The University of Vermont

CS253A QR: Reinforcement Learning: Assignment №3

Ayat Ospanov

September 15, 2018

Contents

1 Exercise 2.6: Mysterious Spikes

When given optimistic initial values, $Q_n(a) > R_n, \forall a$ at the first steps. This means, that

$$Q_{n+1} = Q_n + \alpha(R_n - Q_n) \leq Q_n$$

Therefore, all next step values are less (or equal) than the values at the previous step. As we initialized all starting values with a priori big values, there are lots of equal maximum values. Consequently, there is a big rate of optimal actions (optimal by algorithm) corresponding to those values. When all Q_ns drop under real high values of distributions, the curve stabilizes. This effect was also mentioned in the Exercise 2.5.

2 Exercise 2.7: Unbiased Constant-Step-Size Trick

As we have shown in the exercise 2.4, Q_{n+1} for variable α is:

$$Q_{n+1} = \prod_{i=1}^{n}(1-\alpha_i)Q_1 + \sum_{i=1}^{n}\left(\alpha_i \prod_{j=i+1}^{n}(1-\alpha_j)R_i\right)$$

Here the weight given to R_i is $\alpha_i \prod_{j=i+1}^{n}(1-\alpha_j)$. This is exponential. As we have $\alpha_i = \beta_i$, we have to show that $\beta_i \in [0,1]$ to prove that Q_n is an exponential recency-weighted average. Given $\bar{o}_n = \bar{o}_{n-1} + \alpha(1-\bar{o}_{n-1})$,

$$\beta_n = \frac{\alpha}{\alpha + (1-\alpha)\bar{o}_{n-1}}$$

By definition, $\bar{o}_n \geq 0, \forall n \geq 0$ and $\alpha \in [0,1]$. Thus, $\alpha + (1-\alpha)\bar{o}_{n-1} \geq \alpha$. Knowing this, it is obvious that $\beta_n \in [0,1]$.

Now, let's show that Q_n doesn't have the initial bias. To do this, have a look at Q_2:

$$Q_2 = Q_1 + \beta_1(R_1 - Q_1)$$

Since $\bar{o}_0 = 0$,

$$\beta_1 = \frac{\alpha}{\alpha + (1-\alpha)\bar{o}_0} = \frac{\alpha}{\alpha + (1-\alpha)0} = 1$$

This means, that $Q_2 = Q_1 + R_1 - Q_1 = R_1$.

We have proven that Q_n is an exponential recency-weighted average without initial bias.