Homework Assignment 3

Zhaoyang Xu

September 22, 2020

Contents

1 Policy evaluation (20 points)

From question, we can get

$$\begin{aligned} V_{k+1}(s) &\leftarrow \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V_k(s')] \\ &= \sum_a \pi(s,a) \sum_{s'} P^a_{ss'} R^a_{ss'} + \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}[\gamma V_k(s')] \\ &= \mathbf{r}^\pi + \gamma \mathbf{T}^\pi \mathbf{v}^\pi \quad (vector\ form) \end{aligned}$$

We define F in vector form: $F(\mathbf{v}) = \mathbf{r}^\pi + \gamma \mathbf{T}^\pi \mathbf{v}^\pi$
Now, let's calculate the distance with ∞-norm between any $\mathbf{v}$ and $\mathbf{u}$

$$\begin{aligned} \|F(\mathbf{u}) - F(\mathbf{v})\|_\infty &= \|(\mathbf{r}^\pi + \gamma \mathbf{T}^\pi \mathbf{u}^\pi) - (\mathbf{r}^\pi + \gamma \mathbf{T}^\pi \mathbf{v}^\pi)\|_\infty \\ &= \|\gamma \mathbf{T}^\pi (\mathbf{u}^\pi - \mathbf{v}^\pi)\|_\infty \\ &\leq \|\gamma \mathbf{T}^\pi \|\mathbf{u}^\pi - \mathbf{v}^\pi\|_\infty\|_\infty \\ &\leq \gamma \|\mathbf{u}^\pi - \mathbf{v}^\pi\|_\infty \end{aligned}$$

F has a unique fixed point, and $\mathbf{v}^\pi$ is a fixed point. By contraction mapping theorem, iterative policy evaluation converges on $\mathbf{v}^\pi$.

2 Maze example (20 points)

From One-step Q-learning on page 67 in Slide, we can get

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \qquad (*)$$
$$a = argmax\ Q(s_{t+1}, a)$$

Thus, we can define the final state value

$$Q(s_{t+1}, a) = \frac{\frac{Q(s_t,a_t)-Q(s_t,a_t)}{\alpha} - r_{t+1} + Q(s_t, a_t)}{\gamma}$$
$$a = argmax\ Q(s_{t+1}, a)$$

By this question, it become

$$Q(s_{t+1}, a) = \frac{Q(s_t, a_t) - 0}{0.9}$$
$$Q(s_{t+1}, a) = \frac{Q(s_1, a_1) - 0}{0.9^t}$$

So, we can get

$$Q(s_{t+1}, a) = \frac{5.3}{0.9^6} = 10$$

After we get the goal value 10, the states directly to the left and below can be calculated by $(*)$,

$$Q(s_{t-1}, right) = R(s_{t-1}, right) + 0.9 * \max Q(s,a_t) = 0.9 * 10 = 9$$
$$Q(s_{t-1}, up) = R(s_{t-1}, right) + 0.9 * \max Q(s_t, a_t) = 0.9 * 10 = 9$$

3 Majority Vote (15 points)

3.1

Flipping a coin for three times. If we observe "head", $X_i = 1$; if "tail", $X_i = 0$. And the probability of the coin flip is

$$P(X_i = 1) = 1 \qquad (i = 1, 2, 3)$$
$$P(X_i = 0) = 0$$

So, $\mathbf{X} = \{1, 1, 1\}$. Now, we define $\mathcal{H}$
For the first hypotheses, we assume the first flip is "tail" and others are "heads".
For the second hypotheses, we assume the second flip is "tail" and others are "heads".
For the third hypotheses, we assume the third flip is "tail" and others are "heads".
Thus, we can get

X_i	1	1	1	
h_1	0	1	1	$L(h_1) = 1/3$
h_2	1	0	1	$L(h_2) = 1/3$
h_3	1	1	0	$L(h_3) = 1/3$
MV	1	1	1	$L(MV) = 0$

where $L(MV) = 0$ and $L(h) \geq \frac{1}{3}$ for all h.

3.2

Decision space $\mathbf{X}$ is as same as 3.1, but we change $\mathcal{H}$.
For the first hypotheses, we assume the first flip is "head" and others are "tails".
For the second hypotheses, we assume the second flip is "head" and others are "tails".
For the third hypotheses, we assume the third flip is "head" and others are "tails".
Thus, we can get

X_i	1	1	1	
h_1	1	0	0	$L(h_1) = 2/3$
h_2	0	1	0	$L(h_2) = 2/3$
h_3	0	0	1	$L(h_3) = 2/3$
MV	0	0	0	$L(MV) = 1$

where $L(MV) > L(h)$ for all h.

3.3

MV makes an error equals the probability that at least half of hypothesis $\mathcal{H}$ make an error. Consider a binary classification,

$$\begin{aligned} L(MV) &= \sum_{k=0}^{M/2} \binom{M}{k} (1 - (\frac{1}{2} - \epsilon))^k (\frac{1}{2} - \epsilon)^{M-k} \\ &= \sum_{k=0}^{M/2} \binom{M}{k} (\frac{1}{2} + \epsilon)^k (\frac{1}{2} - \epsilon)^{M-k} \\ &\leq e^{-\frac{M}{2}\,\mathrm{kl}(\frac{1}{2}-\epsilon \| \frac{1}{2}+\epsilon)} \\ &\leq e^{-4M\epsilon^2} \\ &= \frac{1}{e^{4M\epsilon^2}} \overset{M\to\infty}{\longrightarrow} 0 \end{aligned}$$

Hence, we can get $L(MV)$ converges to 0 exponentially fast with the growth of M.

4 Follow The Leader (FTL) algorithm for i.i.d. full information games (45 points)

Follow the leader is an algorithm that at round t uses the best action up to round t. Lets assume $K = 2$, and $a \neq a^*$,

$$a_t = argmax R_t(a)$$

Now, we can write the algorithm down explicitly

FTL Algorithm
for t≤T **do**
Pull arm a_t = $argmax R_t(a)$
end for

By definition on page 59 in notes, and let a^* be an optimal action, and $\Delta = \mu(a^*) - \mu(a)$(assume a game with rewards),

$$\begin{aligned}
\overline{R}_T &= \max_a \mathbb{E}[\sum_{t=1}^{T} r_t^a] - \mathbb{E}[\sum_{t=1}^{T} r_t^{A_t}] \\
&= \sum_a \Delta(a)\mathbb{E}[N_T(a)] \\
&= \sum_a \Delta(a)\mathbb{E}[\sum_{t=1}^{T} \mathbb{1}_{(A_t=a)}] \\
&\leq \sum_a \Delta(a) \sum_{t=1}^{T} \mathbb{P}(\hat{\mu}_{t-1}(a) \geq \hat{\mu}_{t-1}(a^*))
\end{aligned}$$

FTL algorithm chooses the arm with the largest rewards mean in past rounds. Thus, we only need to consider the exploration steps. In exploration steps, it will choose action $a \neq a^*$, when $\hat{\mu}_{t-1}(a) \geq \hat{\mu}_{t-1}(a^*)$ at rounds t. Now our aim is to calculate the probability of this situation. Lets consider about $\mathbb{P}(\hat{\mu}_{t-1}(a) \geq \hat{\mu}_{t-1}(a^*))$.

$$\begin{aligned}
\mathbb{P}(\hat{\mu}_{t-1}(a) \geq \hat{\mu}_{t-1}(a^*)) &\leq \mathbb{P}(\hat{\mu}_{t-1}(a) \geq \hat{\mu}_{t-1}(a^*)) \\
&\leq \mathbb{P}(\hat{\mu}_{t-1}(a) \geq \mu_{t-1}(a) + \frac{1}{2}\Delta(a)) + \mathbb{P}(\hat{\mu}_{t-1}(a^*) \leq \mu_{t-1}(a^*) - \frac{1}{2}\Delta(a)) \\
&\leq 2e^{-2t(\frac{1}{2}\Delta(a))^2} \\
&= 2e^{-\frac{1}{2}t\Delta(a)^2}
\end{aligned}$$

The third line follows hoeffding's inequality. There are two bad events, empirical mean of suboptimal arm greater than expected mean of it adds $\frac{1}{2}\delta$, it means that the empirical mean may show bigger than the optimal arm. Then, the action will choose the suboptimal arm a. Also, for the optimal arm, if empirical mean of optimal arm smaller than expected mean of it minus $\frac{1}{2}\delta$, it will show smaller than the suboptimal.

Hence,

$$\begin{aligned}
\overline{R}_T &\leq \sum_a \Delta(a) \sum_{t=1}^{T} \mathbb{P}(\hat{\mu}_{t-1}(a) \geq \hat{\mu}_{t-1}(a^*)) \\
&\leq \sum_a \Delta(a) \sum_{t=1}^{T} 2e^{-\frac{1}{2}t\Delta^2} \\
&= \sum_a \Delta(a) \frac{2}{1 - e^{-\frac{\Delta(a)^2}{2}}}
\end{aligned}$$

The third step applied geometric series. Now we can get a bound of $\overline{R}_T$.