Homework 6

Daniel Frank

May 30, 2017

1   LSPI Batch Method

1.1   Problem

In chapter 9.1 it says that samples for each run were collected by choosing
random actions for 25 steps. All together there are 50 samples. To generate
samples I took random actions and stored them in one big matrix like

$$
D = \begin{pmatrix}
\text{sample} = 1 & \begin{bmatrix} s_0 & \dots & s_{25} \end{bmatrix} \\
\text{sample} = 1 & \begin{bmatrix} r_0 & \dots & r_{24} \end{bmatrix} \\
\text{sample} = 1 & \begin{bmatrix} a_0 & \dots & a_{24} \end{bmatrix} \\
& \vdots \\
\text{sample} = 50 & \begin{bmatrix} s_0 & \dots & s_{25} \end{bmatrix} \\
\text{sample} = 50 & \begin{bmatrix} r_0 & \dots & r_{24} \end{bmatrix} \\
\text{sample} = 50 & \begin{bmatrix} a_0 & \dots & a_{24} \end{bmatrix}
\end{pmatrix}.
\tag{1}
$$

When I do LSPI I take all the samples in D, the outcome looks almost perfect
after the first iteration shown in figure 1. The Q Values do not change much over
the next iterations and LSPI converges after only three steps taking $\epsilon = 10^{-3}$.
The plots are not as expected and also not as the ones in chapter 9.1, can you
help me here? Is there something wrong with my sample D?

Also I am not sure how to understand the LSTQ-Algorithm. When I update
$\tilde{A}$ which policy should I take in $\phi(s', \pi(s'))$? In my case I took $\text{argmax}_{a \in A} \phi(s, a)^T w$
where I took an initial w in the first step, Is that correct?
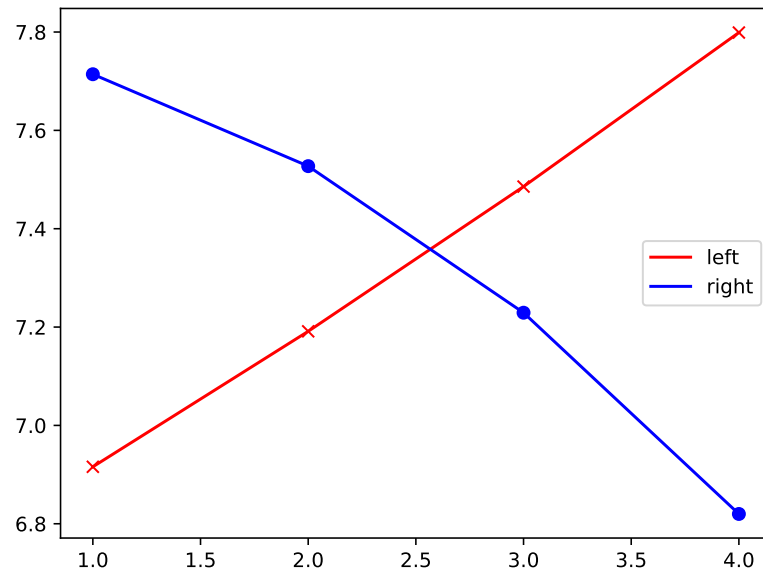
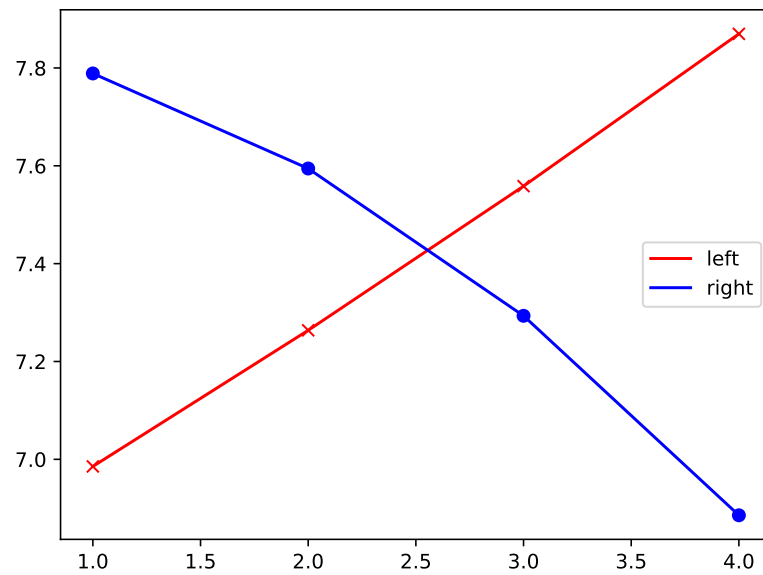Figure 1: LSPI after first iteration
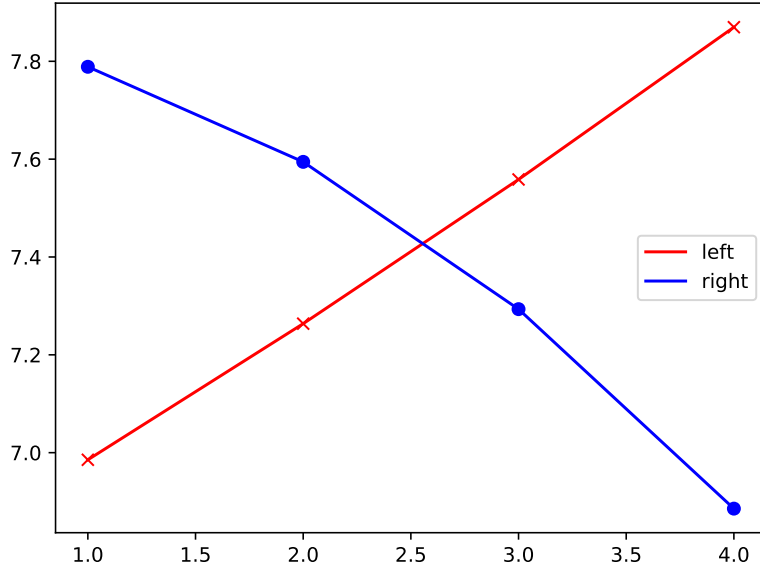

Figure 2: LSPI after second iteration

Figure 3: LSPI after third iteration

1.2 Solution: taking random starting state for samples

1.2.1 4 states

The solution mainly depends on the samples that are generated one solution in shown in table 1.2. There are also samples where it does not converge to the optimal policy, which is I guess on of the downsides of LSPI since it can get stuck in local minima.

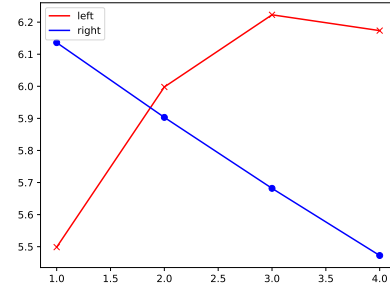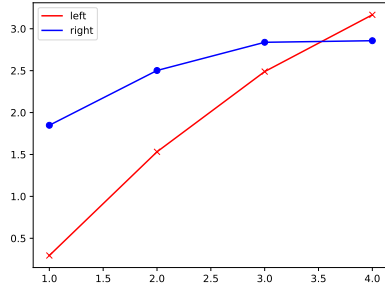Table 1: Evaluation of $Q(s, a; w)$. Ends after 5th iteration

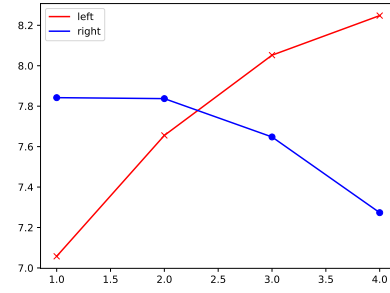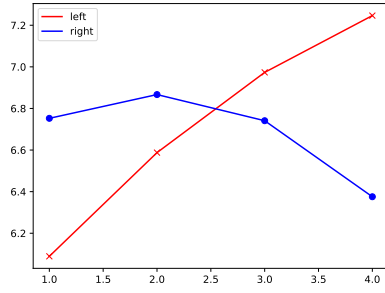Table 2: 1st iteration

Table 3: 2nd iteration
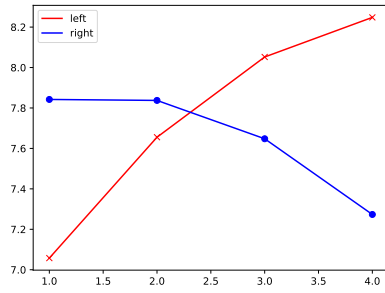

Table 4: 3rd iteration

Table 5: 4th iteration

Table 6: 5th iteration