CS3613

Introduction to Artificial Intelligence

Reinforcement Learning Lab Exercise

Group 30

Nayanathara P.M.C. 210417X

Neluhena M.C. 210419F

Nilakshan A.P.P. 210425U

Ninduwara K.G.M. 210429K

OpenAI Gym's Taxi Environment

Task : Comment on the influence the learning rate and discount rate have on how fast Q-learning can converge. Plot the necessary graphs to justify your answer. State any assumptions clearly.

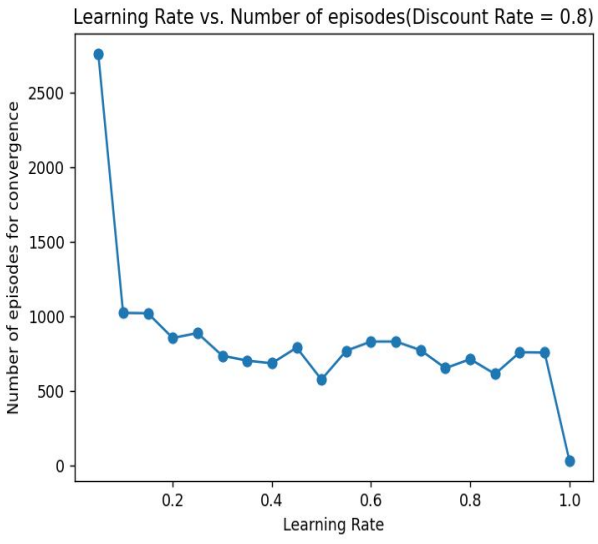
Influence of learning rate on the convergence of Q-learning

The learning rate refers to a hyper-parameter used to govern the pace at which an algorithm updates or learns the values of a parameter estimate.

High learning rates lead to faster initial learning and exploration because a high learning rate allows the agent to quickly adjust its policy based on the most recent experiences. And also, this can lead to high oscillations or divergence(instability) in the q table values.

Learning Rate vs Number of Episodes to converge the Q table

Learning Rate	Discount Rate	Steps
0.05	0.8	2763
0.1	0.8	1024
0.15	0.8	1022
0.2	0.8	856
0.25	0.8	889
0.3	0.8	737
0.35	0.8	704
0.4	0.8	686
0.45	0.8	791
0.5	0.8	577
0.55	0.8	770
0.6	0.8	832
0.65	0.8	832
0.7	0.8	774
0.75	0.8	653
0.8	0.8	714
0.85	0.8	614
0.9	0.8	760
0.95	0.8	758
1	0.8	32



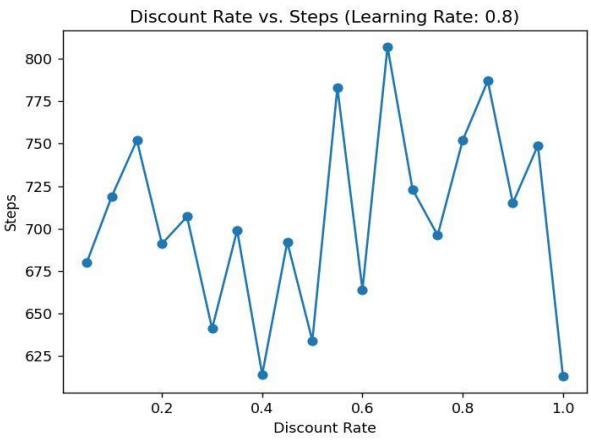
Higher learning rate takes less number of episodes to converge, so high convergence.

Influence of discount rate on the convergence of Q-learning

This hyper-parameter influences the agent's decision-making process. The discount rate determines how much the agent values future rewards compared to the immediate rewards. It plays a significant role in shaping the agent's behavior and impact on the Q-values.

Discount Rate vs Number of Episodes to converge the Q table

Learning R	Discount R	Steps
0.8	0.05	680
0.8	0.1	719
0.8	0.15	752
0.8	0.2	691
0.8	0.25	707
0.8	0.3	641
0.8	0.35	699
0.8	0.4	614
0.8	0.45	692
0.8	0.5	634
0.8	0.55	783
0.8	0.6	664
0.8	0.65	807
0.8	0.7	723
0.8	0.75	696
0.8	0.8	752
0.8	0.85	787
0.8	0.9	715
0.8	0.95	749
0.8	1	613



The impact of the discount rate on the number of episodes required to converge is context-dependent. It depends on factors such as problem complexity, the stability of the environment and the presence of long-term dependencies.