

Maschinelles Lernen:

Wie funktioniert Q-Learning?

Manuel Barbi

Motivation

- **DeepMind Technologies:**
 - gegründet als britisches Start-up in 2010
 - übernommen von Google in 2014
- **Lernen von Atari 2600 Games direkt auf Pixeldaten**
 - durch Kombination von Q-Learning und Convolutional Neural Networks
 - bedeutender Schritt in Richtung General AI
- **Entwicklung von Alpha Go**

Agenda

Maschinelles Lernen

Q-Learning

**Exkurs: Q-Learning
mit Neuronalen Netzen**

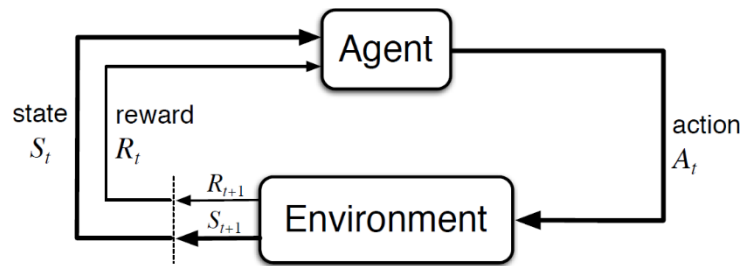
**Q-Learning mit
Anki Overdrive**

Maschinelles Lernen

- **Supervised Learning:**
 - Lernen anhand von „beschrifteten“ Beispielen
 - Bereitstellung von Eingabe-Ausgabe-Paaren (z.B. Bilderkennung)
- **Unsupervised Learning:**
 - Erkennen von Mustern und Features
- **Reinforcement Learning:**
 - Selbständiges Lernen aus Erfahrung
 - anhand von Rewards für jede Aktion (teilweise zeitlich verzögert)

Reinforcement Learning

- Wird häufig mit der Art verglichen, wie Tiere lernen



Quelle: Reinforcement Learning: An Introduction,
Richard S. Sutton and Andrew G. Barto

- **Zustandstransition** (s, a, r, s')
- **Episode** $((s_0, a_0, r_1, s_1), (s_1, a_1, r_2, s_2), \dots, (s_{n-1}, a_{n-1}, r_n, s_n))$
- **Ziel:** Maximiere die Summe der Rewards

$$\sum_{t=1}^n r_t$$

Reinforcement Learning

- **Rewards:** Direktes Feedback, wie gut eine Aktion war
- Abschätzen einer **Value-Funktion:**
 - Langfristiger Wert eines Zustandes bzw. Zustands-Aktions-Paar
- **Discounted Rewards** als Maß für den **Value**

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} = \gamma^0 r_{t+1} + \gamma^1 r_{t+2} + \dots, \gamma \in [0, 1]$$

- Zustandstransitionen im Allgemeinen stochastisch
- Einfluss von Rewards nimmt ab, je weiter diese in der Zukunft liegen

Reinforcement Learning

- **Policy π :** Definiert die Vorgehensweise des Agenten
 - Abbildung eines Zustandes auf eine Aktion oder Wahrscheinlichkeitsverteilung über Aktionen
 - Üblicherweise wird die Aktion mit dem höchsten **Value** ausgeführt

Discounted Rewards

$$\begin{aligned} G_t &= \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} = \\ &\gamma^0 r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} \dots = \\ &r_{t+1} + \gamma \left(r_{t+2} + \gamma (r_{t+3} + \gamma (r_{t+4} \dots)) \right) = \\ &r_{t+1} + \gamma G_{t+1} \end{aligned}$$

Agenda

Maschinelles Lernen

Q-Learning

**Exkurs: Q-Learning
mit Neuronalen Netzen**

**Q-Learning mit
Anki Overdrive**

Q-Learning

- **Modellfreies Reinforcement Learning Verfahren**

- Nähert iterativ eine Value-Funktion an

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

- **Learning-Rate** $\alpha \in [0,1]$
- **Discount-Rate** $\gamma \in [0,1]$
- Schätzt den Erwartungswert der zukünftig erreichbaren Rewards eines Zustands-Aktions-Paares (s, a)

Q-Learning

- **Explore-Exploit-Dilemma**

- Folge dem bisher Gelernten um Rewards zu maximieren
versus
- Erkunde neue Bereiche, um potenzielle Verbesserungen ausfindig zu machen

- **ϵ -Greedy-Policy**








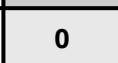
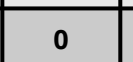


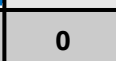

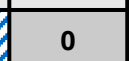

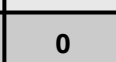


- Führe in der Regel Aktion mit höchstem Q-Value aus
- Führe mit Wahrscheinlichkeit $\epsilon \in [0,1]$ stattdessen eine zufällige Aktion aus

Q-Learning (Ablauf)

- **Der Agent nimmt aktuellen Zustand s wahr**
- **Policy π gibt die nächste Aktion a für Zustand s vor**
 - (In der Regel die Aktion mit dem höchsten Value)
- **Der Agent führt Aktion a aus, erhält Reward r und befindet sich dann im Folgezustand s'**
- **Der Agent aktualisiert den Q-Value für das Zustands-Aktions-Paar (s, a)**
- **Vorgang wird wiederholt bis die Aufgabe erfolgreich gelöst oder gescheitert ist**

Wall-E möchte nach Hause

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
|  | -1 |  -100 |  0 |

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | 0 | 0 |  |  |
| (1,2) | 0 | 0 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | 0 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |








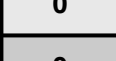
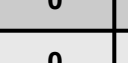

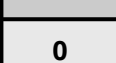
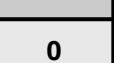

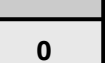
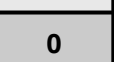



Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
|  -1 | -1 |  -100 |  0 |

$$Q'((1, 1), \text{up})$$

| | up | right | down | left |
|-------|---|---|---|---|
| (1,1) | 0 | 0 |  |  |
| (1,2) | 0 | 0 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | 0 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |








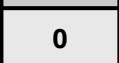
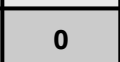


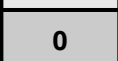

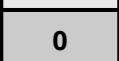

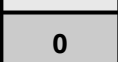

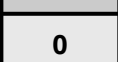
Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
|  | -1 |  -100 |  0 |

$$Q'((1, 1), \text{up}) = 0.9 \times 0 + 0.1(-100 + 0.9 \times 0)$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | 0 | 0 |  |  |
| (1,2) | 0 | 0 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | 0 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |








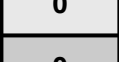
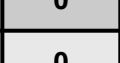


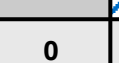

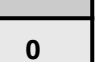
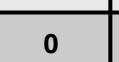
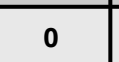


Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
|  | -1 |  -100 |  0 |


$$Q'((1, 1), \text{up}) = 0.9 \times 0 + 0.1(-100 + 0.9 \times 0) = -10$$

| | up | right | down | left |
|-------|---|---|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | 0 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | 0 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |

Wall-E möchte nach Hause









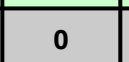


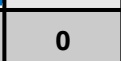

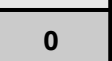
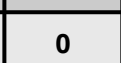
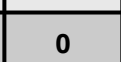

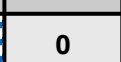
$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((1, 2), \text{right}) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$







| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | 0 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | 0 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |









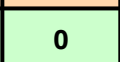


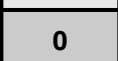

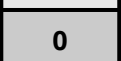

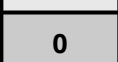

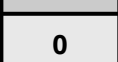
Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|---|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 |  | -1 | -1 |
| -1 | -1 |  -100 |  0 |


$$Q'((2,2), up) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | 0 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |








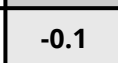
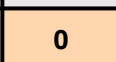

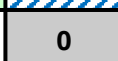


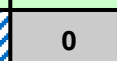

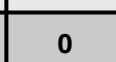

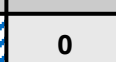
Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|--|---|---|
|  -100 |  -1 | -1 | -1 |
|  -100 |  |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |


$$Q'((2, 3), up) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | 0 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |



















Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|---|--|---|
|  -100 |  | -1 | -1 |
| -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 | -100 |  0 |

$$Q'((2,4), right) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|---|---|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | 0 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |








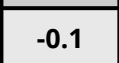
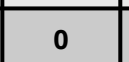


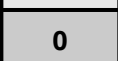

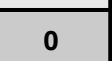

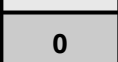


Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 |  | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((3,4), right) = 0.9 \times \text{orange} + 0.1(-1 + 0.9 \times \text{green}) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | 0 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |








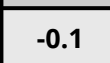
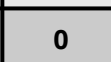


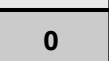

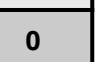

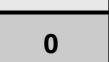


Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 |  |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4, 4), \text{down}) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | 0 | 0 |









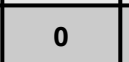


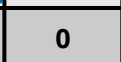

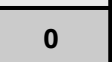
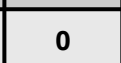
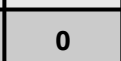


Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 |  |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4, 3), up) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | 0 |  | 0 | 0 |
| (4,4) |  |  | -0.1 | 0 |

Wall-E möchte nach Hause








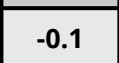
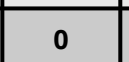


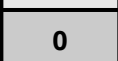

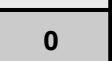

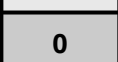


$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 |  |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4,4), \text{down}) = 0.9 \times -0.1 + 0.1(-1 + 0.9 \times 0) = -0.19$$



| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | -0.1 |  | 0 | 0 |
| (4,4) |  |  | -0.1 | 0 |









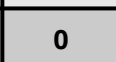


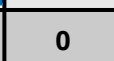

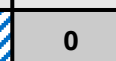

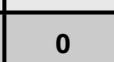

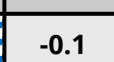
Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 |  |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4, 3), \text{down}) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | -0.1 |  | 0 | 0 |
| (4,4) |  |  | -0.19 | 0 |









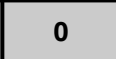








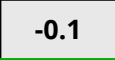
Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 |  |
| -1 | -1 |  -100 |  0 |

$$Q'((4, 2), up) = 0.9 \times 0 + 0.1(-1 + 0.9 \times 0) = -0.1$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | 0 |  | 0 | 0 |
| (4,3) | -0.1 |  | -0.1 | 0 |
| (4,4) |  |  | -0.19 | 0 |









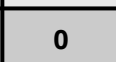


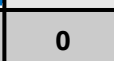

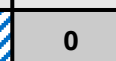

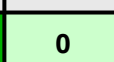

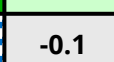
Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 |  |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |




$$Q'((4, 3), left) = 0.9 \times 0 + 0.1(-100 + 0.9 \times 0) = -10$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.1 |  | -0.1 | 0 |
| (4,4) |  |  | -0.19 | 0 |

Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((3,3), up) = 0.9 \times \text{orange} + 0.1(-1 + 0.9 \times \text{green}) = -0.1$$



| | up | right | down | left |
|-------|------|-------|-------|------|
| (1,1) | -10 | 0 | | |
| (1,2) | 0 | -0.1 | 0 | |
| (1,3) | 0 | 0 | 0 | |
| (1,4) | | 0 | 0 | |
| (2,1) | 0 | 0 | | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) | | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 | | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | 0 | 0 | 0 | 0 |
| (3,4) | | -0.1 | 0 | 0 |
| (4,1) | | | | |
| (4,2) | -0.1 | | 0 | 0 |
| (4,3) | -0.1 | | -0.1 | -10 |
| (4,4) | | | -0.19 | 0 |








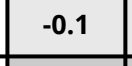
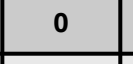

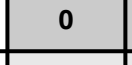
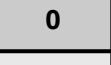

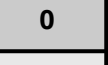




Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 |  | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |







$$Q'((3,4), \text{down}) = 0.9 \times \text{orange} + 0.1(-100 + 0.9 \times \text{green}) = -10$$

| | up | right | down | left |
|-------|---|---|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | 0 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.1 |  | -0.1 | -10 |
| (4,4) |  |  | -0.19 | 0 |









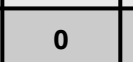


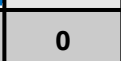

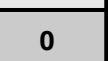
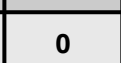



Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((3,3), \text{right}) = 0.9 \times \text{orange} + 0.1(-1 + 0.9 \times -0.1) \approx -0.11$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | 0 | 0 | 0 |
| (3,4) |  | -0.1 | -10 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.1 |  | -0.1 | -10 |
| (4,4) |  |  | -0.19 | 0 |








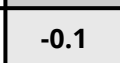
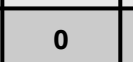


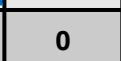

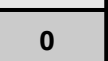
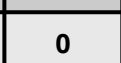
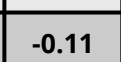


Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 |  |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4, 3), up) = 0.9 \times -0.1 + 0.1(-1 + 0.9 \times 0) = -0.19$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | -0.11 | 0 | 0 |
| (3,4) |  | -0.1 | -10 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.1 |  | -0.1 | -10 |
| (4,4) |  |  | -0.19 | 0 |

Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$








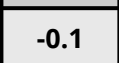
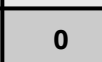


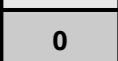

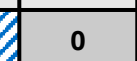

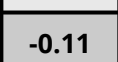


$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 |  |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4, 4), \text{down}) = 0.9 \times -0.19 + 0.1(-1 + 0.9 \times -0.1)$$

$$= -0.28$$



| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | -0.11 | 0 | 0 |
| (3,4) |  | -0.1 | -10 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.19 |  | -0.1 | -10 |
| (4,4) |  |  | -0.19 | 0 |








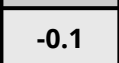
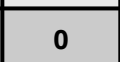


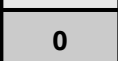

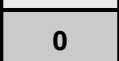

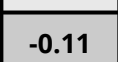


Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 |  |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  0 |

$$Q'((4,3), \text{down}) = 0.9 \times -0.1 + 0.1(-1 + 0.9 \times 0) = -0.19$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | -0.11 | 0 | 0 |
| (3,4) |  | -0.1 | -10 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.19 |  | -0.1 | -10 |
| (4,4) |  |  | -0.28 | 0 |



















Wall-E möchte nach Hause

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \times \max_{a'} Q(s', a'))$$

$$\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.25$$

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 |  |
| -1 | -1 |  -100 |  |









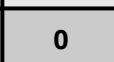


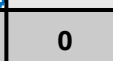

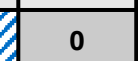

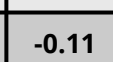


$$Q'((4, 2), \text{down}) = 0.9 \times \text{orange} + 0.1(\text{blue} + 0.9 \times \text{green}) = 0.0$$

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | -0.11 | 0 | 0 |
| (3,4) |  | -0.1 | -10 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0 | 0 |
| (4,3) | -0.19 |  | -0.19 | -10 |
| (4,4) |  |  | -0.28 | 0 |

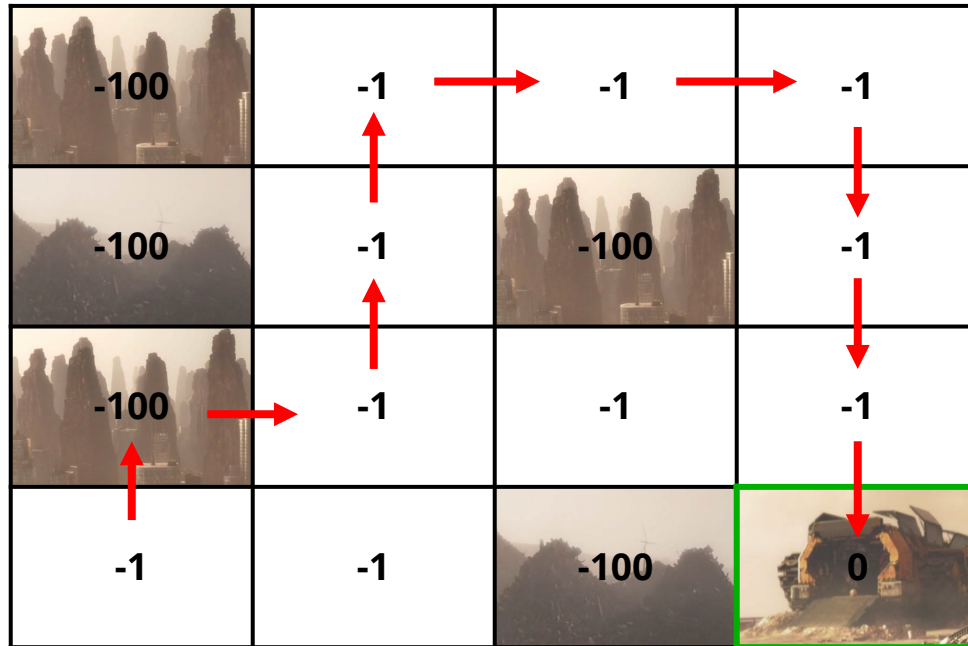
Wall-E möchte nach Hause

| | | | |
|---|----|---|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
| -1 | -1 |  -100 |  |

Puh, endlich geschafft!

| | up | right | down | left |
|-------|--|--|---|---|
| (1,1) | -10 | 0 |  |  |
| (1,2) | 0 | -0.1 | 0 |  |
| (1,3) | 0 | 0 | 0 |  |
| (1,4) |  | 0 | 0 |  |
| (2,1) | 0 | 0 |  | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) |  | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 |  | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | -0.11 | 0 | 0 |
| (3,4) |  | -0.1 | -10 | 0 |
| (4,1) |  |  |  |  |
| (4,2) | -0.1 |  | 0.0 | 0 |
| (4,3) | -0.19 |  | -0.19 | -10 |
| (4,4) |  |  | -0.28 | 0 |

Wall-E möchte nach Hause



Was wurde bisher gelernt?

| | up | right | down | left |
|-------|-------|-------|-------|------|
| (1,1) | -10 | 0 | | |
| (1,2) | 0 | -0.1 | 0 | |
| (1,3) | 0 | 0 | 0 | |
| (1,4) | | 0 | 0 | |
| (2,1) | 0 | 0 | | 0 |
| (2,2) | -0.1 | 0 | 0 | 0 |
| (2,3) | -0.1 | 0 | 0 | 0 |
| (2,4) | | -0.1 | 0 | 0 |
| (3,1) | 0 | 0 | | 0 |
| (3,2) | 0 | 0 | 0 | 0 |
| (3,3) | -0.1 | -0.11 | 0 | 0 |
| (3,4) | | -0.1 | -10 | 0 |
| (4,1) | | | | |
| (4,2) | -0.1 | | 0.0 | 0 |
| (4,3) | -0.19 | | -0.19 | -10 |
| (4,4) | | | -0.28 | 0 |

Wall-E möchte nach Hause



16 Durchläufe später ...

| | up | right | down | left |
|-------|--------|-------|--------|--------|
| (1,1) | -52.32 | -1.98 | | |
| (1,2) | -10 | -0.87 | -0.9 | |
| (1,3) | -10 | -0.57 | -10 | |
| (1,4) | | -0.49 | -10 | |
| (2,1) | -1.66 | -27.1 | | -1.67 |
| (2,2) | -1.25 | -1.24 | -1.28 | -41.1 |
| (2,3) | -1.13 | -19 | -1.12 | -41 |
| (2,4) | | -1.03 | -1.02 | -34.45 |
| (3,1) | -0.1 | 0.0 | | -0.25 |
| (3,2) | -10 | -0.65 | -34.39 | -0.85 |
| (3,3) | -0.22 | -0.11 | -0.1 | -0.1 |
| (3,4) | | -0.92 | -10 | -0.94 |
| (4,1) | | | | |
| (4,2) | -0.1 | | 0.0 | -0.11 |
| (4,3) | -0.42 | | -0.41 | -10 |
| (4,4) | | | -0.67 | -0.79 |

Wall-E möchte nach Hause

| | | | |
|---|----|--|---|
|  -100 | -1 | -1 | -1 |
|  -100 | -1 |  -100 | -1 |
|  -100 | -1 | -1 | -1 |
|  | -1 | -100 |  0 |

Nach 100.000 Durchläufen ...

$$\sum_{t=1}^n \gamma^{t-1} r_t = 0.9^0 \times -1 + 0.9^1 \times -1 + 0.9^2 \times -1 + 0.9^3 \times -1 + 0.9^4 \times 0 \approx -3.44$$

| | up | right | down | left |
|-------|--------|--------|--------|--------|
| (1,1) | -102.4 | -3.44 | | |
| (1,2) | -103.1 | -2.71 | -4.1 | |
| (1,3) | -103.7 | -3.44 | -102.4 | |
| (1,4) | | -4.1 | -95.67 | |
| (2,1) | -2.71 | -99.99 | | -4.1 |
| (2,2) | -3.44 | -1.9 | -3.44 | -102.4 |
| (2,3) | -4.1 | -101.7 | -2.71 | -103.1 |
| (2,4) | | -3.44 | -3.44 | -103.7 |
| (3,1) | -1.9 | 0.0 | | -3.44 |
| (3,2) | -101.7 | -0.99 | -99.99 | -2.71 |
| (3,3) | -3.44 | -1.9 | -1.9 | -3.44 |
| (3,4) | | -2.71 | -101.7 | -4.1 |
| (4,1) | | | | |
| (4,2) | -1.9 | | 0.0 | -1.9 |
| (4,3) | -2.71 | | -0.99 | -101.7 |
| (4,4) | | | -1.9 | -3.44 |

Zwischenfazit

- **Wiederholtes Ausführen möglichst vieler Zustands-Aktions-Paare notwendig**
 - Rückwärtige Propagation geht nur langsam voran
- **Tabellenbasiertes Q-Learning stößt bei großem Zustandsraum schnell an seine Grenzen**
 - Value muss für jedes Zustands-Aktions-Paar einzeln gelernt werden
 - Einsatz eines Funktionsapproximators um Erfahrung zu generalisieren

Agenda

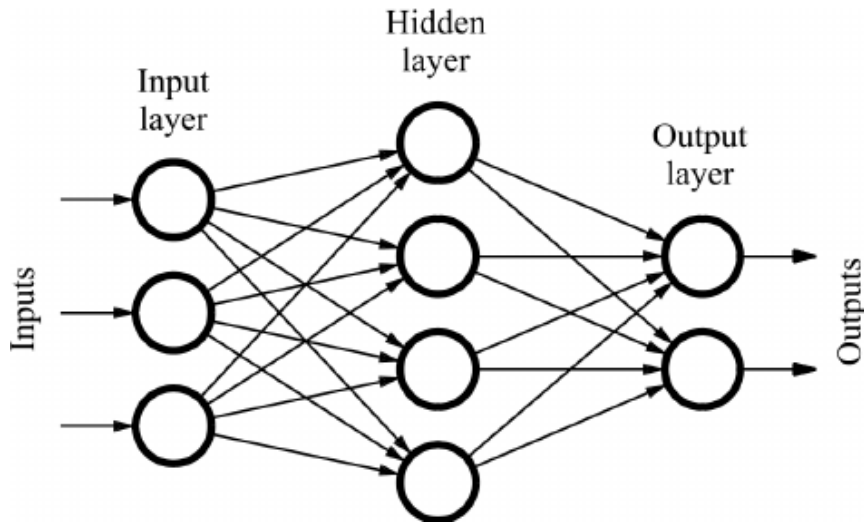
Maschinelles Lernen

Q-Learning

**Exkurs: Q-Learning
mit Neuronalen Netzen**

**Q-Learning mit
Anki Overdrive**

Exkurs: Q-Learning mit Neuronalen Netzen



Quelle: www.researchgate.net

- Gewichte an allen Kanten
- Gewichtete Summe der eingehenden Signale
- Berechnung der Aktivierungsfunktion
- Lernen durch Backpropagation
- Zum Beispiel für Bild-Klassifizierung genutzt

Exkurs: Q-Learning mit Neuronalen Netzen

- **Einsatz eines Neuronalen Netzes als Funktionsapproximator anstatt einer Tabelle**
 - Generalisierender Effekt reduziert die Anzahl der benötigten Zustands-Aktions-Paare
- **Einsatz von Experience Replay**
 - Netz wird mit zufälligen Samples aus dem Speicher gefüttert.
- **Weitere Methoden notwendig, damit das Netz konvergiert**

Exkurs: Q-Learning mit Neuronalen Netzen

- **Convolutional Neural Networks**
 - Arbeiten direkt auf Pixeldaten
 - Extrahieren High Level Features wie Kanten
- **Prioritybased Experience Replay**
 - Bevorzuge Samples mit höherer Wahrscheinlichkeit, aus denen man am meisten Lernen kann
- **Double Q-Learning**
 - Separate Q-Funktionen für Update und Maximumberechnung

Agenda

Maschinelles Lernen

Q-Learning

**Exkurs: Q-Learning
mit Neuronalen Netzen**

**Q-Learning mit
Anki Overdrive**