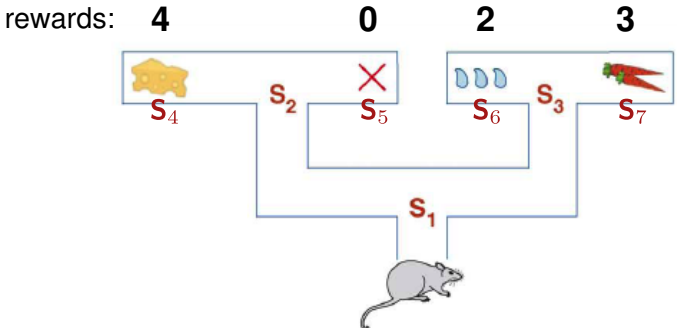
Computational models of goal-directed behavior exercise

Marc Toussaint
Machine Learning & Robotics lab, FU Berlin
Arnimallee, 14195 Berlin, Germany

November 16, 2010

1. Policy Iteration (10 points)

Consider the following T-maze:



We distinguish 7 states $S_1, .., S_7$ in the maze. The first 3 states are the T-junctions; the last 4 states receive rewards $(4, 0, 2, 3)$. At each T-junction we have two possible actions: left, right. Everything is deterministic.

a) Policy Evaluation: In the first iteration we initialize π_0 to be random (50/50 left/right at each junction). Assume $\gamma = 0.5$. What is the value function V^{π_0} for the random policy? First solve this by using the definition (eqn. 1, slide 12). Then solve this by iterating the Dynamic Programming equation (eqn. 2, slide 14).

b) Policy Improvement: What is the improved (deterministic) policy π_1 based on the value function V^{π_0}? (Use the equation given on slide 18.)

c) Policy Evaluation: What is the new value function V^{π_1}?

d) Policy Improvement: What is the improved policy π_2 based on V^{π_1}?

2. Goal-directed vs. habitual – Revaluation (5 points)

Take the same maze as above, but consider too different conditions of the rat:

- in the hunger condition the rewards are (4,0,2,3) as above
- in the thirst condition the rewards are (2,0,4,1) (rewarding the liquid in state S_6 most).

In the first experimental session the rat is always hungry and travels through the maze k times and learns to go to the cheese (as described by exercise 1).

In the second experimental session the rat is thirsty. How will the rat behave depending on k? (Following Niv, Joel Dayan, 2006.)

3. Direct policy evaluation (5 points)

There exists a very simple algorithm which computes the exact value function V^π w.r.t. the current policy in $O(n^3)$ via matrix inversion. Derive this algorithm from the Dynamic Programming equation (eqn. 2, slide 14): Write the equation in vector notation, e.g., defining $\vec{V}^\pi$ to be a vector of size n (n is the size of the state space) which contains all the values $V^\pi(s)$ as elements, and defining $\vec{T}^\pi$ to be an $n \times n$ matrix with elements $T^\pi_{ss'} = P(s' \,|\, \pi(s), s)$.