:
  $V^{\pi_B}(i_{17}) = -16.67$, $V^{\pi_B}(i_{16}) = -15.78$, $V^{\pi_B}(i_{15}) = -14.91$,
  $V^{\pi_B}(i_{14}) = -14.06$, $V^{\pi_B}(i_{13}) = -13.24$, $V^{\pi_B}(i_{12}) = -12.44, \ldots$

Aufgabe 10: Würfeln

- Daher:
  $V^{\pi_B}(i_{17}) = -16.67$, $V^{\pi_B}(i_{16}) = -15.78$, $V^{\pi_B}(i_{15}) = -14.91$,
  $V^{\pi_B}(i_{14}) = -14.06$, $V^{\pi_B}(i_{13}) = -13.24$, $V^{\pi_B}(i_{12}) = -12.44, \ldots$
- Wir erhalten dadurch für π'_B: $\pi'_B(i_k) = G$ für $k \geq 15$ und $\pi'_B(i_k) = W$ für $k \leq 14$.

$$\pi'_B(i_k) = \begin{cases} W & \text{für } k < 15 \\ G & \text{sonst} \end{cases}$$

Aufgabe 10: Würfeln

- Daher:
  $V^{\pi_B}(i_{17}) = -16.67$, $V^{\pi_B}(i_{16}) = -15.78$, $V^{\pi_B}(i_{15}) = -14.91$,
  $V^{\pi_B}(i_{14}) = -14.06$, $V^{\pi_B}(i_{13}) = -13.24$, $V^{\pi_B}(i_{12}) = -12.44, \ldots$
- Wir erhalten dadurch für π'_B: $\pi'_B(i_k) = G$ für $k \geq 15$ und $\pi'_B(i_k) = W$ für $k \leq 14$.

$$\pi'_B(i_k) = \begin{cases} W & \text{für } k < 15 \\ G & \text{sonst} \end{cases}$$

- Es gilt also offensichtlich $\pi'_B = \pi'_A$, und insbesondere auch $\pi'_B = \pi'_A = \pi^*$.