SUB
DATE

CISC 856 - Reinforcement Learning
Assignment 1

Name: Hanan Fared Mohamed OMara.
ID: 20398559

→ Exercise 1 ←

1- first, without knowing which case (A or B) you are facing at each step:

- The expected value of action 1 is
  0.5 * 0.1 + 0.5 * 0.9 = 0.5
- The expected value of action 2 is
  0.5 * 0.2 + 0.5 * 0.8 = 0.5

So, the best expected success, we can achieve is 0.5 (by choosing either action with equal probability).

2- Second, knowing which case you are in at each step:

- in case A, the optimal action is action 2 with expected value 0.2
- in case B, the optimal action is action 1 with expected value 0.9.

⇒ by choosing the optimal action for each case, the best expected success is 0.5 * 0.2 + 0.5 * 0.9 = 0.55

SUB
DATE

The summary → without knowing the facts of situation, the best predicted success is 0.5 which may be obtained by doing both actions equally

- Knowing the case, the best predicted success is 0.55, which may be obtained by selecting the best action for each case (action 2 for case A, action 1 for case B).

* Exercise 2
from lecture 3 in slide 22
→ we went to Maximize the return of rewards :-

$$G_t \doteq R_{t+1} + R_{t+2} + R_{t+3} + \cdots R_T$$

where T is the final time step (or horizon)
→ since there is no terminal time T, we need to use

$$G_t \doteq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

* where $0 \leq \gamma \leq 1$ is the discount rate (or factor). Then

$$G_t \doteq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \cdots$$

$$= R_{t+1} + \gamma (R_{t+2} + \gamma R_{t+3} + \gamma^2 R_{t+4} + \cdots)$$

$$= R_{t+1} + \gamma G_{t+1}$$

in qestion say → assume that the Reward $R_t = 1$
Demonstrate that :

$$G_t = \sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma}$$

SUB
DATE

This Previous equation even in The Continuous Task Case, G_t is bounded, if The reward is always $= 1$ for $0 \leq \gamma \leq 1$

→ The result for Reinforcement learning :-

If reward always $= 1$ and $\gamma = 1$

Because This equation here, (This Summation) = geometric Series, we should use gamma factor less than one because if equals to one This will given $\frac{1}{0}$, That is going to "Infinity" but using values less Than 1 will have a bounded return

→ let's Show The example given in lecture :-

gamma $= 0.95$ and All rewards equal to 1 → The result $= \frac{1}{1-0.95} = \frac{1}{0.05}$

→ And if $\gamma = 0$ Then The ratio is Constant $= 1$

Then Now I will Demonstrate This equation

$$G_t = \sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma}$$

→ If we say as (question) say That The reward at time step t (R_t) is always equal to 1, Then we can Simplify The expression for discounted future return G_t as follow:

$$G_t = R_{t+1} + \gamma G_{t+1} = 1 + \gamma G_{t+1}$$

→ we can Then substitue This expression into itself recursively as $G_{(t+1)} = 1 + \gamma G_{(t+2)}$

$$G_{(t+2)} = 1 + \gamma G_{(t+3)}$$

$$G_{(t+3)} = 1 + \gamma G_{(t+4)}$$

SUB
DATE

Subsitituting These values back into the original equation

$$G_t = 1 + \gamma(1 + \gamma(1 + \gamma\, G_{(t+4)}))$$

we Can See The Pattern where The term γ^k is multiplied by $G_{(t+k)}$, Let's Consider The limit as K approaches infinity:

$$G_t = 1 + \gamma\left(1 + \gamma\left(1 + \gamma(1 + \ldots)\right)\right)$$

→ This infinite sum Can be written as a geometric Series with a Common ratio γ. The sum of a geometric series with a Common ration between -1 and 1 is given by:

$$\Sigma(\gamma^k) = \frac{1}{1-\gamma}$$

in This Case, Since $0 < \gamma < 1$, The Sum Converges, and we have:-

$$\Sigma(\gamma^k) = \frac{1}{(1-\gamma)}$$

$$\therefore G_t = \frac{1}{(1-\gamma)}$$

→ The Significance of This result will Provide a closed form expression for Calculating The expected Sum of discounted rewards in an episodic Task with a Constant reward of 1 and discount factor Gamma. This allows RL Algorithm to estimate and optimize The expected return.

SUB
DATE

* Exercise 3 :-

inclass we discussed the concept of "Exponential Weighted Average". Demonstrate that the right hand side of

$$Q_{n+1} = (1-\alpha)^n Q_1 + \sum_{i=1}^{n} \alpha (1-\alpha)^{n-i} R_i$$

is an exponential weighted average. Explain why?

"From the Record of Lecture 2 and Slide ⇒
⇒ first consider that a single action has been selected $n-1$ times. The estimate Q_n of this action value is

$$Q_n = \frac{R_1 + R_2 + \cdots R_{(n-1)}}{(n-1)}$$

As $(n-1)$ is the total number of R

→ This equation can be implemented incrementally

* incremental implementation

$$Q_{n+1} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{1}{n}\left(R_n + \sum_{i=1}^{n-1} R_i\right)$$

$$= \frac{1}{n}\left(R_n + (n-1)\frac{1}{n-1}\sum_{i=1}^{n-1} R_i\right)$$

$$= \frac{1}{n}\left(R_n + (n-1) Q_n\right) = \frac{1}{n}\left(R_n + n Q_n - Q_n\right)$$

$$= Q_n + \frac{1}{n}\left[R_n - Q_n\right]$$

SUB
DATE

The update rule is

New Estimate ← old Estimate + Step size [Target − old Estimate]

* We can use a constant step size:

$$Q_{n+1} \overset{\text{def}}{=} Q_n + \alpha\left[R_n - Q_n\right]$$

* Where $\alpha \in (0, 1]$ is constant. Then

$$Q_{n+1} = Q_n + \alpha\left[R_n - Q_n\right] = \alpha R_n + (1-\alpha) Q_n$$

$$= \alpha R_n + (1-\alpha)\left[\alpha R_{n-1} + (1-\alpha) Q_{n-1}\right]$$

$$= \alpha R_n + (1-\alpha)\alpha R_{n-1} + (1-\alpha)^2 Q_{n-1}$$

$$= \alpha R_n + (1-\alpha)\alpha R_{n-1} + (1-\alpha)^2 \alpha R_{n-2} +$$

$$\cdots + (1-\alpha)^{n-1}\alpha R_1 + (1-\alpha)^n Q_1$$

$$\boxed{Q_{n+1} = (1-\alpha)^n Q_1 + \sum_{i=1}^{n} \alpha (1-\alpha)^{n-i} R_i.}$$

This equation is called an exponential weight average as This is an average because if you add This number and all other numbers, The summation is going to be equal to 1, also The exponential weighted average is a weighted average that assigns exponentially decreasing weights to past values, resulting in smoothed out average that is more responsive to recent values.