

Reinforcement-Lernen

Wintersemester 10/11

Aufgabenzettel 4

Abgabe: 14.01.2011 **vor** der Vorlesung.

Aufgabe 7 (20%). Funktionsapproximatoren:

Gib ein Beispiel für einen *nicht-linearen* Funktionsapproximator an, der für eine Wertefunktion im Reinforcement Learning verwendet werden kann. Gib auch die Ableitung nach den Parametern des Funktionsapproximators an.

Aufgabe 8 (50%). Eligibility Traces:

1. Gib einen Pseudo-Code für SARSA(λ) mit “replacing traces” an.
2. Man betrachte einen complex Reward mit $R_{complex}(s) = \frac{1}{2}R^{TD(0)}(s) + \frac{1}{2}R^{MC}(s)$, wobei $R^{TD(0)}(s)$ den TD(0) und $R^{MC}(s)$ den first visit Monte Carlo Return bezeichne. Wie sähe dafür die eligibility trace im backwards view aus? (Tipp: Spalte die eligibility trace in zwei Terme auf.)

Aufgabe 9 (30%). Planen:

1. Was muesste man tun, um ein stochastisches Modell zu lernen?
2. Welche Vorteile bzw. Nachteile hat Dyna-Q im Vergleich zu Dynamischen Programmieren?