

Mountain Car with Q Learning Algorithm

Abstract:

In this study, Q learning algorithm was used to solve the mountain car problem. The algorithm containing the basic components of Q learning has been adapted for the mountain car from the Gymnasium environment list. Training with a basic Q table was tested using different learning parameters. The values of the Q table in binary form are saved to the “pkl” file while the code is running. At the end of the episodes, findings about the learning process are recorded in png format. The findings are shared in the results and analysis sections. There are clear explanations in the comment lines in the code file.

Parameter Exploration:

Learning Rate (α): The learning rate in Q-learning, alpha, determines how much the learned value is updated; A value of 0 means that learning did not occur, and the error is calculated by subtracting the old value from the learned value and multiplying it by the learning rate.

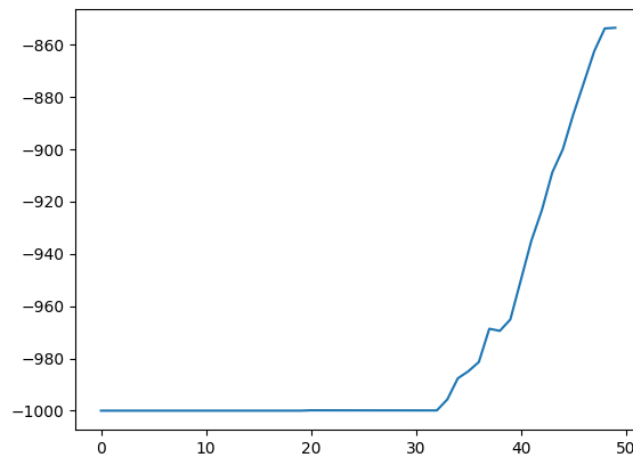
Discount Factor (γ): Choosing the discount factor smaller makes the strategy less short-sighted.

Experimental Setup:

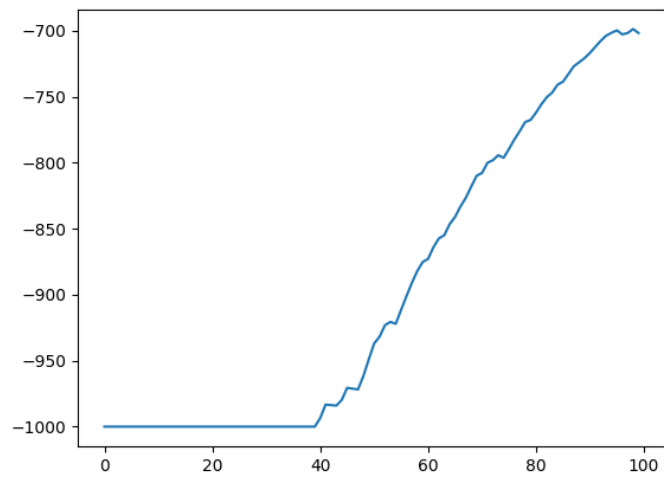
Information about setting up the experimental environment and running the code is described in READ.me.

Results and Analysis:

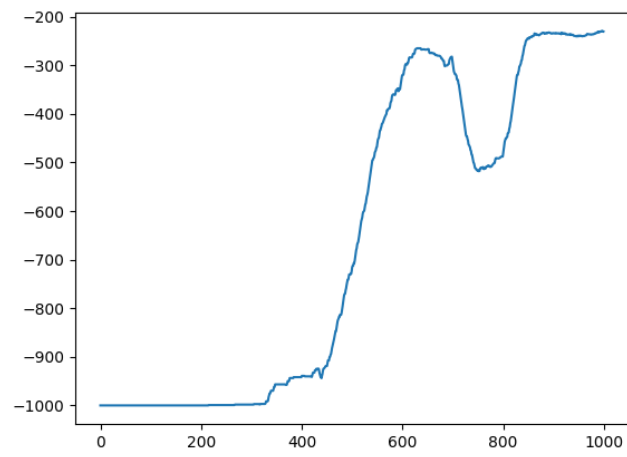
When `learning_rate_a = 0.9` (alpha or learning rate), `discount_factor_g = 0.9` (gamma or discount factor) are selected:



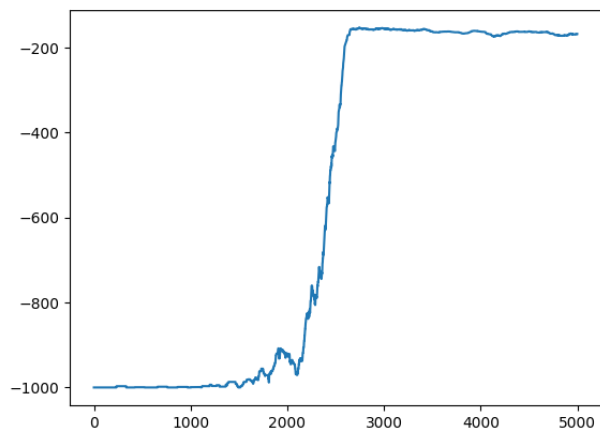
Episodes value is 50



Episodes value is 100

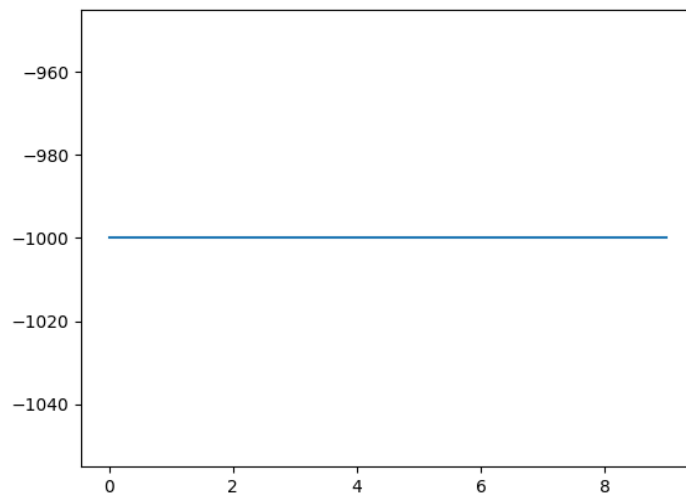


Episodes value is 1000

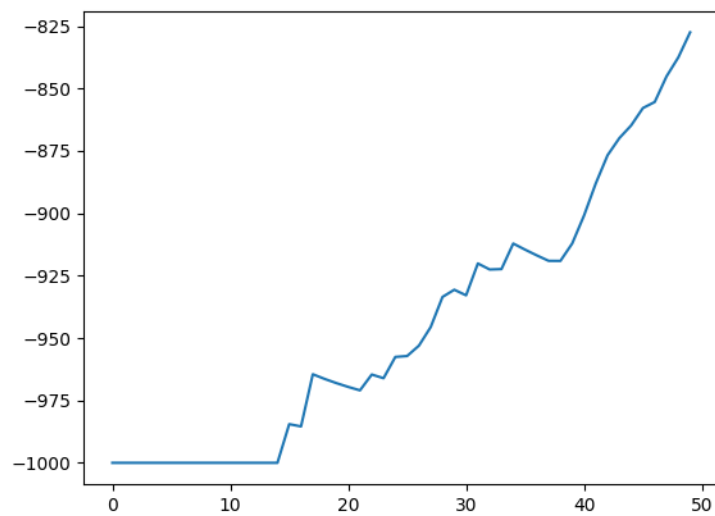


Episodes value is 5000

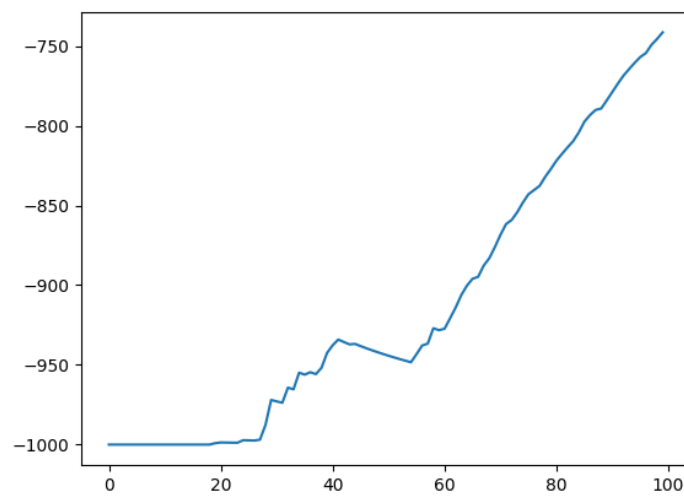
When $\text{learning_rate_a} = 0.3$ (alpha or learning rate), $\text{discount_factor_g} = 0.9$ (gamma or discount factor) are selected:



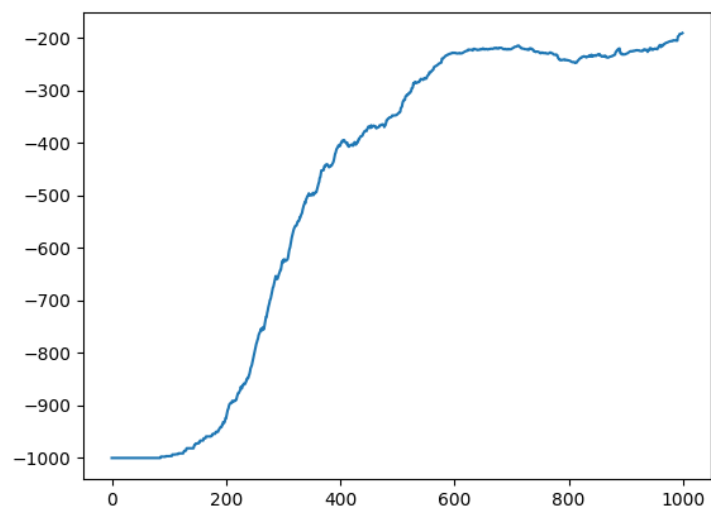
Episodes value is 10



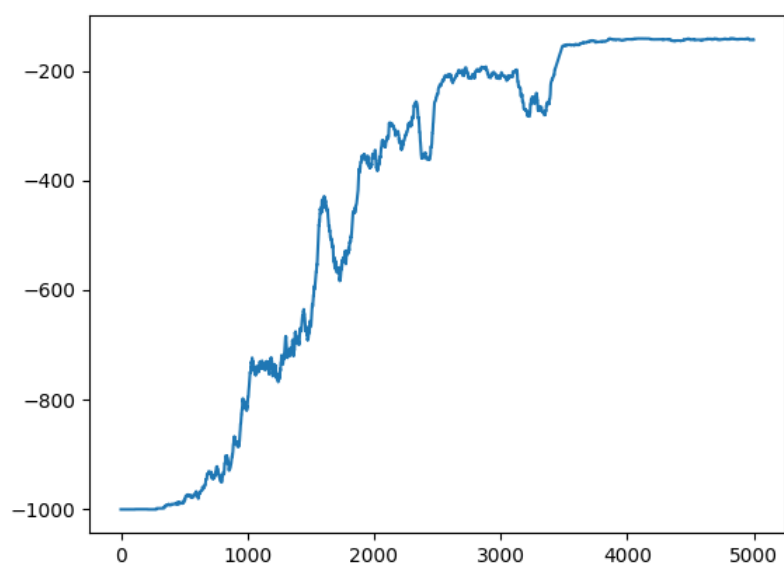
Episodes value is 50



Episodes value is 100

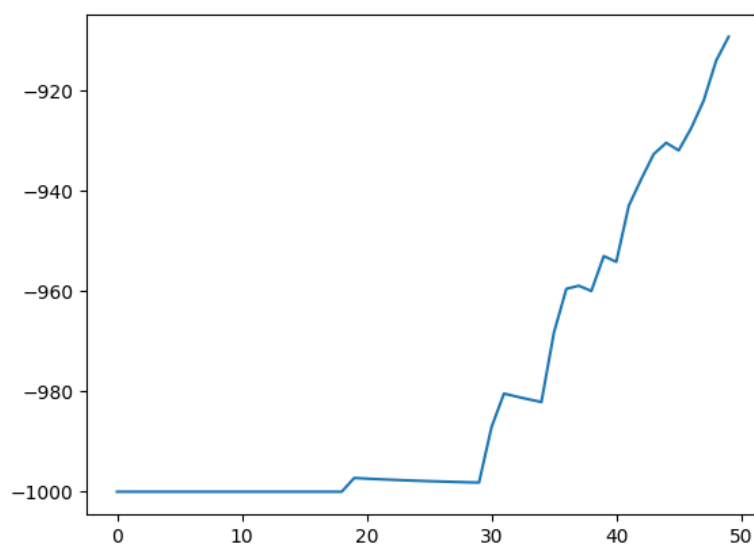


Episodes value is 1000

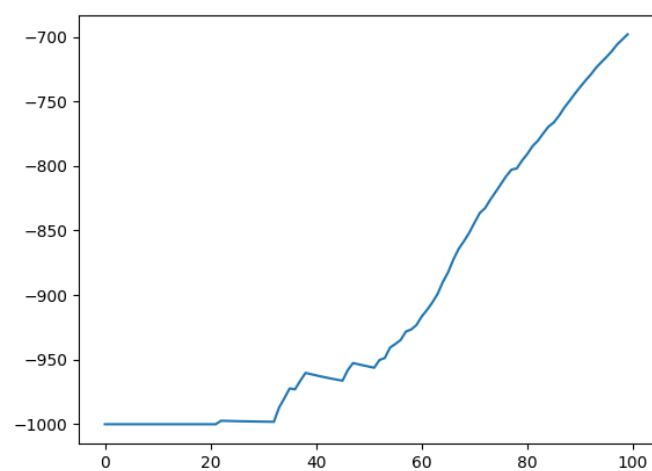


Episodes value is 5000

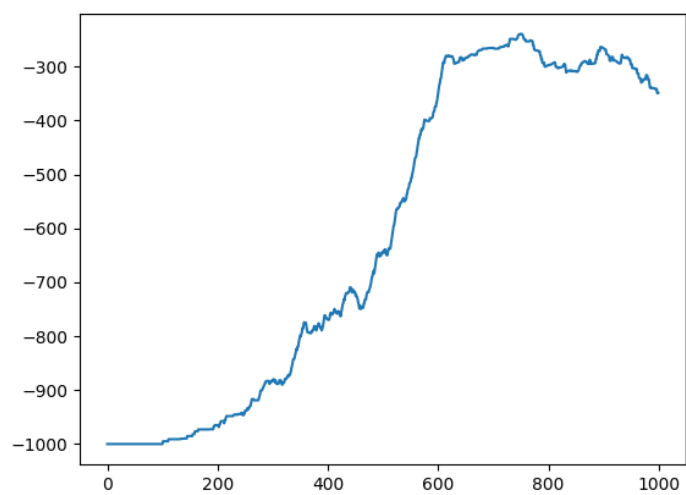
When $\text{learning_rate_a} = 0.3$ (alpha or learning rate), $\text{discount_factor_g} = 0.3$ (gamma or discount factor) are selected:



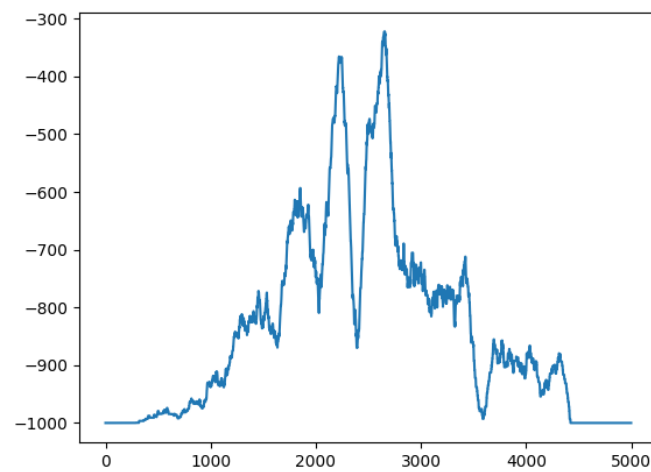
Episodes value is 50



Episodes value is 100

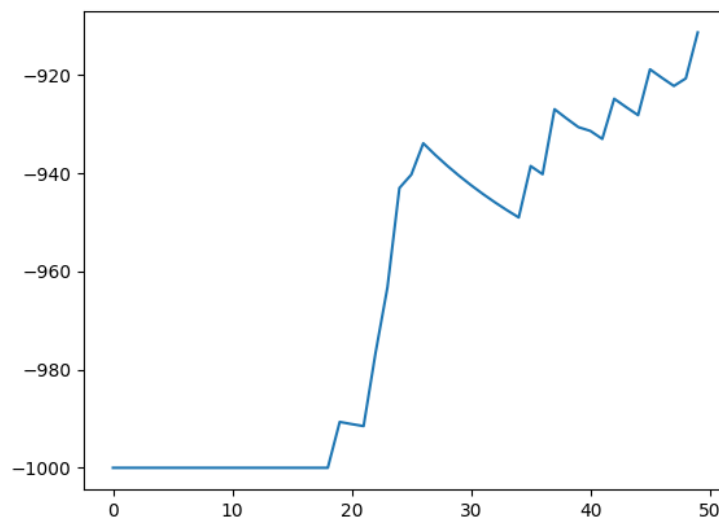


Episodes value is 1000

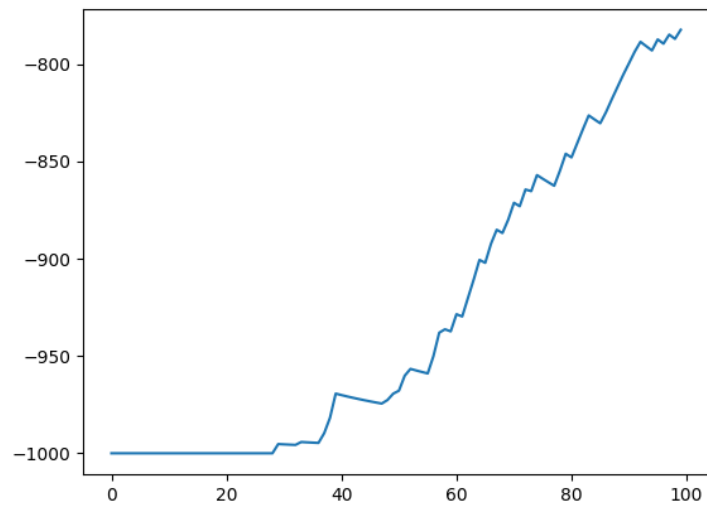


Episodes value is 5000

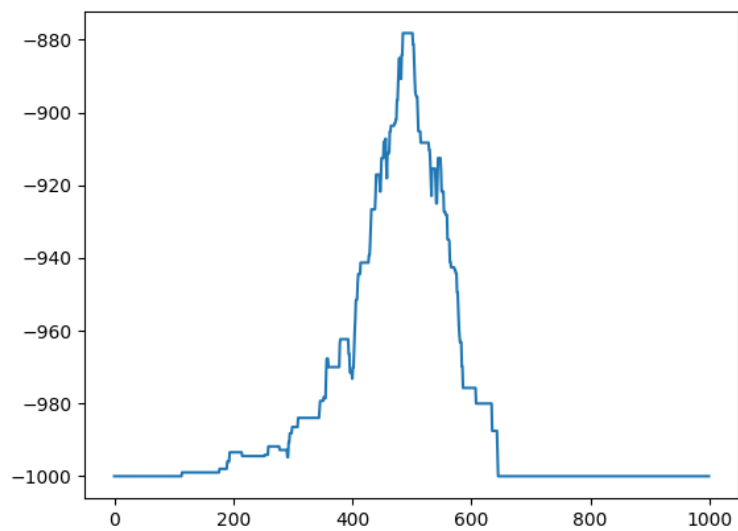
When $\text{learning_rate_a} = 0.9$ (alpha or learning rate), $\text{discount_factor_g} = 0.3$ (gamma or discount factor) are selected:



Episodes value is 50



Episodes value is 100



Episodes value is 1000

Conclusion:

When the result graphs from the Q learning algorithm tested with different parameters were examined, the most optimal parameters were recorded when 0.9 was selected for the learning rate and discount factor.

More tests could have been done with different parameters, but due to the long learning period, only these results were reached.