Project 3

6.2 Observation and Report

Version 1: Vanilla ϵ-greedy

The closest I was able to get this to value iteration is difference in 3 states with `alpha` = 0.2 and `epsilon` = 0.6

With alpha dropping to 0.2, it allows the step to exploit large amount of prior knowledge while retaining a little bit of most recent information. Since we do not know the transition probability and the environment dynamics, it's important to have some flexibility on how much to be updated with some new information while using the previous information. And epsilon is slightly increased for greater threshold of flexibility

Version 2: Decay ϵ-greedy

The closest I was able to get this to value iteration is difference in 2 states with `alpha` = 0.2 `epsilon_0` = 0.65

Alpha was lowered for same reason as above. Since we've lowered the steps with the information needed to step through, initial epsilon (threshold) is also lowered since it's already decaying based on current episode for more accurate reinforcement learning, there isn't a need for it to be so high at the first place

Version 3: Effective Exploiting ϵ-greedy

The closest I was able to get this to value iteration is difference in 2 states with `alpha` = 0.5 `epsilon` = 0.6

Threshold is raised slightly higher (0.5 → 0.6) since we're taking more aggressive approach of picking best action - Q values of state not all zeros and sampling (instead of only sampling like version 1), it allows for more accurate results since there is another factor we're checking before picking the action