Ishaan Bassi
2016238

Date / /

Assignment 1 Theory

Q1 (Exercise 2.5)

In the stationary case, our estimates Q converge to q^* due to the law of large numbers. However, in non-stationary case, q^* (the expected value of reward) changes at each step. In such a case, using constant step-size results in recent rewards getting more weight. Hence in the long run constant step size (α) performs better than sample average ($\frac{1}{n}$)

Q4. Stationary Case

In stationary case, it is observed ~~by~~ that optimistic greedy performs best followed by UCB and then epsilon greedy. As the q^* i.e the expected value of rewards don't change with time, the optimal action remains same. As discussed in Ex 2.6 answer (Q3), the optimistic greedy agent explores (due to low rewards) and takes the optimal action for rest of the simulation. UCB ~~takes~~ explores but also takes uncertainity in the estimates into account. Optimistic finds optimal action faster and with straightforward approach (without exploration).

Date **Non stationary** (UCB v/s optimistic greedy v/s ϵ-gre...

Q4.(b) In non stationary case, the optimal action can change with time, hence **exploration** is **needed**. ~~B~~Hence optimistic greedy that relies on ~~expen~~ exploring only in the initial stage performs worst. UCB is again very slightly better that epsilon greedy.

_/_/_

Q 2. Exercise 2.6

In optimistic greedy approach, the agent tends to explore more in the initial steps. This is due to the dissatisfying rewards it recieves i.e the actual rewards recieved are much less than the optimistic action value estimates and hence different actions are tried. As optimal action would also be chosen, along with others hence no. of times it is chosen (across several runs) in the initial steps gives spikes.

(Exercise 2.7)

Q3. The incremental rule for estimated reward Q is given by :

$$Q_{n+1} = Q_n + \underbrace{\alpha}_{\text{step size}} [R_n - Q_n]$$

with step size $\beta_n = \frac{\alpha}{\bar{o}_n}$, we have and $\bar{o}_n = \bar{o}_{n-1} + \alpha(1 - \bar{o}_{n-1})$

$$\begin{aligned}
Q_{n+1} &= Q_n + \beta_n [R_n - Q_n] \\
&= \beta_n R_n - \beta_n Q_n + Q_n \\
&= \beta_n R_n + (1 - \beta_n) . Q_n \\
&= \beta_n R_n + (1-\beta_n)\left(Q_{n-1} + \beta_{n-1}(R_{n-1} - Q_{n-1})\right) \\
&= \beta_n R_n + (1-\beta_n)\beta_{n-1} R_{n-1} + (1-\beta_n) Q_{n-1} - \beta_{n-1} Q_{n-1}(1-\beta_n) \\
&= \beta_n R_n + (1-\beta_n)\beta_{n-1} R_{n-1} + (1-\beta_n)(1-\beta_{n-1}) Q_{n-1} \\
&\;\vdots \\
&= \beta_n R_n + (1-\beta_n)(\beta_{n-1} R_{n-1}) - - - + \prod_{i=1}^{n}(1-\beta_i) Q_1
\end{aligned}$$

Now as $\bar{o}_0 = 0 \therefore \bar{o}_1 = 0 + \alpha(1-0) = \alpha$

$$\Rightarrow \boxed{\beta_1 = 1}$$

$\therefore \prod_{i=1}^{n}(1-\beta_i) = 0 \quad \therefore$ Initial bias is eliminated.