Reinforcement Learning Homework 5

Deadline 19th November 2024
Students Sahiti Chebolu, Surabhi S Nath, Xin Sui

Answers to the coding questions are in the associated Jupyter notebook. Answers to theory questions are as follows.

Theory Question 1: Linear regret for ϵ-greedy

To prove: $\mathcal{R}_{\mathcal{V}}(T) \geq \epsilon \frac{K-1}{K} \Delta_{min} T$

- We know regret is given by $\mathcal{R}_{\mathcal{V}}(T) = \sum_{a=1}^{K} \Delta_a \mathbb{E}[N_a(T)]$ where $a \in 1, ..., K$ are the arms available

- The expected number of times a is chosen is given by: $\mathbb{E}[N_a(T)] = \mathbb{E}[\sum_{t=1}^{T} \mathbb{1}(A_t = a)]$

- For an ϵ-greedy policy, each arm is chosen at least with probability $\frac{\epsilon}{K}$ at every round t. Hence, $\mathbb{E}[N_a(T)] = \mathbb{E}[\sum_{t=1}^{T} \mathbb{1}(A_t = a)] \geq \frac{\epsilon T}{K}$

- The difference in means of an arm and the best arm, Δ_a is at least as big as the difference between the best and second best arm, Δ_{min}. Further $\Delta_a = 0$ for the best arm

- Hence

$$\begin{aligned}
\sum_{a=1}^{K} \Delta_a \mathbb{E}[N_a(T)] &= \sum_{a \neq a*} \Delta_a \mathbb{E}[N_a(T)] \text{ since } \Delta_a = 0 \text{ for } a = a* \\
&\geq \sum_{a \neq a*} \Delta_{min} \frac{\epsilon T}{K} \\
&\geq \Delta_{min} \frac{\epsilon T}{K} \sum_{a \neq a*} 1 \\
&\geq \Delta_{min} \frac{\epsilon T}{K} K - 1
\end{aligned}$$

Theory Question 1: Explore-Then-Commit (ETC)

(a)

To find: $\mathbb{E}[N_a(T)]$ for ETC

- $\mathbb{E}[N_a(T)] = \sum_{t=1}^{T} P(A_t = a)$

- Until $mK < T$ rounds, each arm is chosen m times. After mK rounds, the arm with the best empirical mean is chosen. $P(\hat{a} = a)$ is the porbablity that arm a is committed on. So we can split the interval $t \in 1, ..., T$ into $t \in 1, ..., mK$ and $t \in mk + 1, ..., T$:

$$
\begin{aligned}
\mathbb{E}[N_a(T)] &= \sum_{t=1}^{T} P(A_t = a) \\
&= \sum_{t=1}^{mK} P(A_t = a) + \sum_{t=mK+1}^{T} P(A_t = a) \\
&= m + \sum_{t=mK+1}^{T} P(A_t = a) \\
&= m + \sum_{t=mK+1}^{T} P(\hat{a} = a) \\
&= m + (T - mK)P(\hat{a} = a)
\end{aligned}
$$

(b)

We want to bound $P(\hat{a} = a) \leq P(\hat{\mu}_a \geq \hat{\mu}_{a*})$

- This can be re-written as: $P(\hat{\mu}_a - \hat{\mu}_{a*} \geq 0) = P(\hat{\mu}_a - \hat{\mu}_{a*} + \mu_a - \mu_{a*} \geq \mu_a - \mu_{a*}) = P(\hat{\mu}_{a*} - \hat{\mu}_a + \Delta_a \leq \Delta_a)$
- $\hat{\mu}_{a*} - \hat{\mu}_a$ is sub-Gaussian with mean Δ_a
- By the Hoeffding's inequality:

$$
P(\hat{\mu}_{a*} - \hat{\mu}_a + \Delta_a \leq \Delta_a) \leq exp(-\frac{m\Delta_a^2}{2\sigma^2})
$$

where m is the number of samples.

(c)

As $m \to \infty$, the smaller the bound.