Excercise 1

Kai Schneider

April 25, 2021

Task 1

a.)

$k = 2$, $\varepsilon = 0.5$
$\rightarrow P(\text{greedy}) = 1 - \varepsilon + \frac{\varepsilon}{k} = 1 - 0.5 + \frac{0.5}{2} = 0.75$
$\rightarrow P(\text{non-greedy}) = \frac{\varepsilon}{k} = \frac{0.5}{2} = 0.25$

b.)

$k = 4 \ \rightarrow a_i$ with $i = 1:4,\ Q_1(a_i) = 0$

with $A_t = \underset{a}{\text{argmax}} Q_t(a)$ as the greedy policy and $Q_t(a) = \frac{\sum\limits_{i=1}^{t-1} R_{i,a_i=a}}{n(a)}$ and the given data:

$$\begin{aligned} A_1 &= 1 & R_1 &= 1 \\ A_2 &= 2 & R_2 &= 1 \\ A_3 &= 2 & R_3 &= 2 \\ A_4 &= 2 & R_4 &= 2 \\ A_5 &= 3 & R_5 &= 0 \end{aligned}$$

I:

Step 1 (from Q_1 to Q_2) was definitely a random step because $Q_1(a_i) = 0\ \forall i$, therefore the selection was arbitrary.

	a_1	a_2	a_3	a_4	action
Q_1	0	0	0	0	$A_1 = 1$
Q_2	1	0	0	0	$A_2 = 2$
Q_3	1	1	0	0	$A_3 = 2$
Q_4	1	3	0	0	$A_4 = 2$
Q_5	1	5	0	0	$A_5 = 3$
Q_6	1	5	0	0	-

A random selection also has to be occured in step 2 ($Q_2 \rightarrow Q_3$), because $A_2 = 2$ despite the $\underset{a}{\operatorname{argmax}}$ being 1. Also in the fifth step ($Q_5 \rightarrow Q_6$) action A_3 was selected, which had to be a random selection too.

II:

In general a random step could have occured at any other point too. Especially step 3 ($Q_3 \rightarrow Q_4$) is a likely candidate, because the *argmax* is either 1 or 2. But even if the chosen A_i is the $\underset{a}{\operatorname{argmax}} Q_t(a)$, it is still possible that this was a random selection.