Problem 1.

Consider a system with $\gamma = 0.9$ the following state and action spaces: $S = \{-1, 1, 2\}$, $A = \{-1, 0, 1\}$. The available batch data are as follows:

$$D = \{(s_0 = 1, a_0 = 1, r_1 = 1, s_1 = 2), (s_1 = 2, a_1 = 0, r_2 = -1, s_2 = 1)$$
$$(s_2 = 1, a_2 = -1, r_3 = 0, s_3 = -1)\}$$

Consider the basis function $\Phi(s, a) = a^2 s + as + a$ with initial weights $\omega^0 = 1$. Perform LSPI to compute ω^1 and ω^2, and policy associated to ω^2.

* In case of tie for action seletion, give the preference to -1, then 0 and finally +1.

Example $\longrightarrow$ $\underset{a \in \{-1, 0, 1\}}{\arg\max} \left\{ \frac{-1}{2} \quad \frac{0}{2} \quad \frac{+1}{2} \right\} = -1$

Problem 2.

Repeat problem 1 using the basis function $\Phi(s,a)=\begin{bmatrix} as+a \\ a^2 s \end{bmatrix}$ with initial weights $w^0=\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Perform LSPI to compute w^1 and w^2, and policy associated to w^2. Is the final policy (i.e., π^2) different from problem 1? Can two basis functions, in general, lead to different policies?

Problem 3.

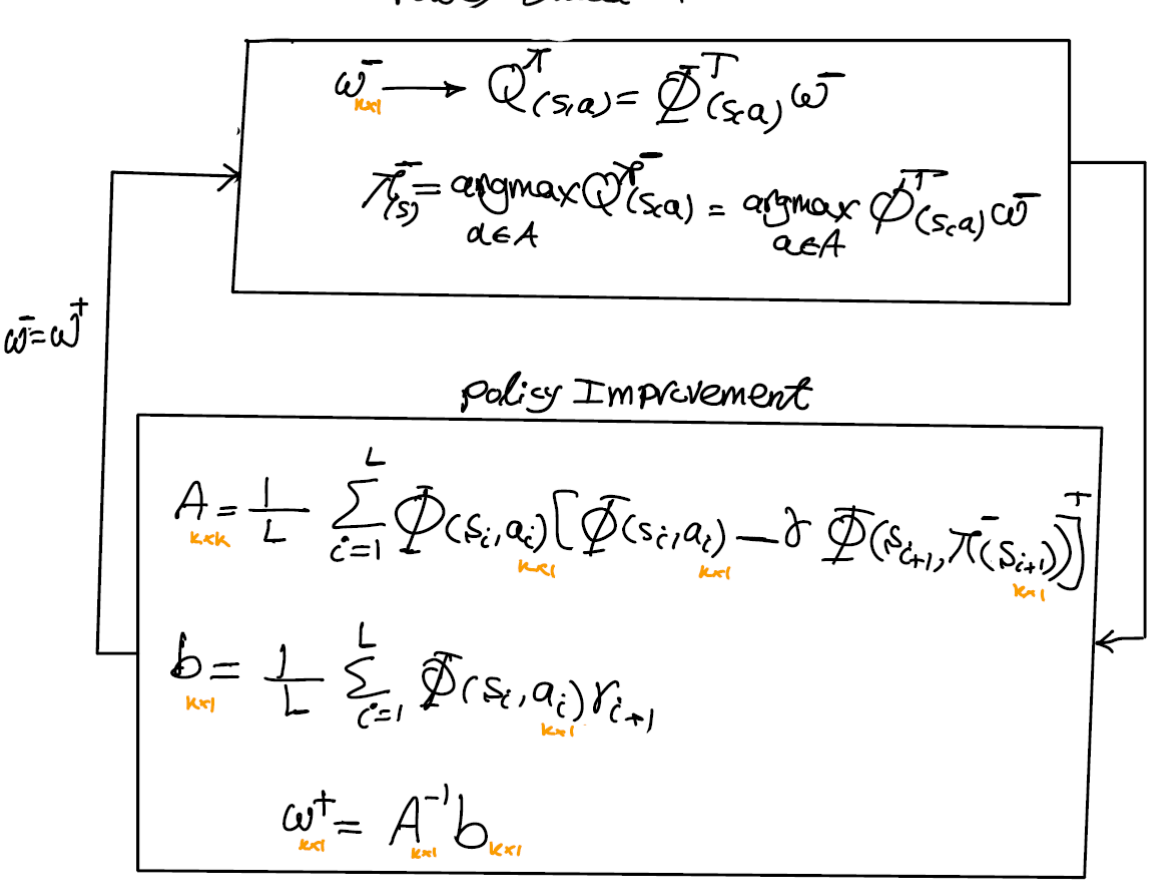
Consider a system with the following continous state and action spaces: $S=[-2 \quad 2]$, $A=[-1.5 \quad 2]$. The available batch data is as:

$$D=\{(s_0=2, a_0=1, r_1=1, s_1=1), (s_1=1, a_1=-1, r_2=2, s_2=-1)\}$$

Consider $\gamma=0.9$, the basis function $\Phi(s,a)=\begin{bmatrix} a \\ s \times a \end{bmatrix}$ with initial weights $w^0=\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$. Perform LSPI to compute w^1 and w^2. Compute the policy associated with w^2 for any $s \in [-1 \quad 1]$.

Least-squares policy Iteration

Batch Data: $D = \{(s_1, a_1, r_2, s_2), \ldots, (s_L, a_L, r_{L+1}, s_{L+1})\}$



Questions about the HW should be directed to TA, Begum Taskazan, at taskazan.b@northeastern.edu.