## Grundlagen der Künstlichen Intelligenz

Prof. Dr. B. Nebel, Prof. Dr. W. Burgard
Dr. A. Kleiner, R. Mattmüller
Sommersemester 2008

Universität Freiburg
Institut für Informatik

# Übungsblatt 11

**Abgabe: Freitag, 18. Juli 2008**

**Aufgabe 11.1** (Komplexität der exakten Inferenz)

Beweisen Sie, dass das 3-SAT-Problem auf das Problem der exakten Inferenz in Bayesschen Netzen reduziert werden kann und dass damit exakte Inferenz NP-schwer ist.
*Hinweis:* Betrachten Sie ein Netz mit einer Zufallsvariable für jede Aussagenvariable, einer für jede Klausel, und einer für die Konjunktion der Klauseln.

**Aufgabe 11.2** (Value-Iteration-Algorithmus)

Betrachten Sie die folgende Gitterwelt. Die $u$-Werte stehen für den Nutzen eines Zustandes, nachdem die *Value Iteration* konvergiert ist, $r$ für die Belohnung, die ein Zustand erbringt. Nehmen Sie an, dass $\gamma = 1$ und dass ein Agent vier mögliche Aktionen ausführen kann: **Nord**, **Süd**, **Ost** und **West**. Mit Wahrscheinlichkeit 0,7 erreicht der Agent den Zustand, den er erreichen will, mit Wahrscheinlichkeit 0,2 bewegt er sich nach rechts und mit Wahrscheinlichkeit 0,1 nach links von der beabsichtigten Richtung.

| | | |
|---|---|---|
| $u = 8$ | $u = 15$ | $u = 12$ |
| $u = 2$ | $r = 2$ | $u = 10$ |
| $u = 7$ | $u = 16$ | $u = 11$ |

Welches ist die beste Aktion, die ein Agent ausführen kann, der sich im zentralen Zustand der Gitterwelt aufhält? Erklären Sie Ihre Antwort. Welchen Nutzen hat der zentrale Zustand damit?

**Aufgabe 11.3** (Konvergenz der Value-Iteration)

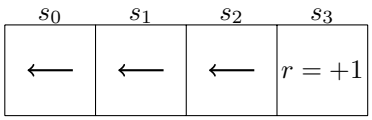
Betrachten Sie das Bellman-Update als einen Operator $B$, der, angewandt auf einen Nutzenvektor (eine Komponente für jeden Zustand), einen neuen Nutzenvektor liefert. Dann kann das Bellman-Update als $U_{i+1} \leftarrow BU_i$ geschrieben werden. Sei $|| \cdot ||$ die Max-Norm, d. h. $||U|| = \max_s |U(s)|$. Dann gilt für den wahren Nutzen $U$ (ohne Beweis)

$$||BU_i - U|| \leq \gamma ||U_i - U||.$$

Ferner gilt (wieder ohne Beweis), dass $|U(s)| \leq R_{\max}/(1-\gamma)$ für alle Zustände $s$, wobei $R_{\max}$ die maximale Belohnung in einem Zustand ist. Berechnen Sie, wieviele Iterationen $N$ des Value-Iteration-Algorithmus höchstens notwendig sind, bis der Fehler $||U_N - U||$ kleiner oder gleich einem gegebenen $\epsilon > 0$ wird. Sie dürfen $\gamma < 1$ annehmen.

**Aufgabe 11.4** (Policy-Iteration-Algorithmus)

Sei nun $\gamma = 0{,}5$ und die einzigen Aktionen seien **Ost** und **West**. Mit Wahrscheinlichkeit 0,9 erreicht der Agent den Zustand, den er erreichen will (bzw. bleibt stehen, falls die Aktion ihn über den Rand des Gitter hinausführen würde), und mit Wahrscheinlichkeit 0,1 bewegt er sich in die entgegengesetzte Richtung. Die Belohnung in den drei westlichen Zuständen ist jeweils $-0{,}05$.




Führen Sie einen Schritt der *Policy Iteration* durch, wobei die initiale Policy $\pi_0$ durch die Pfeile in den Zuständen gegeben ist. Geben Sie das lineare Gleichungssystem für die erste *Policy Evaluation* und eine Lösung des Gleichungssystems sowie die erste verbesserte Policy $\pi_1$ an.

Die Übungsblätter dürfen und sollten in Gruppen von drei (3) Studenten bearbeitet werden. Bitte füllen Sie das Deckblatt[1] aus und heften Sie es an Ihre Lösung.

[1] `http://www.informatik.uni-freiburg.de/~ki/teaching/ss08/gki/coverSheet-german.pdf`