Reinforcement Learning (SS18) - Exercise 1

Daniel Hennes

18.04.2018

Bandits

1. Consider ϵ-greedy action selection for a bandit with two actions ($k = 2$) and $\epsilon = 0.5$. What is the probability that the greedy action is selected?
2. Consider a k-armed bandit problem with $k = 4$ actions, denoted 1, 2, 3, and 4. Consider applying to this problem a bandit algorithm using ϵ-greedy action selection, sample-average action-value estimates, and initial estimates of $Q_1(a) = 0$, for all a. Suppose, you observe the following sequence of actions and rewards: $A_1 = 1, R_1 = 1, A_2 = 2, R_2 = 1, A_3 = 2, R_3 = 2, A_4 = 2, R_4 = 2, A_5 = 3, R_5 = 0$. On some of these time steps the ϵ case may have occurred, causing an action to be selected at random.
   (a) On which time steps did this definitely occur?
   (b) On which time steps could this possibly have occurred?

Markov-chains

3. A Markov chain with state space $\mathcal{S} = \{1, 2, 3\}$ has the transition probability matrix:

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

   Note that $P_{i,j} = \Pr\{S_{t+1} = j \mid S_t = i\}$ for any t.
   (a) Draw the transition diagram of the Markov chain.
   (b) Assume the process starts in state 1 (i.e., $S_1 = 1$). What is the conditional probability $\Pr\{S_3 = 3, S_2 = 2 \mid S_1 = 1\}$?
   (c) Determine the conditional probability $\Pr\{S_4 = 3, S_3 = 2 \mid S_2 = 1\}$.
   (d) What is the stationary distribution for this Markov chain?
   (Find the probability distribution over states that remains unchanged as time progresses.)
4. A Markov chain with state space $\mathcal{S} = \{1, 2, 3\}$ has the transition probability matrix:

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}$$

   (a) Is state 3 absorbing?
   (b) What is the expected time to reach state 3?
5. A fair coin is tossed repeatedly. Find the expected number of tosses till the pattern *heads, tails, heads* (HTH) appears first.