Assignment 12

Miao-Chin Yen

March 8, 2022

Problem 2

Reference td_lambda_prediction.py.

Problem 3

$$G_t - V(S_t) = \sum_{u=t}^{T-1} \gamma^{u-t} \cdot (R_{u+1} + \gamma \cdot V(S_{u+1}) - V(S_u))$$

Proof:

$$G_t - V(S_t) = R_{t+1} + \gamma \cdot R_{t+2} + \gamma^2 \cdot R_{t+3} + ... + \gamma^{T-t-1} \cdot R_T - V(S_t) + \sum_{j=1}^{T-t} \gamma^j \cdot [V(S_{t+j}) - V(S_{t+j})]$$

$$= R_{t+1} + \gamma \cdot R_{t+2} + \gamma^2 \cdot R_{t+3} + ... + \gamma^{T-t-1} R_T +$$

$$-V(S_t) + \gamma \cdot V(S_{t+1}) - \gamma \cdot V(S_{t+1}) + \gamma^2 \cdot V(S_{t+2}) - \gamma^2 V(S_{t+2}) + ... + \gamma^{T-t} V(S_T) - \gamma^{T-t} V(S_T)$$

We can reorder these terms.

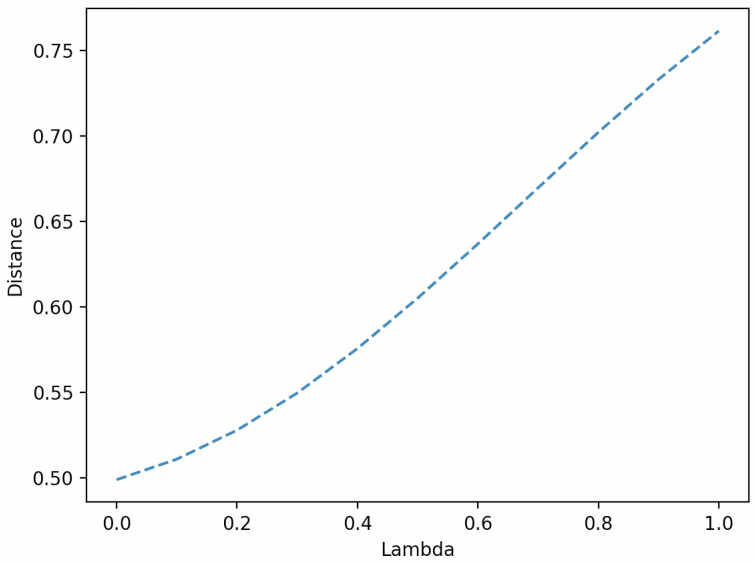
$$G_t - V(S_t) = R_{t+1} + \gamma \cdot V(S_{t+1}) - V(S_t) + \gamma \cdot R_{t+2} + \gamma^2 \cdot V(S_{t+2}) - \gamma \cdot V(S_{t+1})$$

$$+ ... + \gamma^{T-t-1} R_T + \gamma^{T-t} \cdot V(S_T) - \gamma^{T-t-1} \cdot V(S_{T-1})$$

$$= \sum_{u=t}^{T-1} \gamma^{u-t} \cdot (R_{u+1} + \gamma \cdot V(S_{u+1}) - V(S_u))$$

Problem 4

We again test on frog puzzle problem in assignment 2. Since we will use Monte-Carlo and the SimpleInventoryMRPFinite is not episodic. Therefore, we can not get trace experiences from this easily. I remembered that Sven told me Monte Carlo can only be applied to episodic process.
Reference td_lambda_prediction.pyand frog_puzzle_mrp.py
We notice that when we are given more data (traces), we would get a solution close to the one given by DP. We test for different λ and the result is as follows. As λ is getting closer to 1, the value convergence becomes lower.(The distance between the solution we get and the actual solution becomes higher.) I think that when λ gets closer to 1, it is like Monte-Carlo method which is weighted sum method. Hence, our result is supported. We also uses the script we have in assignment 11 and found that $TD(1)$ behaves just like Monte-Carlo and $TD(0)$ behaves just like TD method. However, the distance is a little bit larger than expected. I think this is the same issue as in assignment 11.


0.75
0.70
0.65
0.60
0.55
0.50
Distance
0.0
0.2
0.4
0.6
0.8
1.0
Lambda