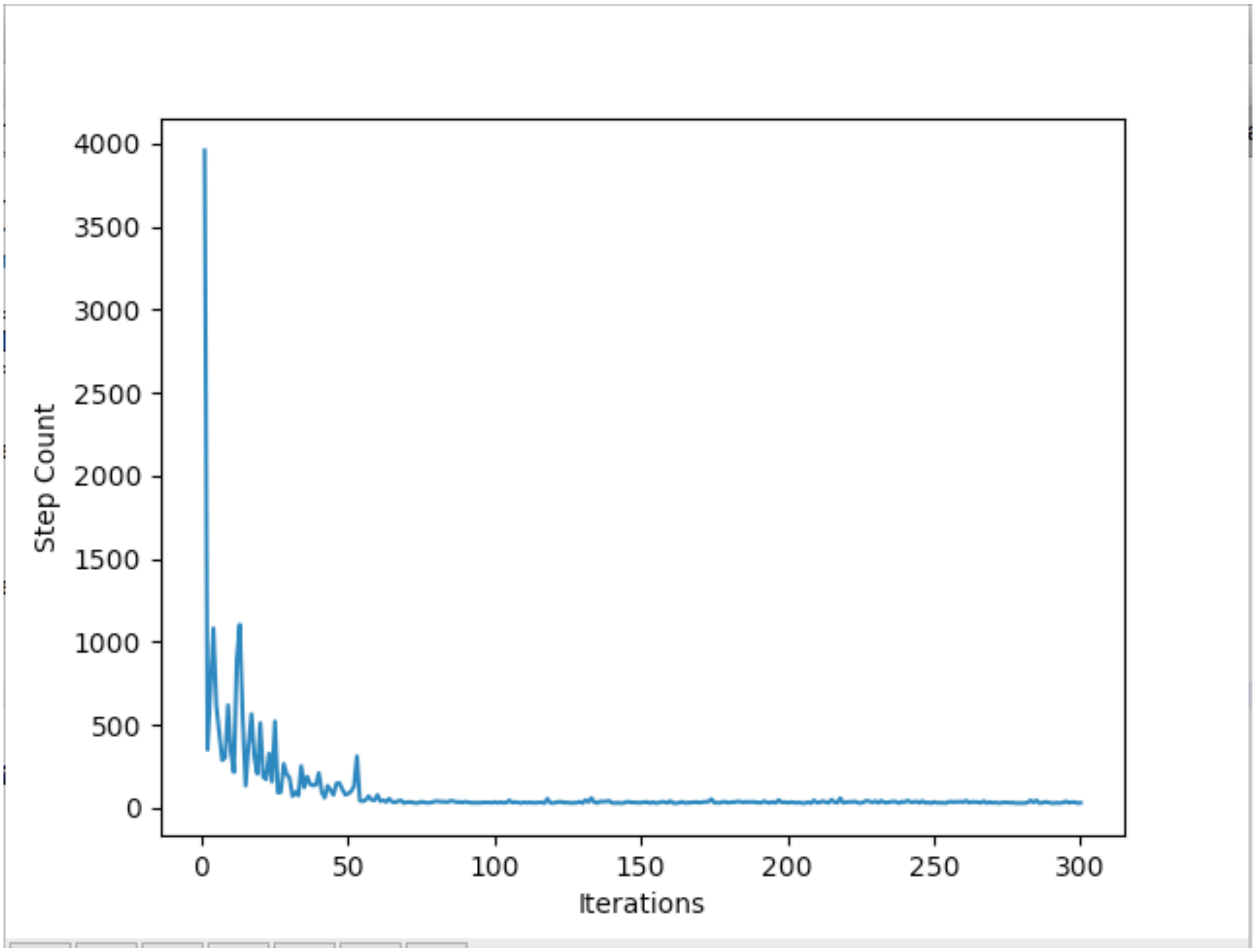
List of the parameters used in learning:

Reward=100 if final state is reached 0 otherwise

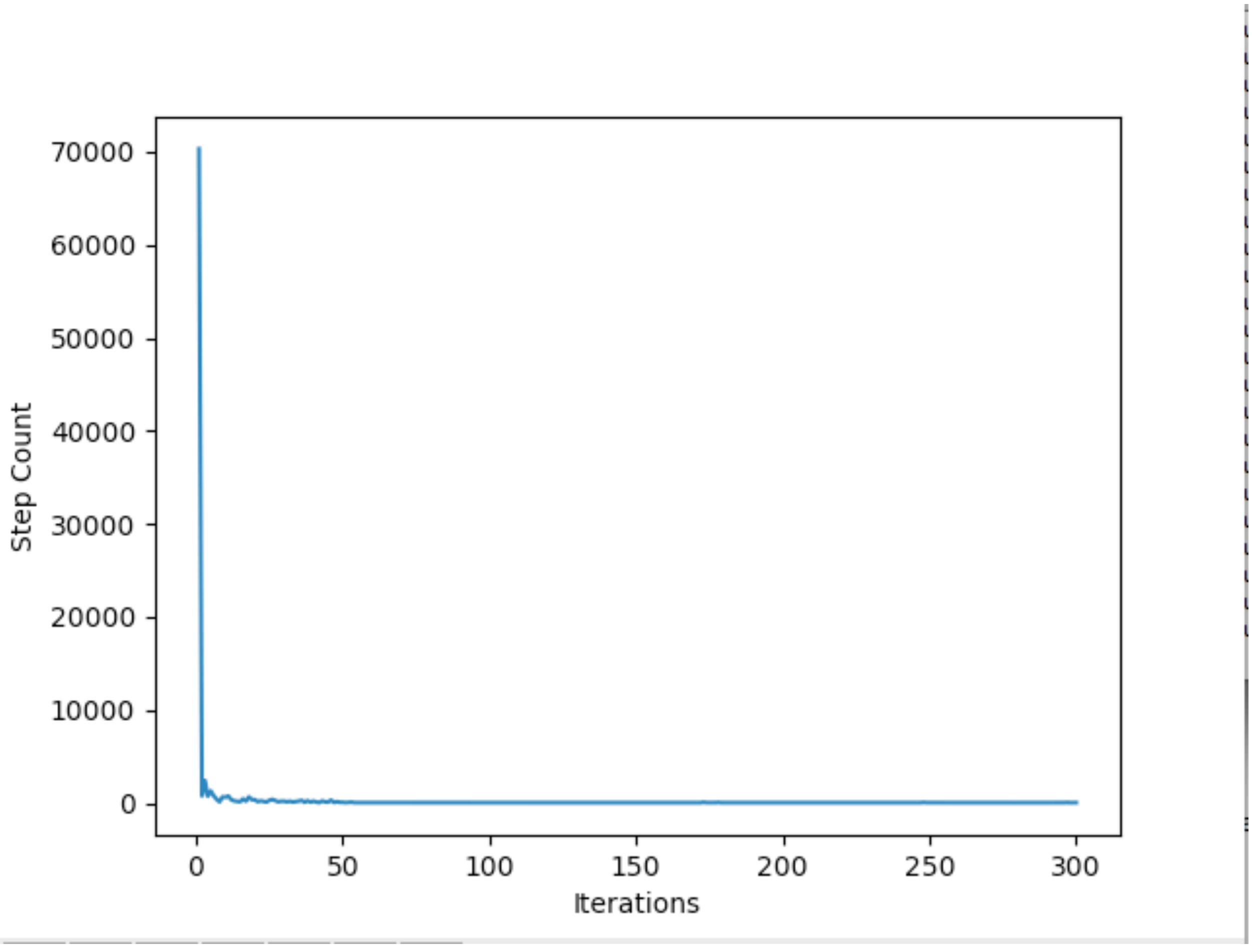
Discount Factor =0.75

Epsilon Greedy =0.1

Learning Curve:



Learning curve when the Q table is initialized with 1:



The learning curve for this converges quickly as compared to previous one. The reason is that when the Q table is initialized with 1, the reward propagates more quickly as compared to when it is zero as discount factor times the reward is zero for all cases here except the final state. But when initialized with 1, this is not the case.