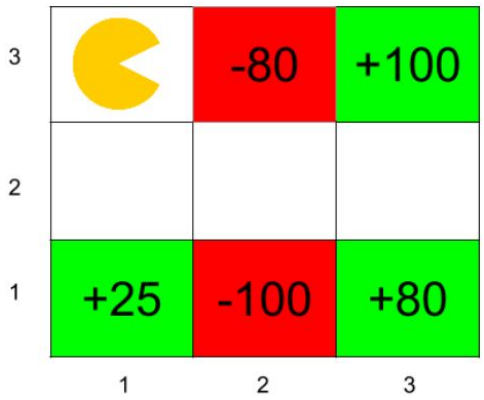
Übungsblatt

Grundlagen der Künstlichen Intelligenz

01.12.2023, DHBW Lörrach

- Reinforcement Learning -

Betrachten Sie Pacman, der gerade versucht, die optimale Strategie π^* im dargestellten $3x3$-Gitter zu erlernen. Wenn eine Aktion dazu führt, dass Pacman auf einen der schattierten Flächen landet, erhält er die angezeigte Belohnung. Alle schattierten Zustände sind absorbierend, d.h. der Markov-Entscheidungsprozess terminiert, sobald Pacman sie erreicht. Für die anderen Zustände sind die Aktionen North (N), East (E), South (S) oder West (W) verfügbar, die Pacman deterministisch in den entsprechenden Nachbarzustand bewegen. Nehmen Sie einen Diskontierungsfaktor $\gamma = 0.5$ an. Pacman startet in Zustand (1,3).



Pacman-Welt

a) Berechnen Sie den Wert der optimalen Wertefunktion V^{π^*} für folgende Zustände:

$V^{\pi^*}(3,2) =$ _____ $V^{\pi^*}(2,2) =$ _____ $V^{\pi^*}(1,3) =$ _____

Hinweis: Verwenden Sie das Optimalitätskriterium $V^{\pi^*}(s) \geq V^{\pi}(s)$ aus der Vorlesung.

b) Der Agent startet von der oberen linken Ecke und durchläuft folgende Episoden der Pacman-Welt. Jeder Eintrag einer Episode stellt ein Tuple mit (s, a, s', r) dar.

Episode 1	Episode 2	Episode 3
(1,3), S, (1,2), 0	(1,3), S, (1,2), 0	(1,3), S, (1,2), 0
(1,2), E, (2,2), 0	(1,2), E, (2,2), 0	(1,2), E, (2,2), 0
(2,2), S, (2,1), -100	(2,2), E, (3,2), 0	(2,2), E, (3,2), 0
	(3,2), N, (3,3), +100	(3,2), S, (3,1), +80

Berechnen Sie die Q-Werte mit der Iterationsvorschrift für Q-Lernen aus der Vorlesung:

$Q((3,2),N) =$ _____ $Q((1,2),N) =$ _____ $Q((2,2),N) =$ _____