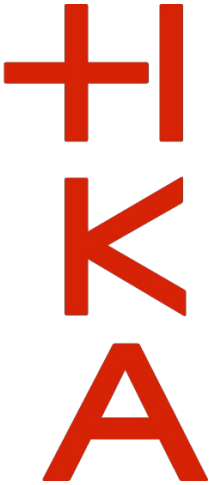
inovex



KI Labor - Wintersemester 2022

Reinforcement **L**earning
Sprintwechsel & Vorstellung Assignment

Jochen Gietzen, Stefan Käser, Maximilian Blanck, Pascal Fecht, **Adrian Westermeier, Tim Bossenmaier**

Karlsruhe, 16. Dezember 2022

Schedule

Datum	Thema	Inhalt	Präsenz
30. Sept.	Allg.	Organisation, Teamfindung, Vorstellung CV	Ja
7. Okt.	Ausfall (DMA Techday)		
14. Okt.	CV	Q&A Sessions	Nein
21. Okt.	CV	Sprintwechsel, Vorstellung Assignment	Ja
28. Okt.	CV	Q&A Sessions	Nein
4. Nov.	CV / NLP	Abgabe CV, Vorstellung NLP	Ja
11. Nov.	NLP	Q&A Sessions	Nein
18. Nov.	NLP	Sprintwechsel, Vorstellung Assignment	Ja
25. Nov.	NLP	Q&A Sessions	Nein
2. Dez.	Ausfall (Winter Plenum)		
9. Dez.	**NLP / RL**	**Abgabe NLP, Vorstellung RL**	**Ja**
16. Dez.	RL	Sprintwechsel, Vorstellung Assignment	Nein
23. Dez.	RL	Q&A Sessions (optional) ?	Nein
13. Jan.	**RL**	**Abgabe RL, Abschluss KI Labor**	**Ja**
20. Jan.	**-**	**kein Termin**	**-**

Agenda

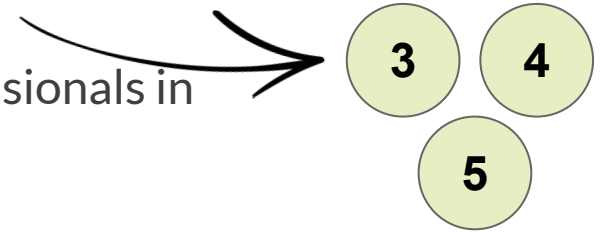
- **Theorie**
  - Deep Q-Network
  - Experience Replay
  - Target Model
  - Vorverarbeitung für Pixel-basierte Atari Games (Framestacking, etc.)

- **Praxis**
  - CartPole Gym mit Deep Q-Learning (Aufgabe 3)
  - Pong (Pixel-basiert) mit Deep Q-Learning (Aufgabe 4)
  - Assignment

Reinforcement Learning

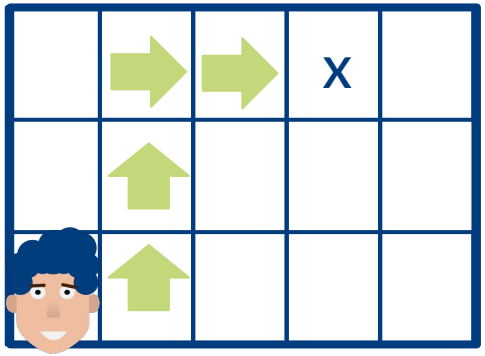
Meilensteine im Reinforcement Learning

1950’s	Optimal control, dynamic programming
1970-80’s	Temporal difference learning
1989	Q-Learning
1992	TD Gammon
2013	Playing atari with deep reinforcement learning
2016	AlphaGo vs. Lee Sedol
2019	AlphaStar & OpenAI Five beat professionals in Starcraft and Dota 2

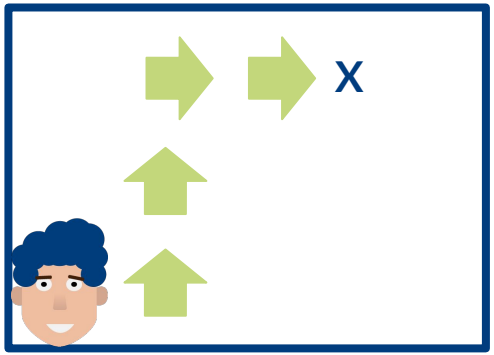


inovex

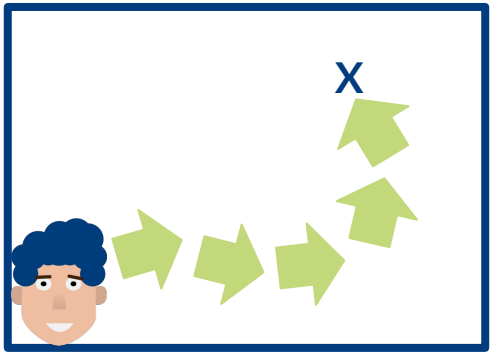
Recap



Zustände & Aktionen diskret



Zustände kontinuierlich & Aktionen diskret

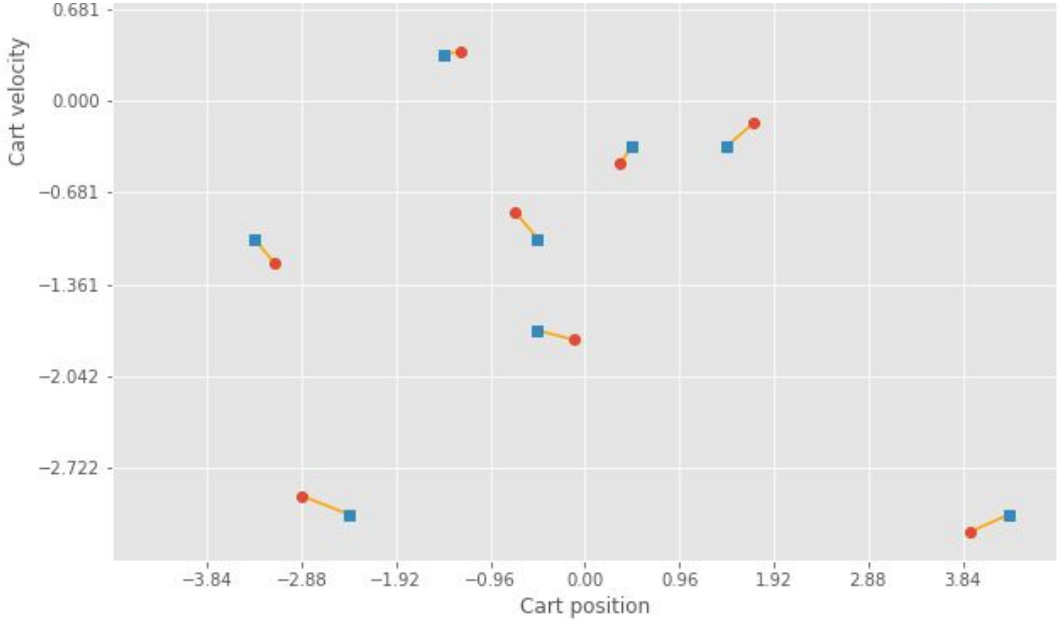


Zustände & Aktionen kontinuierlich

inovex

Was bisher geschah ...

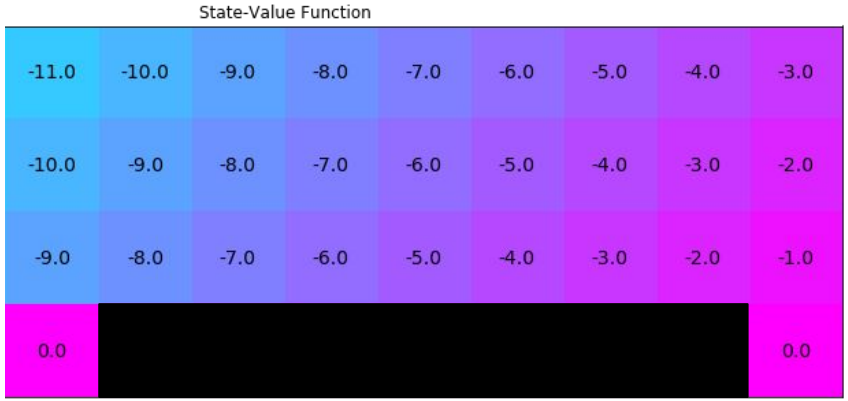
Diskretisierung

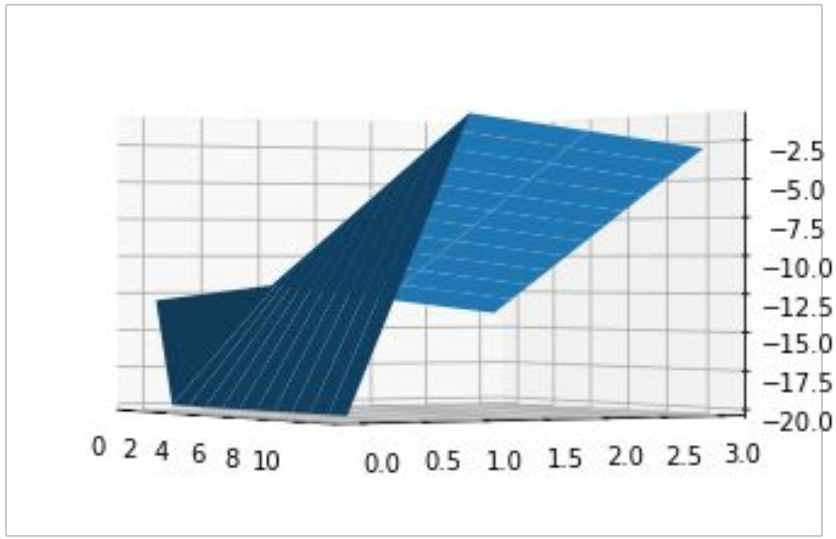


inovex

Funktionsapproximation

Beispiel Cliffwalking





$$\hat{v}(s,\theta) \approx v_\pi(s), \theta \in \mathbb{R}^d$$

Approximation — Value-Function — Gewichtsvektor

inovex

Funktionsapproximation

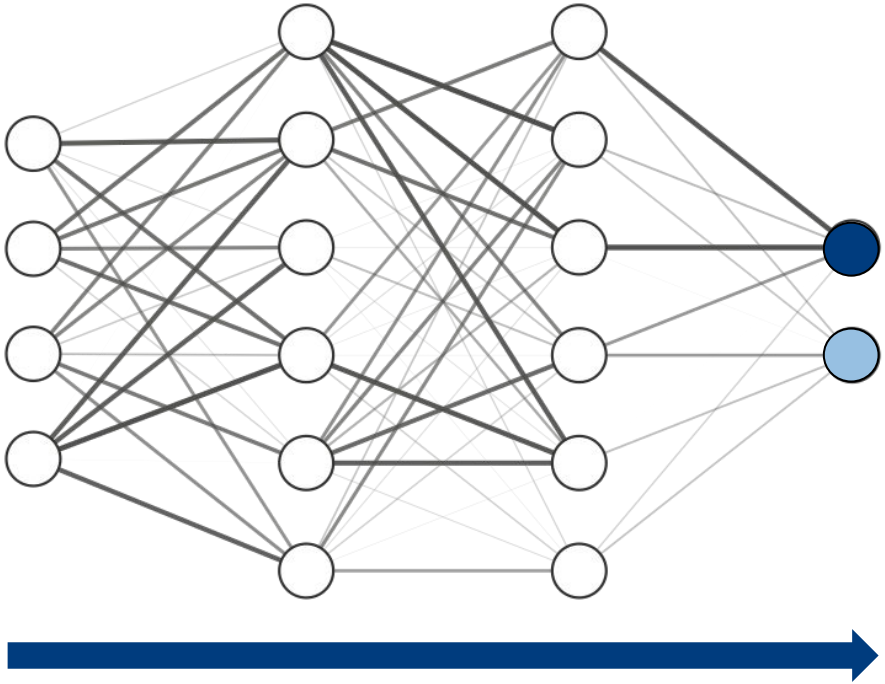
Beispiel CartPole: **Forward-Pass**

Cart Position

Cart Velocity

Pole Angle

Pole Velocity at Tip



Q(➡)

Q(⬅)

inovex

Funktionsapproximation

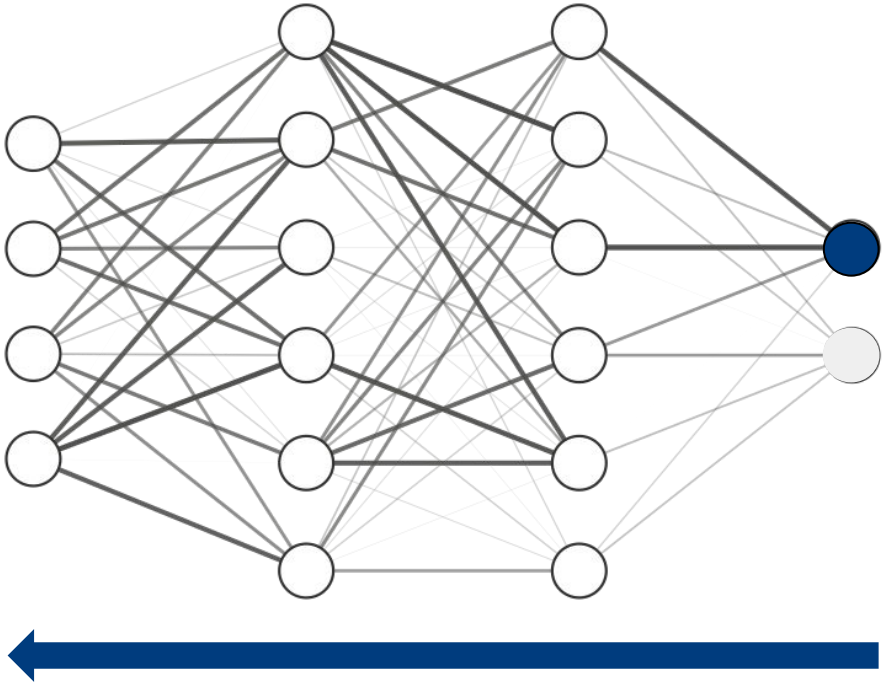
Beispiel CartPole: **Backward-Pass**

Cart Position

Cart Velocity

Pole Angle

Pole Velocity at Tip

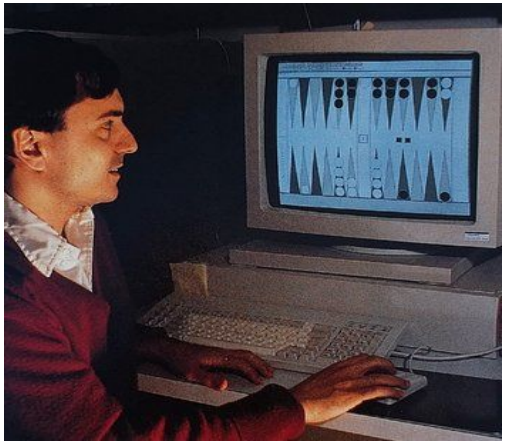


Action: right selected

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

inovex

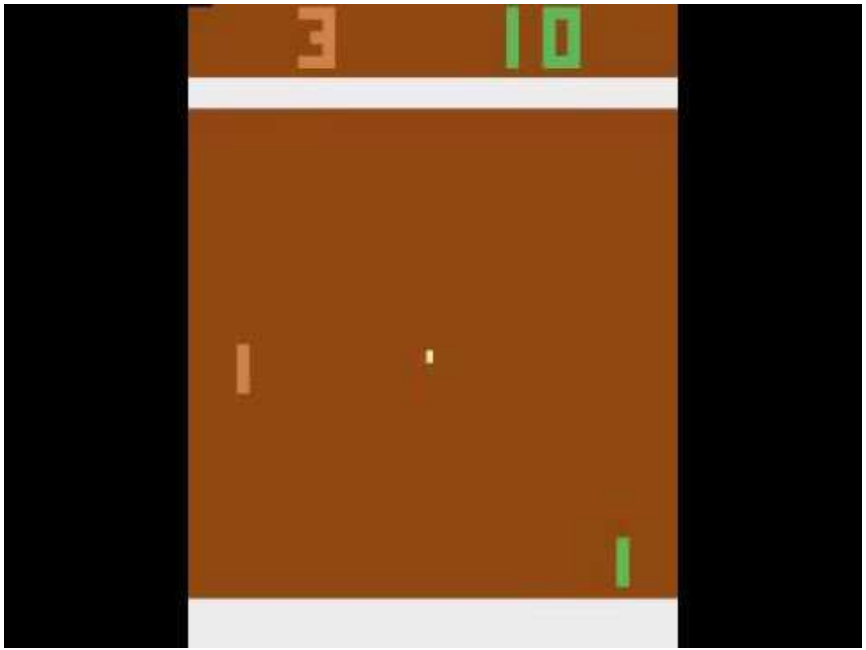
RL mit Neuronalen Netzen



- TD-Gammon (1995)
  - erste erfolgreiche Anwendung von NNs in RL
  - ähnliche Ansätze für andere Spiele
    nicht erfolgreich
  - etwa 20 Jahre keine Fortschritte in diesem Bereich
- Ansatz erfährt durch Deep Learning und Fortschritte in Computer Vision wieder Aufmerksamkeit
  - RL kann Ende-zu-Ende gelernt werden (direkter Input von Bildern anstelle von erstellten Features) -> großer Fortschritt

Bildquelle: IBM Research - TD Gammon (https://researcher.watson.ibm.com/researcher/view_page.php?id=6853)

Atari 2600 Spiele

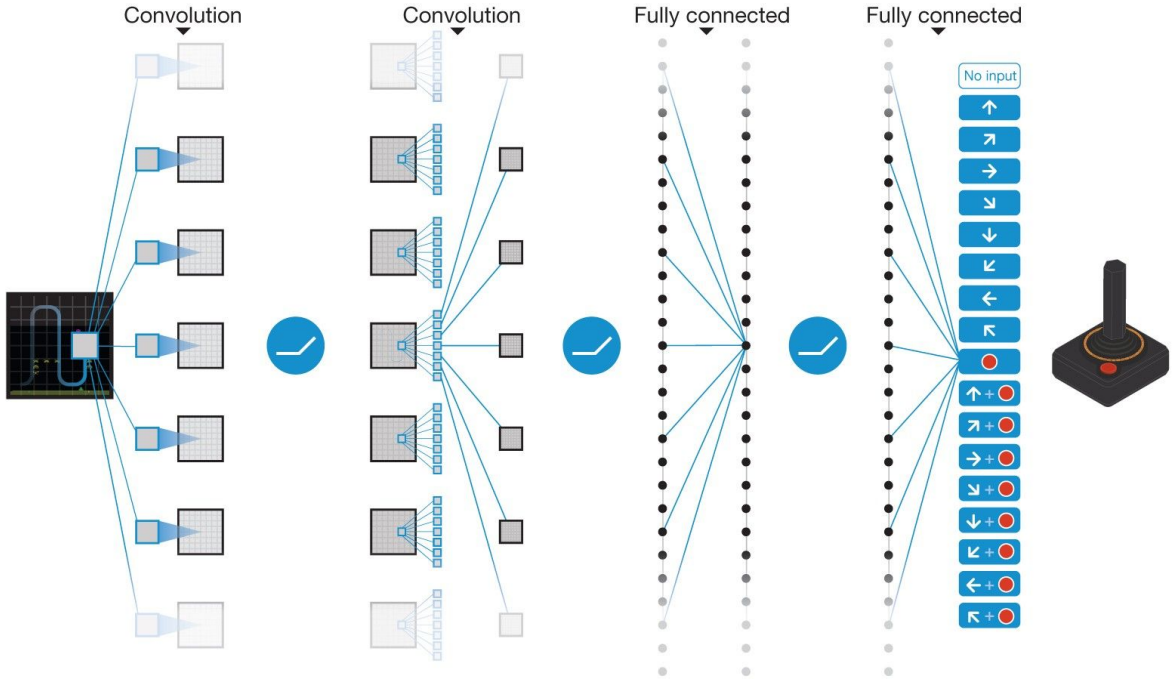


inovex

Deep Q-Network

Pixel-basierte Atari Games

Input:
84 x 84 x n



Mnih et al. (2015) - Human-level control through Deep Reinforcement Learning

Funktionsapproximation

Deep Q-Network

- Approximation der Q-Funktion mit NN
  - Optimierung mit stochastic gradient descent (SGD)
  - Minimierung des Abstands zwischen Schätzer und Target

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim U(D)}[(y_i - \hat{q}(s,a,\theta_i))^2]$$

Loss L_i **Weights** θ_i **in Iteration i**

Experience Replay $(s,a,r,s')\sim U(D)$

Target Q-Value y_i

approximierter Q-Value $\hat{q}(s,a,\theta_i)$

Mnih et al. (2015) - Human-level Control through Deep Reinforcement Learning

Target Network

- Kopie der eigentlichen Architektur mit fixen Gewichten
- Update alle c Iterationen

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim U(D)}[(y_i - \hat{q}(s, a, \theta_i))^2]$$

Q-Network

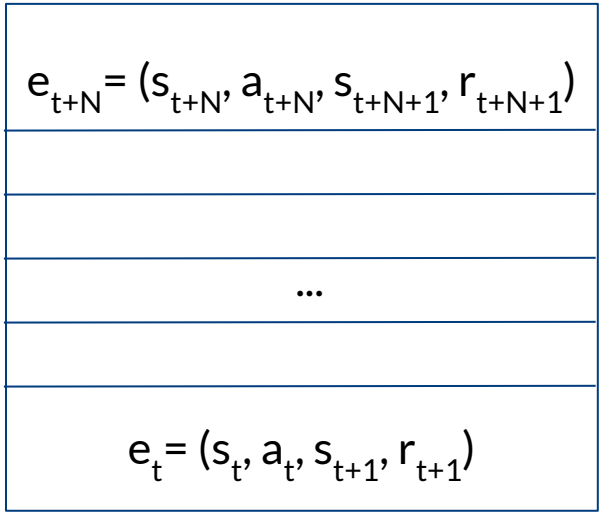
$$r + \gamma \max_{a'} Q(s', a', \theta_i^-)$$

Target Network

Mnih et al. (2015) - Human-level Control through Deep Reinforcement Learning

Experience Replay

- Problem
  - Starke Korrelation der States erschwert das Lernen
- Lösung
  - Letzte *N* Experiences werden in **Replay Memory** gespeichert
  - Random Uniform Sampling



Replay Memory

Mnih et al. (2013) - Playing Atari with Deep Reinforcement Learning

Deep Q-Learning

Algorithmus

Algorithm 1 Deep Q-learning with Experience Replay

Initialize replay memory $\mathcal{D}$ to capacity N
Initialize action-value function Q with random weights
for episode $= 1, M$ **do**
  Initialise sequence $s_1 = \{x_1\}$ and preprocessed sequenced $\phi_1 = \phi(s_1)$
  for $t = 1, T$ **do**
    With probability ϵ select a random action a_t
    otherwise select $a_t = \max_a Q^*(\phi(s_t), a; \theta)$
    Execute action a_t in emulator and observe reward r_t and image x_{t+1}
    Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
    Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in $\mathcal{D}$
    Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from $\mathcal{D}$
    Set $y_j = \begin{cases} r_j & \text{for terminal } \phi_{j+1} \\ r_j + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$
    Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ according to equation 3
  end for
end for

Mnih et al. (2013) - Playing Atari with Deep Reinforcement Learning

Beispiele für Preprocessing

Warp Frame

- Konvertierung der Frames in Graustufen
- Downsampling / Cropping

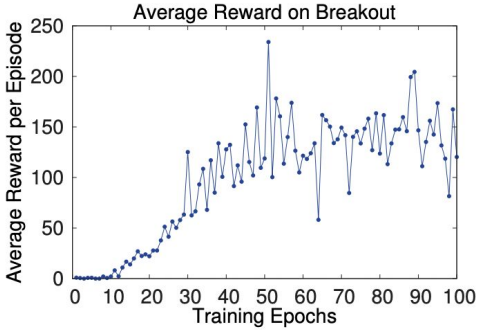
Framestakking

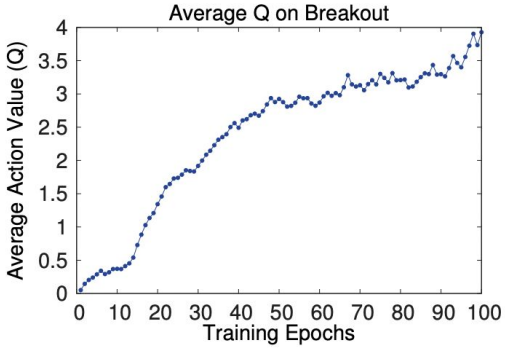
- mehrere aufeinanderfolgende Frames als Eingabe, um Bewegung nachvollziehen zu können und für kürzeres Training
- Ausführung der gewählten Aktion für gesamten Framestack

Mnih et al. (2013) - Playing Atari with Deep Reinforcement Learning

Evaluierung von Deep RL

- Im Vergleich zu Supervised Learning deutlich herausfordernder (kein Vergleich von training und validation möglich)
- Zwei grundlegende Metriken:
  - **Average Reward:** mittlerer, erzielter Reward je Episode
  - **Average Q-Value:** mittlerer Q-Wert für eine vor dem Training zufällig gewählte Menge an Zuständen je Episode

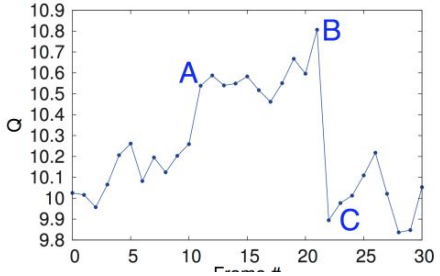




Mnih et al. (2013) - Playing Atari with Deep Reinforcement Learning

Deep Q-Network

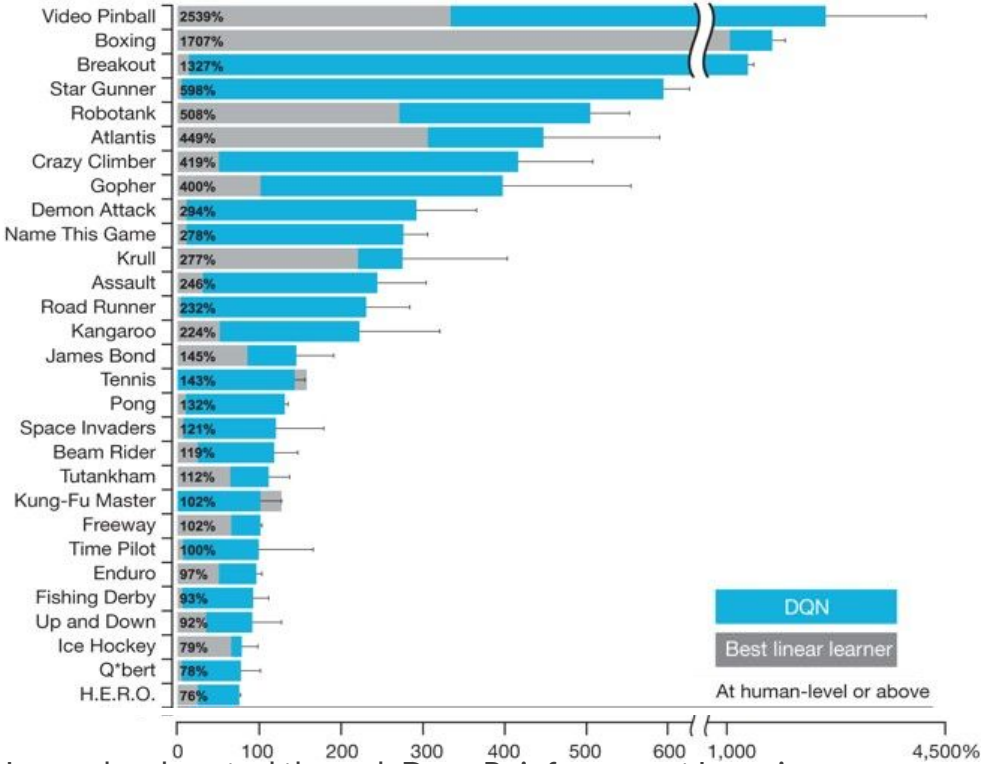
Beispiel für eine gelernte Q-Funktion



Mnih et al. (2015) - Human-level Control through Deep Reinforcement Learning

Deep Q-Network

Pixel-basierte Atari Games



Mnih et al. (2015) - Human-level control through Deep Reinforcement Learning

inovex

Aufgaben

Aufgaben

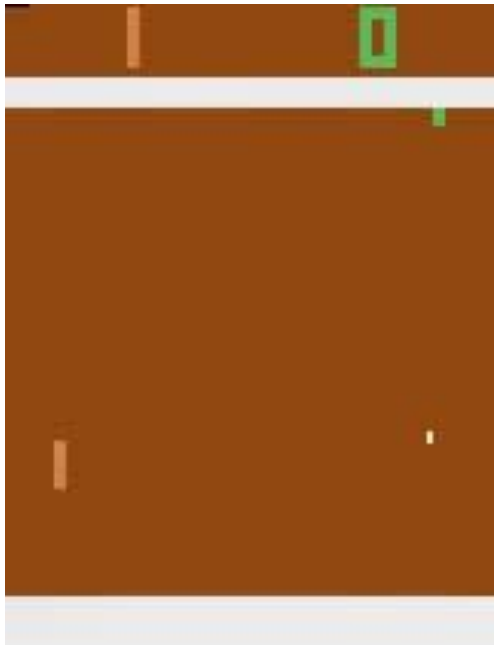
- Aufgabe 3: CartPole Gym mit Deep Q-Learning (freiwillig)
- Aufgabe 4: Pong mit Deep Q-Learning (freiwillig)
- Aufgabe 5: Assignment (Bewertungsgrundlage)

Aufgabe 3: CartPole Gym mit Deep Q-Learning

- **Freiwillige Bearbeitung** als Vorbereitung auf Assignment
- Jupyter Lab Notebook

Aufgabe 4: Pong mit Deep Q-Learning

- **Freiwillige Bearbeitung** als Vorbereitung auf Assignment
- Jupyter Lab Notebook



Aufgabe 5: Assignment

- **Assignment dient als Bewertungsgrundlage = Pflicht**
- ein Atari Spiele zur freien Auswahl
- Jupyter Lab Notebook
- Freie Wahl des Ansatzes

inovex

Als kleine Motivation

Infomaterial

- Kostenlose "Standard"-Lektüre für den Einstieg in RL:
  Reinforcement Learning: An Introduction (Sutton and Barto), siehe
  http://incompleteideas.net/book/RLbook2018.pdf
- Ausführlich und gut erklärter Einstieg in RL (Video-Lektionen):
  UCL Course on RL (David Silver, Google DeepMind), siehe
  https://www.davidsilver.uk/teaching/
- *Algorithms in Reinforcement Learning* von Csaba Szepesvári, siehe
  https://sites.ualberta.ca/~szepesva/papers/RLAlgsInMDPs.pdf
- Blog mit Videos zum Einstieg in RL und Q-Learning, DQN und vieles mehr:
  Reinforcement Learning – Introducing Goal Oriented Intelligence, siehe
  https://deeplizard.com/learn/video/nyjbcRQ-uQ8
- *David Silver - AlphaGo, AlphaZero, and Deep Reinforcement Learning*, siehe Lex Fridman Podcast #86 https://lexfridman.com/david-silver/

Vielen Dank

Adrian Westermeier
adrian.westermeier@inovex.de

Tim Bossenmaier
tim.bossenmaier@inovex.de