Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Matthias Schubert
Sebastian Schmoll, Sabrina Friedl

Munich, 07.01.2019

Deep Learning and Artificial Intelligence

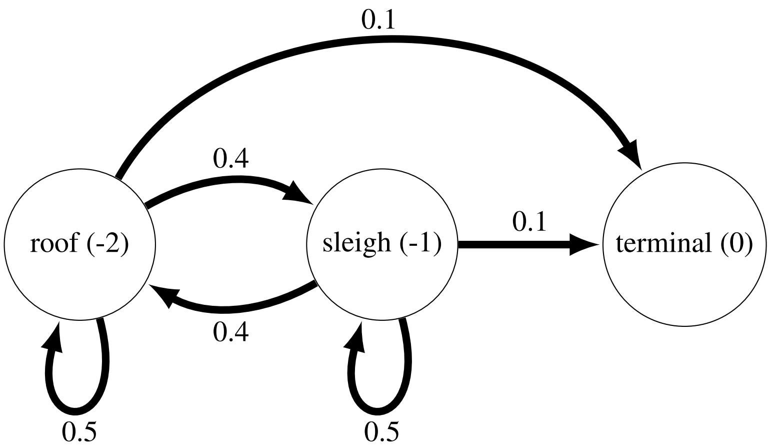
WS 2018/19

Exercise 10: Markov Decision Processes

Exercise 10-1 (Santa's) Markov Reward Process

On Christmas Eve, Santa has to deliver all presents to their recipients. To do so, he moves from house to house. Santa can try to throw a present into the chimney either directly from his sleigh or from the house's roof. After each throw, Santa decides that he has delivered enough presents and is done with the current house (terminal state) with probability 0.1. Otherwise, he switches his position from the sleigh to the roof or vice versa (probability 0.4) or stays put (probability 0.5).

Throwing the presents from the sleigh is a bit difficult, since Santa has to consider all winds and turbulences in the air. However, he is very experienced so that he will always hit the chimney successfully. The cost (negative reward) for throwing presents from the sleigh is thus -1. By contrast, the roofs are usually slippery and it may happen that Santa slips and falls to the ground so that he has to climb up again to deliver the present. Therefore, the cost for delivering presents from the roof is -2. The terminal state has a cost of 0. Below you can see Santa's (Markov) reward process representing Santa's work:



(a) Name the transition probability matrix P that is defined by the above Markov reward process.

(b) Compute the expected discounted future rewards (utilities) for the states "roof", "sleigh" and "terminal" using the Bellman equation (with $\gamma = 1$):

$$U(s) = R(s) + \gamma \sum_{s'} P(s'|s)U(s'), \quad \forall s \in S.$$

Exercise 10-2 Policy Iteration

Santa wants to know whether he can do better. He can now decide at any time (not just after a throw) whether he wants to change his position (i.e. switch from the roof to the sleigh or vice versa, action c) or throw a present (action t). Due to the wind and the slippy roof, changing position only succeeds with probability 0.8. With probability 0.2, he just stays where he was. Moreover, each throw leads to Santa being done (reaching the terminal state) with probability 0.1 and staying put with probability 0.9.

Tasks:

(a) Write down the transition probabilities for each action and state that are not 0 (you can also specify the transition matrix for each action). Moreover, draw the MDP described above, i.e. draw a node for each state and an edge for each possible transition. Use different colors for the two actions and annotate the edges with their transition probabilities and the nodes with their rewards.

(b) What can be said intuitively about the optimal policy in states x_s and x_r?

(c) Apply policy iteration to determine the optimal policy and the state-values of x_s and x_r. Assume the initial policy π_0 has action t in both states. For the policy-evaluation step, use the simplified Bellman equation for a fixed policy π_t :

$$U^{\pi_t}(x) = U^t(x) = R(x) + \gamma \sum_{s'} P(x' \mid x, \pi_t(x))\, U^t(x').$$

Why does Santa have such a big belly?

(d) What happens if the initial policy has action c in both states? Does discounting help?

Exercise 10-3 MDPs with Python

You will find a Jupyter notebook on the lecture web-page that contains a programming exercise on MDPs. Please follow the instructions in the notebook file.