Einführung in die Künstliche Intelligenz

Gedächtnisprotokoll 2022 - 30. August

Aufgabe 1:

a) Geben Sie das Bayes-Theorem an
b) Geben Sie P(X | Y) mit Hilfe von P(X, Y, Z) an
c) Sei A(D) eine Funktion, welche den Durchschnitt μ zurückgibt,
z.B.: Für D = 1,2,3,4,5 = A(D) = 3.
Erklären Sie, wie Sie mit Hilfe von Bootstrapping Samples erstellen können, und geben Sie Beispiele für Samples an, und damit die Varianz schätzen können und geben Sie eine geeignete Formel für die Berechnung der Varianz an.

Aufgabe 2:

Erklären Sie die jeweiligen Branchings von:

a) 1 Spieler Spiel, welches deterministisch ist
b) MDP Spiel
c) POMDP
d) 2 Spieler Spiel, welches deterministisch ist, aber man beobachtet weder den Initialzustand noch andere Folgezustände, dafür aber die Aktionen des jeweils anderen Spielers

Aufgabe 3:

Berechnen Sie von den drei Banditen den jeweiligen UCB1 Wert, wobei $ln = log_2$.

1. Bandit: -1, 0, 2, 3 => μ = 1
2. Bandit: -1, 1 => μ = 0
3. Bandit 0, 2 => μ = 1

(Die μ's waren nicht gegeben, dennoch füge ich sie ein, da dies die Werte waren, die man eh mit berechnen musste)
(Ja, es gab keine Frage zu "welchen Banditen sollte man wählen, wenn …")

Aufgabe 4:

2 Spieler Spiel, 0 oder 1 wählbar pro Spieler, Spieler sehen auch die Aktionen des jeweils anderen.

1. Rollout Policy nimmt immer die 0.
2. Man wählt nach Expansion immer zuerst die 0.

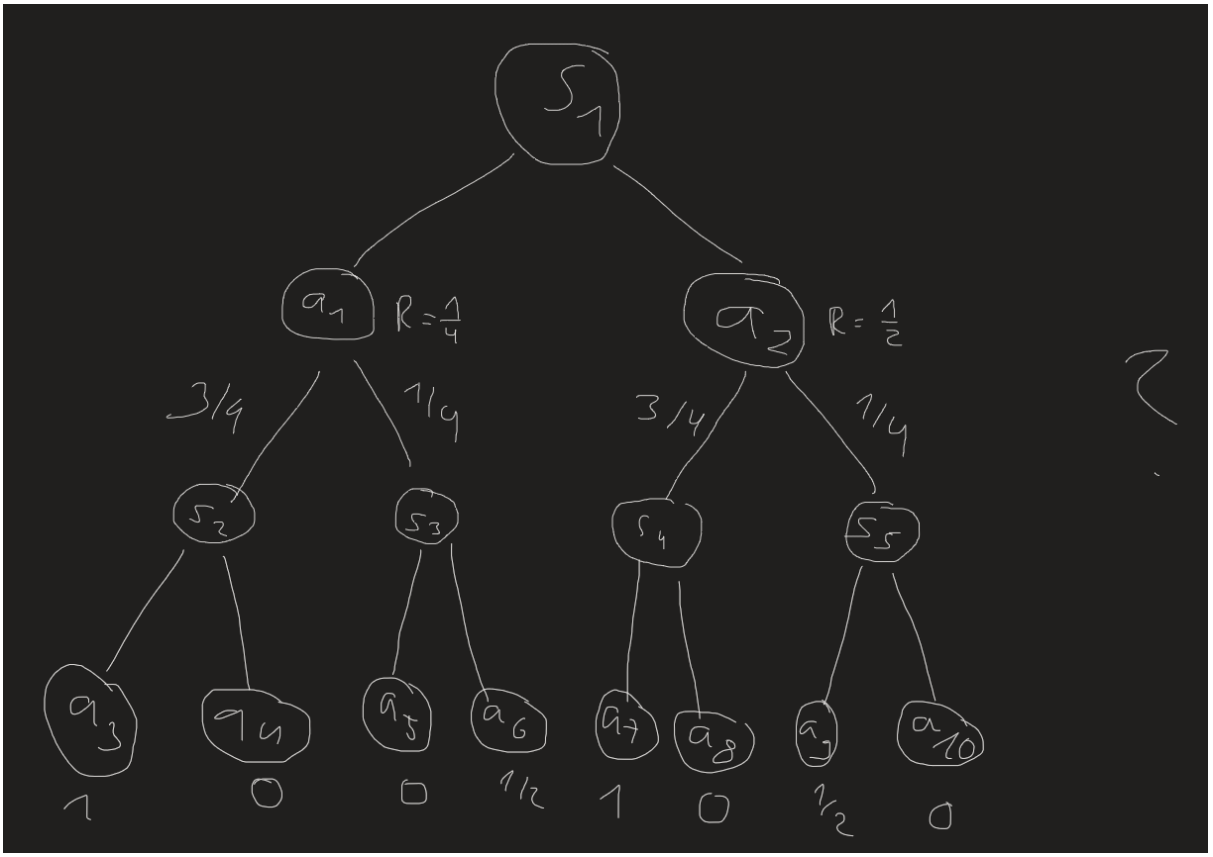
4 Iterationen zu diesem MCTS.
Aktionen a = [0, 1].
Wenn nach der zweiten Runde die Summe der gewählten Aktionen gerade ist, hat man einen Reward von -1, sonst bei ungerade einen Reward 1.

Standard MCTS nach den 4 Prinzipien: Selection, Expansion, Rollout und Backpropagation.

Jeweils relevant sind die UCB1 Werte ab der 2. Iteration des Algorithmus'
Die Werte von Q_0 sind in jeder Iteration auch gefragt.

Aufgabe 5:

1. Q-Iterations Formel angeben mit immediate Reward und Übergangswahrscheinlichkeiten P(s' | s, a).
2. Berechnen Sie die 2 Q-Iterationswerte für Q(s_1, a_1) und Q(s_1, a_2)



Aufgabe 6

TD-Learning und Q-Learning

4 States s = [N, S, C, X], wobei X ein terminal state ist.
3 Aktionen a = [Unten, Oben, Exit]

Initial mit 0, sonst machen wir die folgenden Beobachtungen:

State s	Aktion a	Reward r	Folgestate s’
C	Unten	0	S
S	Exit	24	X
C	Oben	0	N
N	Exit	28	X
C	Unten	0	S
S	Exit	24	X
C	Oben	0	N
N	Exit	28	X
…	…	…	…

a) Unter den ersten Beobachtungen (Tabelle) was sind die Werte V(S), V(C) und V(N)

b) Gegen was wird V(C) konvergieren für eine passend kleine Lernrate α.
c) Gegen was wird Q(C, Unten) konvergieren für eine passend kleine Lernrate α