


# Deep Q-Learning

Janot George, Maurice Borries

Hochschule für Technik und Wirtschaft Berlin  
Studiengang Wirtschaftsmathematik (Bachelor)  
Seminar

02.02.2022

# Inhaltsübersicht

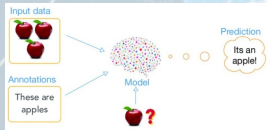
- 
- 1 Konzept
  - 2 Q-Learning Bestandteile
  - 3 Erweiterung zum Deep-Q-Learning
  - 4 Abbildungs- und Quellenverzeichnis

# Inhaltsübersicht

- 
- 1 Konzept
  - 2 Q-Learning Bestandteile
  - 3 Erweiterung zum Deep-Q-Learning
  - 4 Abbildungs- und Quellenverzeichnis

# Arten des Maschinen Learnings

## Supervised



Vorhersagen treffen

## Unsupervised



Muster entdecken

## Reinforcement



Aus Fehlern selbstständig lernen

# Was ist (Deep) Q-Learning



- Algorithmus trainiert eine KI
- Zuckerbrot-und-Peitschen-Prinzip
- KI startet ohne Daten
- Konzept: Daten erzeugen und Funktionen live optimieren
- Ziel: Annäherung an optimale Lösung des Problems

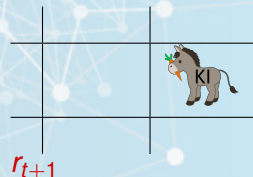
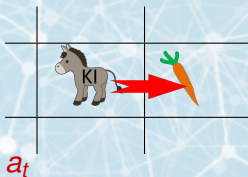
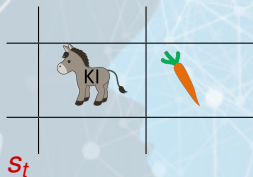
# Inhaltsübersicht

- 
- 1 Konzept
  - 2 Q-Learning Bestandteile**
  - 3 Erweiterung zum Deep-Q-Learning
  - 4 Abbildungs- und Quellenverzeichnis

# Markov-Entscheidungsprozess

## Definitionen:

- Zustand der Umgebung zum Zeitpunkt  $t$ :  $s_t$
- Aktion der KI (des Agenten) zum Zeitpunkt  $t$ :  $a_t$
- Belohnung der Aktion  $a_t$ :  $r_{t+1}$



# Exploration versus Exploitation

## Exploration

- Erkundung des Umfeldes
- zufällige Aktionen ausführen
- neue Daten sammeln

## Exploitation

- Ausnutzung gesammelter Daten
- Daten überprüfen
- Daten aktualisieren

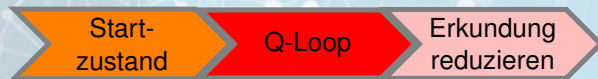
Optimale Anwendung der Erkundungsrate  $\epsilon$  erfordert:

- $\epsilon_{max}$  : Erkundungsrate zu Beginn
- $\epsilon_{red}$  : Reduktionswert der Erkundungsrate
- $\epsilon_{min}$  : Minimale Erkundungsrate

$$\epsilon = \epsilon_{min} + \frac{\epsilon_{max} - \epsilon_{min}}{e^{\epsilon_{red} \cdot \text{Episodennummer}}}$$



# Q-Learning Episode



Q-Loop:



# Q-Funktionen

Aktionsbewertungsfunktion: (Diskontierungsfaktor  $\gamma$ )

$$Q(s_t, a_t) = \mathbb{E} \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \right]$$

Bellman-Gleichung:

$$Q^*(s_t, a_t) = \mathbb{E} \left[ r_{t+1} + \gamma \max_{a'} [Q_T(s_{t+1}, a')] \right]$$











Loss-Funktion:

$$Loss = Q^*(s_t, a_t) - Q_T(s_t, a_t)$$

Annäherung mit Lernrate  $\alpha$ :

$$Q_{T(neu)}(s_t, a_t) = (1 - \alpha) \cdot Q_T(s_t, a_t) + \alpha \left[ r_{t+1} + \gamma \max_{a'} [Q_T(s_{t+1}, a')] \right]$$

# Q-Tabelle

						
	0,23	0,34	1	-0,24	-0,39	0
	-0,52	-0,49	-0,32	-0,05	-0,03	0
	0,24	0,45	-0,38	0,55	1	0
	-0,59	-0,15	-0,15	-0,34	-0,03	0

# Beispiele



# Grenzen des Q-learning

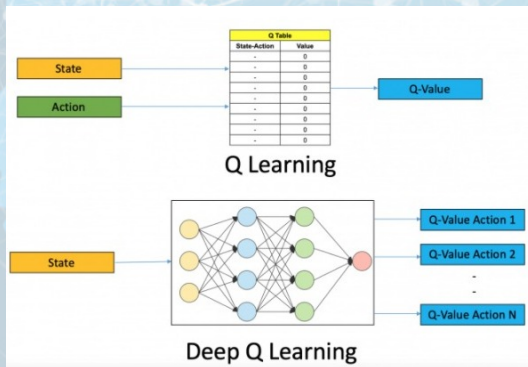
- Komplexität des Problems
- Trainingsgeschwindigkeit
- Größe der Q-Tabelle

# Inhaltsübersicht

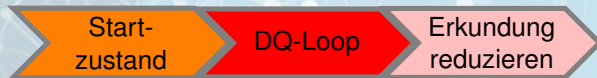
- 
- 1 Konzept
  - 2 Q-Learning Bestandteile
  - 3 Erweiterung zum Deep-Q-Learning**
  - 4 Abbildungs- und Quellenverzeichnis

# Änderungen

- Q-Tabelle durch neuronales Netzwerk
- Ergänze Erinnerungsspeicher



# Deep-Q-Learning Episode



DQ-Loop:



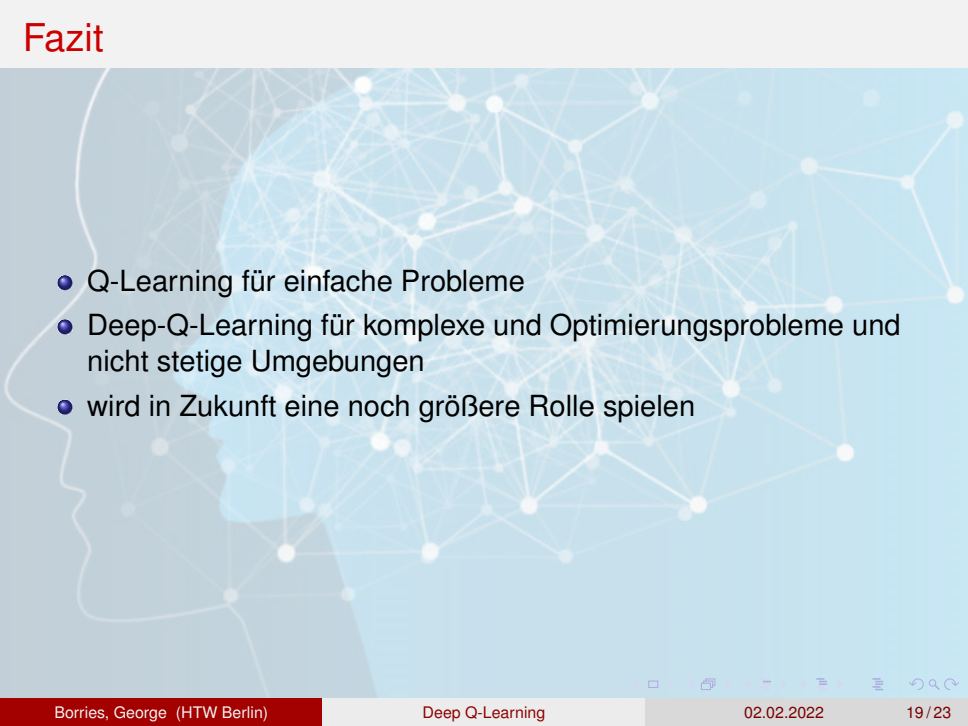


# Deep-Q-Learning Arten

	Deep-Q-Learning (DQL)	Double Deep-Q-Learning (DDQL)	Clipped Deep-Q-Learning (CDQL)
Strategie-Netzwerk	Exploitation $Q(s_t, a_t)$ $\max [Q(s_{t+1}, a')]$	Exploitation $Q(s_t, a_t)$	Exploitation $Q(s_t, a_t)$ $\max [Q(s_{t+1}, a')]$
Target-Netzwerk		$\max [Q(s_{t+1}, a')]$	$\max [Q(s_{t+1}, a')]$

# Beispiele



- 
- Q-Learning für einfache Probleme
  - Deep-Q-Learning für komplexe und Optimierungsprobleme und nicht stetige Umgebungen
  - wird in Zukunft eine noch größere Rolle spielen

# Inhaltsübersicht

- 
- 1 Konzept
  - 2 Q-Learning Bestandteile
  - 3 Erweiterung zum Deep-Q-Learning
  - 4 Abbildungs- und Quellenverzeichnis**

# Abbildungsverzeichnis

- Algorithmus: <https://www.allaboutlean.com/employee-motivation-1/carrot-and-stick/>; Zuletzt aufgerufen: 23.11.2021
- KI (Esel): <https://de.cleanpng.com/png-b1ndvx/>; Zuletzt aufgerufen: 22.11.2021
- Hintergrund: <https://it-talents.de/it-wissen/machine-learning-accuracy-und-precision/>; Zuletzt aufgerufen: 7.11.2021
- reinforcement <https://www.kdnuggets.com/2019/10/mathworks-reinforcement-learning.html>; Zuletzt aufgerufen: 30.1.2022
- unsupervised [https://www.researchgate.net/figure/Supervised-learning-and-unsupervised-learning-Supervised-learning-uses-annotation\\_fig1\\_329533120](https://www.researchgate.net/figure/Supervised-learning-and-unsupervised-learning-Supervised-learning-uses-annotation_fig1_329533120) ; Zuletzt aufgerufen: 30.1.2022
- supervised <https://www.analyticsvidhya.com/blog/2019/04/introduction-deep-q-learning-python/> ; Zuletzt aufgerufen: 30.1.2022

# Quellenverzeichnis

- Richter, S. (2019) Statistisches und maschinelles Lernen, Berlin, Springer.
- K.-L. Du and M. N. S. Swamy, Neural Networks and Statistical Learning.
- Ilyas, Agakishiev: METIS: Reinforcement Learning, Humboldt-Universität zu Berlin, [IRTG1792.HU-Berlin.de](https://www.irtg1792.hu-berlin.de)
- <https://www.learndatasci.com/tutorials/reinforcement-q-learning-scratch-python-openai-gym/> , Satwik Kansal, Brendan Martin, zuletzt abgerufen: 7.12.2021
- <https://ichi.pro/de/einfuehrung-in-das-reinforcement-learning-markov-entscheidungsprozess-75613770348762> , ICHI.PRO, Laurenz Wuttke, zuletzt abgerufen: 7.12.2021
- <https://hci.iwr.uni-heidelberg.de/system/files/private/downloads/541645681/dammann-reinforcement-learning-report.pdf> , Patrick Dammann, zuletzt abgerufen: 7.12.2021

# Quellenverzeichnis

- <http://www.informatik.uni-ulm.de/ni/Lehre/SS05/RL/vorlesung/rl03.pdf> , F. Schwenker, zuletzt abgerufen: 7.12.2021
- <https://deeplizard.com/> , Chris and Mandy, zuletzt abgerufen: 7.12.2021
- <https://datasolut.com/neuronale-netzwerke-einfuehrung/> ,Laurenz Wuttke, zuletzt abgerufen: 7.12.2021
- <https://www.samyzaf.com/ML/rl/qmaze.html> , zuletzt abgerufen: 7.12.2021
- [https://www-aitude-com.translate.goog/supervised-vs-unsupervised-vs-reinforcement/?\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=de&\\_x\\_tr\\_hl=de&\\_x\\_tr\\_pto=op,sc](https://www-aitude-com.translate.goog/supervised-vs-unsupervised-vs-reinforcement/?_x_tr_sl=en&_x_tr_tl=de&_x_tr_hl=de&_x_tr_pto=op,sc) ; zuletzt abgerufen: 27.1.2022
- <https://www.trendreport.de/anwendung-des-machine-learning-bei-der-analyse-von-kapitalmaerkten/> , Tobias Waning, Alexander Brun, Hendrik von der Haar, zuletzt abgerufen: 27.1.2022