Universität Bielefeld

Technische Fakultät
AG Neuroinformatik
Robert Haschke

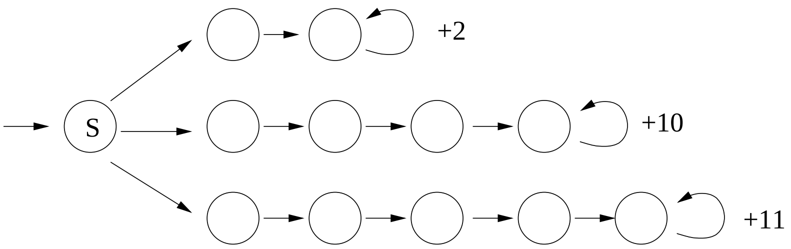
Sekretariat: P. Udelhoven
Raum M7-113
Tel. 106-6891

Übungen zu "Robotik II"

WS 2003/04 Blatt 10

Ausgabe: 8.1.2004, **Abgabe:** 15.1.2004

Aufgabe 10.1, Kumulierte Rewards: Betrachten Sie den im folgenden Graphen dargestellten deterministischen Markovschen Entscheidungsprozess.



Im Startzustand S kann der Agent drei Aktionen wählen, die ihn in die drei unterschiedlichen Zweige des Graphen führen, ansonsten ist immer nur eine Aktion (angedeutet durch die Pfeile) möglich. Erst in den Endzuständen des Graphen bekommt der Agent einen positiven Reward.

Für den Agenten gibt es drei mögliche policies – nämlich die Auswahl eines der drei Zweige. Berechnen Sie für jede dieser drei policies den kumulierten Reward unter Benutzung der verschiedenen Reward-Funktionen:

- discounted reward mit Zeithorizont $h = 5$
- finite-horizon reward mit $\gamma = 0.9$
- average reward

Nutzen Sie die Identität $\sum_{t=0}^{\infty} \gamma^t = \frac{1}{1-y}$.
Für welchen Zweig würde sich der Agent aufgrund dieser Werte jeweils entscheiden?

Aufgabe 10.2, : Zeigen Sie, dass der Absolutwert eines rewards r_t nicht entscheidend ist, sondern es nur auf die Differenz zwischen positivem und negativem Feedback ankommt. Beweisen Sie dazu, dass die Addition einer Konstante C zu allen rewards r_t lediglich einen konstanten Offset zu allen Werten $V^{\pi}(s)$ hinzufügt.