Week 3 mainly focus on Monte Carlo method which generally is a model-free learning methods that learn without prior knowledge of the state to next state transitions given our actions. Lab 3 contains 4 files each one represents different approach of finding the optimal policy. The first file “policy from values” performs iterative policy evaluation of the input policy. Iterative policy evaluation loop through all state giving a specific policy to compute the value of each state. After getting all values, the function takes those values to find the policy, which is looping through all states, and then loop through all possible actions at the state and get the destination value for the current action. And computing the value using (reward + gamma\*value at destination).

Policy iteration file, used after value of each state is being computed. It examines rewards and values to find better policy, and it keeps examine until no better policy has been found. How it works is by updating values by performing iterative policy evaluation then updating the policy using the updated values, then see if the policy changed it keeps updating until the policy does not change. I have noticed that it takes resources for looping and computing the policy by going through each state then going to each possible action and keep doing it. I think it is not applicable to large scale applications.

Last file is Value iteration, which do the same process as policy iteration, but the difference is that value iteration keeps update the value until converge occurs, then it computes the new policy.

Honestly, I find Value iteration file is confusing and there are areas I do not understand.