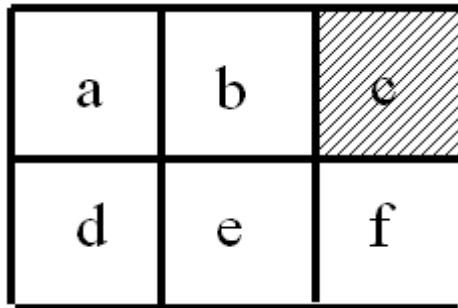


# Maschinelles Lernen und Data Mining

Übungsblatt für den 25.1.2005

## Aufgabe 1

Ein Agent bewegt sich in einer einfachen Welt, die wie folgt angeordnet ist:



Der Agent kann sich jeweils ein Feld nach oben, unten, rechts, oder links bewegen, falls sich in der jeweiligen Richtung ein Feld befindet. Wenn der Agent im Feld c rechts oben landet, erhält er einen Reward von 1.0, und kann sich von dort nicht mehr wegbewegen, an allen Feldern erhält er einen Reward 0.

- Formulieren Sie die Reward-Funktion.
- Stellen Sie die optimale Bewertungsfunktion  $V^*(s)$  auf, unter der Annahme, daß der Discount-Faktor  $\gamma = 0.9$  ist.
- Bestimmen Sie die Q-Funktion.
- Geben Sie eine optimale Policy an.
- Versuchen Sie, mit Q-learning die Q-Funktion zu lernen, indem Sie den Agenten auf ein zufällig gewähltes Anfangsfeld stellen, und jeweils die beste Aktion nach der momentanen Q-Funktion ausführen (bei Gleichheit zufällige Auswahl), bis der Agent am Zielfeld angekommen ist, und das ganze bis zur Konvergenz wiederholen. Die Lernrate  $\alpha$  können Sie mit 1 annehmen.