Assignment 4: Markov Decision Process

Shailesh Tappe: stappe3@gatech.edu

Objective: The objective of the assignment is to analyze and explore Markov Decision Process (MDP) technique of reinforcement learning (RL) with two different problem sets to explore MDP. An analysis is performed to understand policy iteration (PI), value iteration (VI) and Q learning (QL) algorithms of MDP and understand how it perform with the problem set.

Environments: The experiment uses 2 problem set (environments) one with grid world and one without grid to understand MDP with different algorithms and its analysis

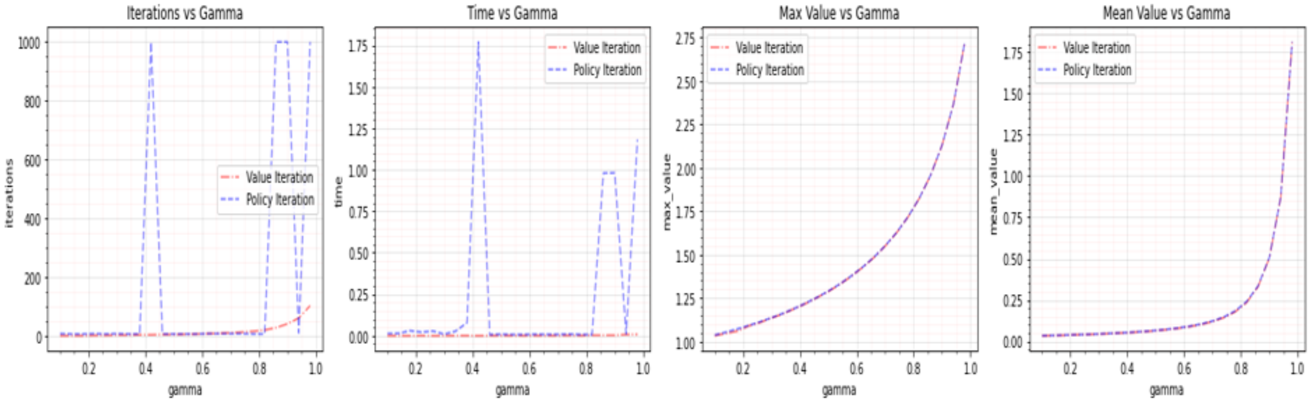
1. **Frozen Lake Problem:** Frozen lake is grid world example from openAI gym python library, where objective is to walk on tiles represented as ice just like   frozen lake. Environment is stochastic and is setup with thickness probability as 95%. Frozen ice have 4 different types of state i..e. start state (S), end goal state (G), tile frozen state (F) and fall in hole (H) set with 4 different actions in directional moves i.e. left, down, right, up represented as 0,1,2,3 respectively. The objective of game environment is to walk from start state to end goal state, without falling into terminate state (hole state).
   Understanding grid problem and solving it stochastically with MDP to find optimal policy can be helpful in resolving grid problems in real time examples like industrial robotics.
2. **Forest Management Problem:** To understand how MDP algorithms perform with non-deterministic non grid example, the experiment uses forest management example from python library mdptoolbox. The example game is played to determine whether to cut forest (Cut action) to gain profit of cutting forest or to wait (Wait action) to maintain old forest for wildlife conservation in non-deterministic state space like forest. The example helps in analyzing not-deterministic state action space using MDP like in real world scenario like autonomous cars, or forest management as specified in example.

Algorithms: Markov Decision Process (MDP) in reinforcement learning (RL) is mathematical model of decision making in which for any discreate or continuous space S, an action A is defined. Agent can be solved by
value iteration (VI), in which optimal policy $\pi(s)$ is calculated with value function. It is iterative algorithm to calculate optimal value $V_{i+1}(s) := \max_a \left\{ \sum_{s'} P_a(s'|s)\,(R_a(s,s') + \gamma V_i(s')) \right\}$
policy iteration (PI) where value is calculated with random selection of policy and optimizing policy iteratively with each selection of value $\pi'(s) := \arg\max_a \left(R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') V_\pi(s') \right)$
Q Learning (QL), in which policies are generated based on Q function values for any given state action pair Q(s,a) and reward is predicted by Q function $Q(s,a): S \times A \rightarrow \mathbb{R}$. It explores to learn optimal policy $\pi(s)$

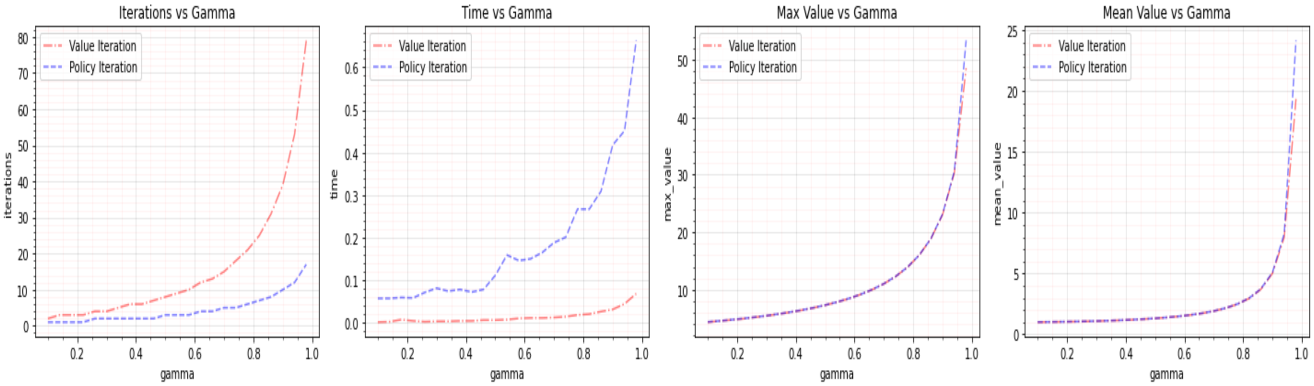
Experiments:

Value Iteration and Policy Iteration discount factor (Gamma) Analysis: To analyze and compare value iteration and policy iteration, 23 discount factors ranging from 0.11 to 0.99 was used with 100000 iterations for both frozen lake and forest management experiments.
Observing discount factor (gamma) with iterations to converge, it is found that for frozen lake although value iteration convergence is very consistent with policy iteration and converging at low value of iteration, but policy iteration convergence seems to be fluctuating to higher convergence iterations at 0.5 gamma, and in between 0.8 and 0.95. Time taken for each gamma is also in line with iteration to converge. Although inconclusive because of variation in policy iteration convergence, it can be observed that value iteration convergence is taking more iterations with increasing gamma compare to policy iterations. Policy iteration tends to take less time compare to value iteration because it carries out

maximizing action at each step. This behavior is observed in forest management example, where value iteration is taking more time to converge over increasing gamma, as compared to policy iteration Observing forest management example, policy iteration is taking less iterations to convergence over the range of gamma (discount factor) as compared to value iteration.



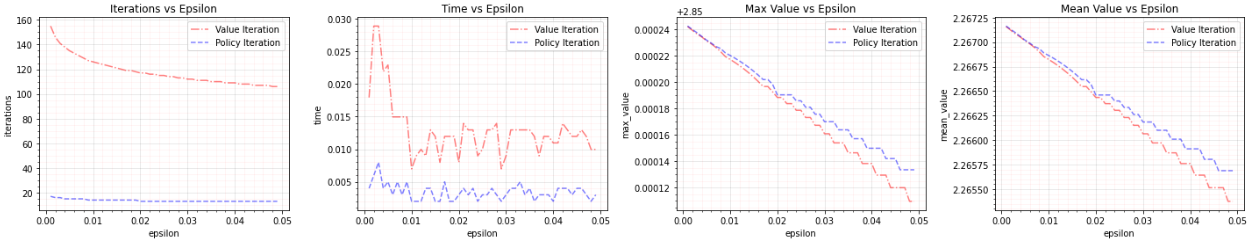
Frozen Lake Gamma Analysis



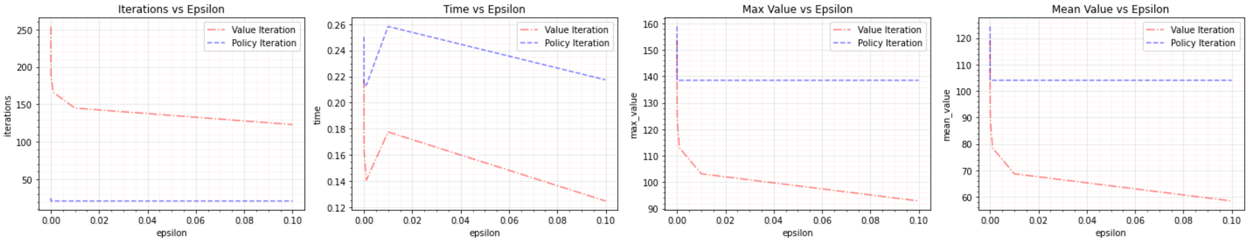
Forest Management Gamma Analysis

Observing optimal and average rewards over gamma, with higher gamma value both algorithms tends to get more rewards and with discount factor. With intuitive observation from graph, it seems that for frozen lake tends to lend more reward for discount factor of 0.99 and both frozen lake and forest management environments.

Value Iteration and Policy Iteration epsilon Analysis: One of method to determine if algorithm is converged and essentially stop running iterations is to observe delta at iteration and compare it with epsilon.



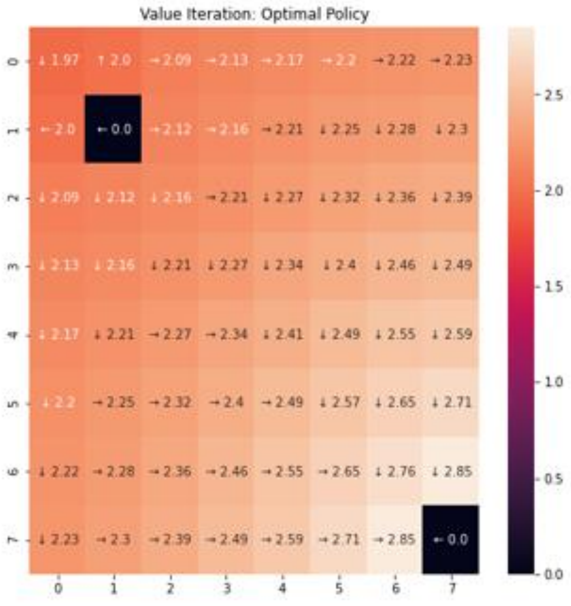
Frozen Lake Epsilon Analysis

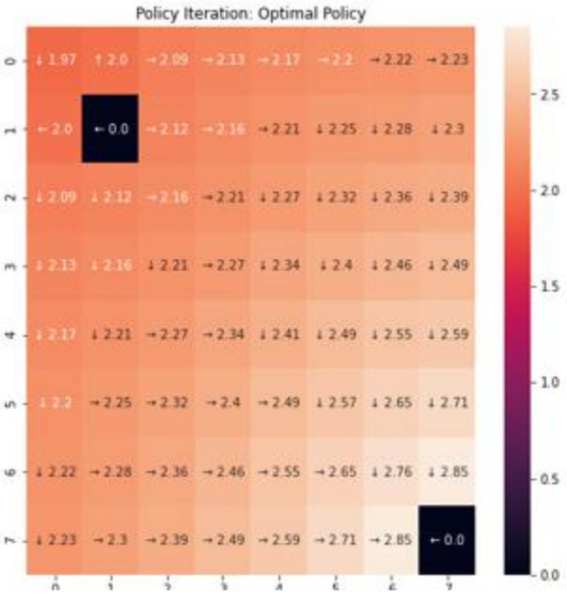


Forest Management Epsilon Analysis

Observing epsilon with converged iterations, it is observed that low value of epsilon takes more steps to converge. Value iteration for both environments is taking more steps to converge, as compare to policy iteration, although policy iteration produces higher optimal value. Higher epsilon value with sun optimal rewards with range of epsilon indicates that higher epsilon tends to produce sub optimal policy. Intuitively optimal epsilon value for frozen-lake is close to 0.01 and 0.0001 for forest management environment.

Optimal Policy for Frozen Lake: With discount factor of 0.99 and epsilon value of 0.01 shows optimal policy for value iteration and policy iteration.



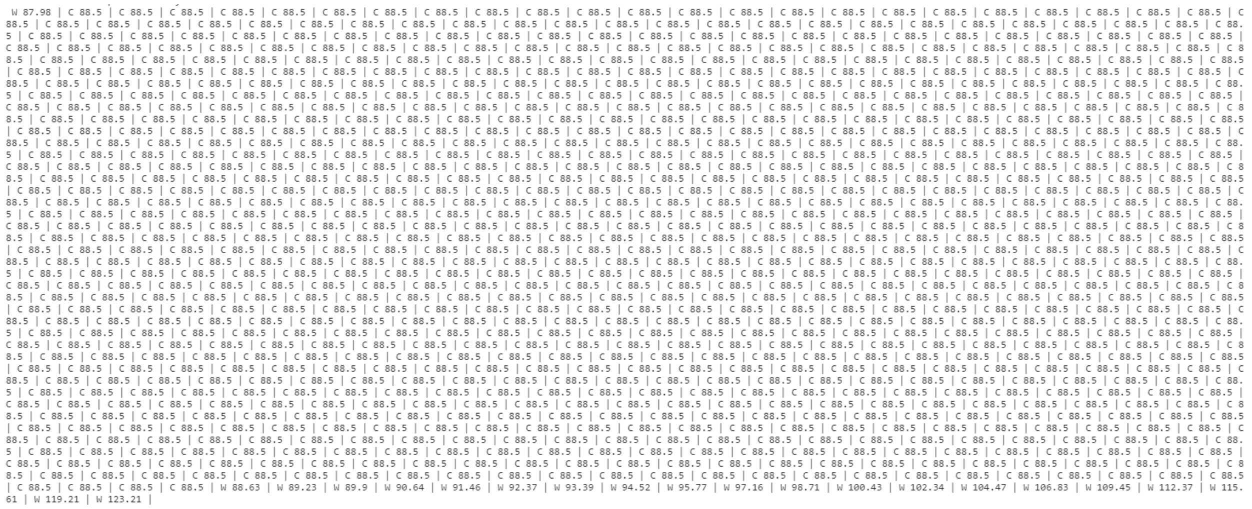


Value Iteration Frozen Lake

Policy Iteration Frozen Lake

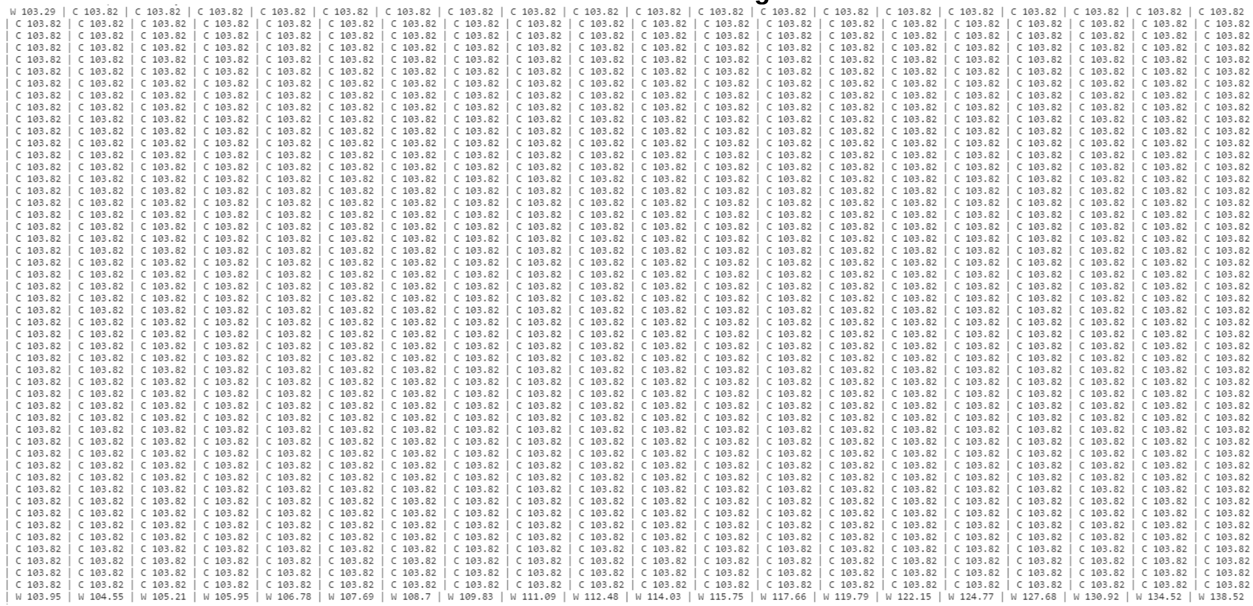
Observing heatmap of optimal value and policy action to with state space of 8X8 grid, it is observed that both value iteration and policy iteration algorithm are able to converge and produce same optimal policy.

Optimal Policy for Frozen Lake: With discount factor of 0.99 and epsilon value of 0.0001 shows optimal policy for value iteration and policy iteration



```
 W 87.98 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C
88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.
5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 |
C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 8
8.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5 | C 88.5
| C 88.5 | C 88.5 | C 88.5 | C 88.5 | W 88.63 | W 89.23 | W 89.9 | W 90.64 | W 91.46 | W 92.37 | W 93.39 | W 94.52 | W 95.77 | W 97.16 | W 98.71 | W 100.43 | W 102.34 | W 104.47 | W 106.83 | W 109.45 | W 112.37 | W 115.
61 | W 119.21 | W 123.21 |
```

Value Iteration Forest Management

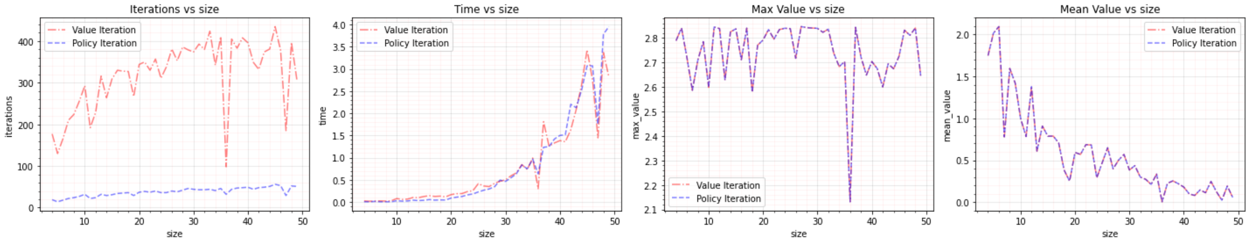


```
 W 103.29 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82 | C 103.82
| W 103.95 | W 104.55 | W 105.21 | W 105.95 | W 106.78 | W 107.69 | W 108.7 | W 109.83 | W 111.09 | W 112.48 | W 114.03 | W 115.75 | W 117.66 | W 119.79 | W 122.15 | W 124.77 | W 127.68 | W 130.92 | W 134.52 | W 138.52
```

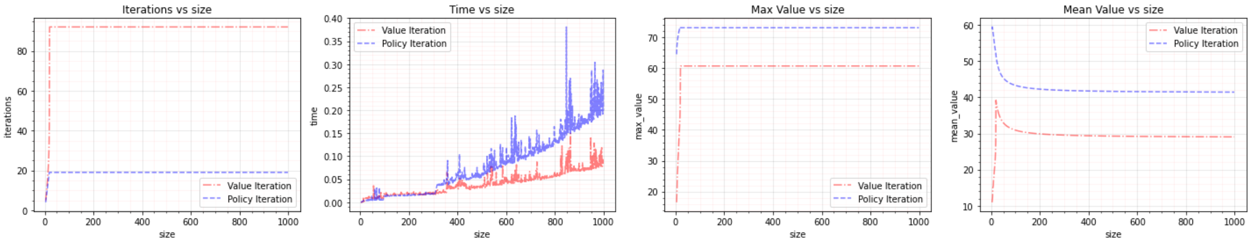
Policy Iteration Forest Management

Observing value iteration and policy iteration for forest management example, although optimal rewards are close to each other they do not match perfectly. This constitutes hyper parameter tunning, as in case of non-grid world like this example state space is more continuous and need more observations of next states.

Value Iteration and Policy Iteration varying state size Analysis: To analyze how value iteration and policy iteration performs in varying degree of states (small states to large states) in both grid (frozen lake) and non-grid world (forest management, the experiment was performed iterating sizes i.e. 4X4 to 50X50 for frozen lake and 100 t to 1000 for forest management environment.


Iterations vs size
Time vs size
Max Value vs size
Mean Value vs size

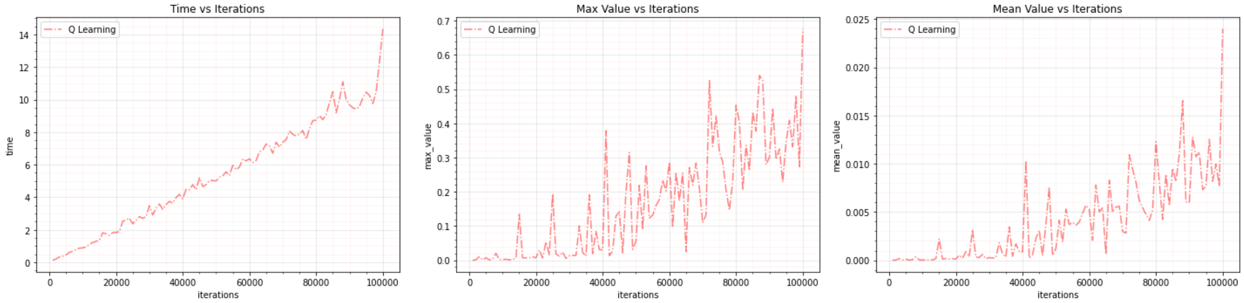
Frozen Lake State Size Analysis


Iterations vs size
Time vs size
Max Value vs size
Mean Value vs size

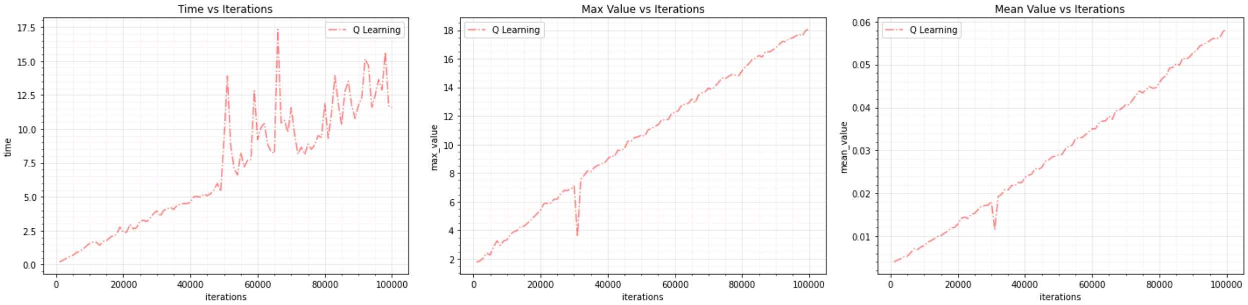
Forest Management State Size Analysis

Observing iterations to converge for each size, it is observed that value iteration takes long time as compared to policy iteration. Time taken for each step is increasing with each step indicates with complexity of increase state, algorithm tends to take more time to converge. Observing rewards both for max (optimal) and average reward, it seems that with deterministic environment like frozen lake, complexity is less with known state action and action, but non deterministic environments (non-grid), complexity is more with more continuous state space and action, like observed in reward values of forest management example. Policy iteration in this example produces more rewards i.e. with non-deterministic environment selecting policy randomly and optimizing it, yields more optimal result.

Q Learning: The experiment carries Q Learning with epsilon-greedy action selection to explore and exploit in state space. In the experiment. Agent takes maximum greedy action with value of epsilon otherwise takes random action with value of 1 – epsilon. Epsilon value of 0.9 and discount factor of 0.99 was used in experiment for both environments.

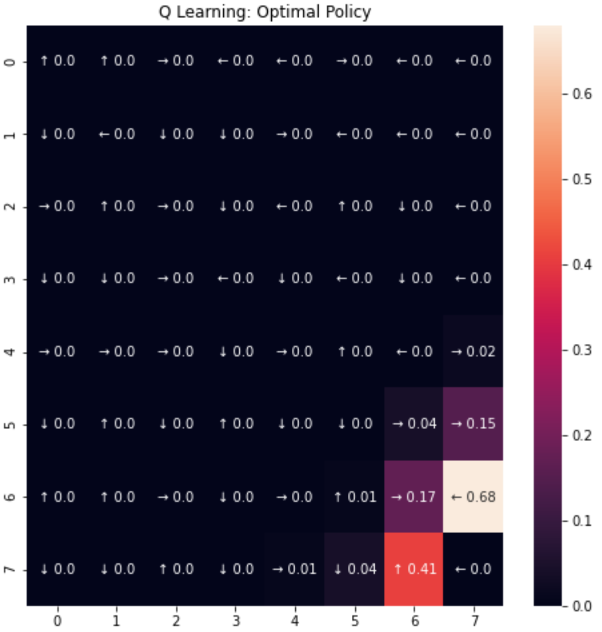

Time vs Iterations
Max Value vs Iterations
Mean Value vs Iterations

Frozen Lake Q Learning with Epsilon Greedy Exploration-Exploitation

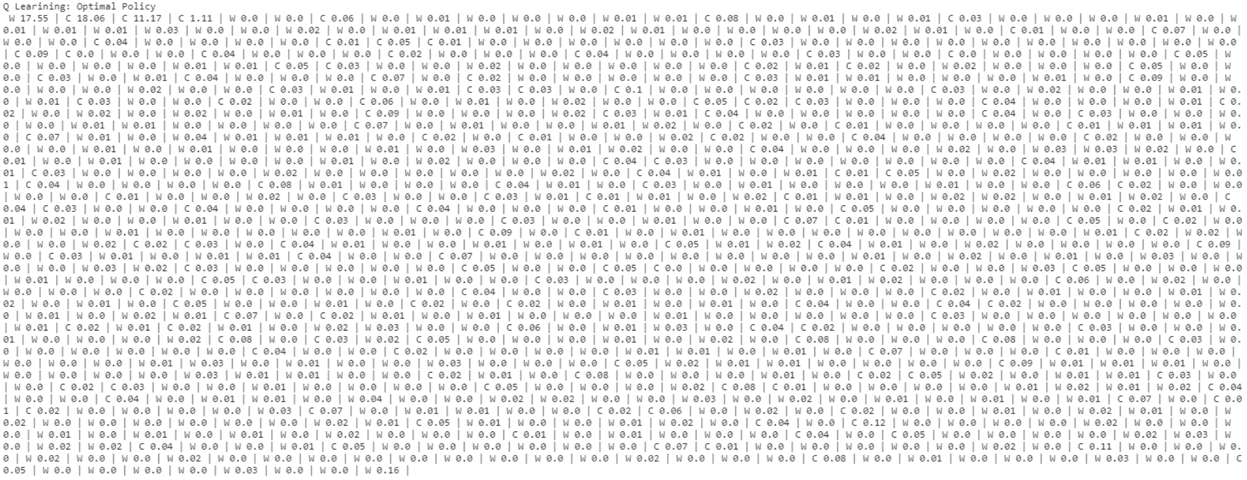

Time vs Iterations
Max Value vs Iterations
Mean Value vs Iterations

Forest Management Q Learning with Epsilon Greedy Exploration-Exploitation

Observing reward values with steps, it is observed that with increasing number of episodes rewards is increasing and converging linearly.



Frozen Lake Q Learning optimal policy



```
Q Learining: Optimal Policy
 W 17.55 | C 18.06 | C 11.17 | C 1.11 | W 0.0 | W 0.0 | C 0.06 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.01 | C 0.08 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W
0.01 | W 0.01 | W 0.01 | W 0.03 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.01 | W 0.01 | W 0.01 | W 0.0 | W 0.02 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.01 | W 0.0 | C 0.01 | W 0.0 | W 0.0 | C 0.07 | W 0.0 |
W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.01 | C 0.05 | C 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0
 | C 0.09 | C 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | C 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | W
0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.01 | C 0.05 | C 0.03 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.01 | C 0.02 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | W 0.0 | W
0.0 | C 0.03 | W 0.0 | W 0.01 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | C 0.07 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.01 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | C 0.09 | W 0.0 | W
0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | C 0.03 | W 0.01 | W 0.0 | W 0.01 | C 0.03 | C 0.03 | W 0.0 | C 0.1 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.01 | W 0.
0 | W 0.01 | C 0.03 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | C 0.06 | W 0.0 | W 0.01 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | C 0.05 | C 0.02 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | C 0.
02 | W 0.0 | W 0.02 | W 0.0 | W 0.02 | W 0.0 | W 0.01 | W 0.0 | C 0.09 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | C 0.03 | W 0.01 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.
0 | W 0.0 | W 0.01 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.07 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.01 | W 0.02 | W 0.0 | C 0.02 | W 0.0 | C 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.01 | W 0.01 | W 0.01 | W 0.
0 | C 0.07 | W 0.01 | W 0.0 | W 0.04 | W 0.01 | W 0.01 | W 0.01 | W 0.0 | C 0.02 | W 0.0 | C 0.01 | W 0.0 | W 0.0 | W 0.02 | C 0.02 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W
0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.03 | W 0.0 | W 0.01 | W 0.02 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.03 | W 0.03 | W 0.02 | W 0.0 | C
0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.01 | W 0.01 | W 0.0 | W 0.
01 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | C 0.04 | W 0.01 | W 0.0 | W 0.01 | C 0.01 | C 0.05 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0
1 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.08 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.01 | W 0.0 | C 0.03 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | C 0.06 | C 0.02 | W 0.0 | W 0.0
 | W 0.0 | W 0.0 | C 0.01 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | C 0.03 | W 0.01 | C 0.01 | W 0.01 | W 0.0 | W 0.02 | C 0.01 | W 0.01 | W 0.0 | W 0.02 | W 0.02 | W 0.0 | W 0.01 | W 0.02 | W 0.0 | C
0.04 | C 0.03 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | W 0.0 | C 0.01 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | C 0.05 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.01 | W 0.
01 | W 0.02 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | C 0.07 | C 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | W 0.0 | C 0.02 | W 0.0
 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | C 0.09 | W 0.0 | C 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | C 0.02 | W 0.02 | W
0.0 | W 0.0 | W 0.02 | C 0.02 | C 0.03 | W 0.0 | C 0.04 | W 0.01 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | C 0.05 | W 0.01 | W 0.02 | C 0.04 | W 0.01 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.09 |
W 0.0 | C 0.03 | W 0.01 | W 0.0 | W 0.01 | W 0.01 | C 0.04 | W 0.0 | W 0.0 | C 0.07 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.02 | W 0.0 | W 0.01 | W 0.0 | W 0.03 | W 0.0 | W
0.0 | W 0.0 | W 0.03 | W 0.02 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | W 0.0 | W 0.0 | C 0.05 | C 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W 0.03 | C 0.05 | W 0.0 | W 0.0 | W 0.0
 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | C 0.03 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.01 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.06 | W 0.0 | W 0.02 | W 0.0 |
W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.01 | W 0.
02 | W 0.0 | W 0.01 | W 0.0 | C 0.05 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | C 0.02 | W 0.0 | C 0.02 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | C 0.04 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.
0 | W 0.01 | W 0.0 | W 0.02 | W 0.01 | C 0.07 | W 0.0 | C 0.02 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0
 | W 0.01 | C 0.02 | W 0.01 | C 0.02 | W 0.01 | W 0.0 | W 0.02 | W 0.03 | W 0.0 | W 0.0 | C 0.06 | W 0.0 | W 0.01 | W 0.03 | W 0.0 | C 0.04 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.0 | W 0.0 | W 0.
01 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | C 0.08 | W 0.0 | C 0.03 | W 0.02 | C 0.05 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.02 | W 0.0 | C 0.08 | W 0.0 | W 0.0 | W 0.0 | C 0.08 | W 0.0 | W 0.0 | W 0.0 | C 0.03 | W 0.
0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | C 0.07 | W 0.0 | W 0.0 | W 0.0 | C 0.01 | W 0.0 | W 0.0 | W 0.0 |
W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.03 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.03 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | W 0.02 | W 0.01 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.09 | W 0.01 | W 0.01 | W 0.01 | W 0.0 |
W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.03 | W 0.01 | W 0.01 | W 0.0 | W 0.0 | C 0.02 | W 0.01 | W 0.0 | C 0.08 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | C 0.02 | C 0.05 | W 0.02 | W 0.0 | W 0.01 | W 0.01 | C 0.03 | W 0.0
 | W 0.0 | C 0.02 | C 0.03 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.05 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | C 0.08 | C 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.01 | W 0.02 | W 0.01 | W 0.02 | C 0.04
 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | W 0.01 | W 0.01 | W 0.0 | W 0.04 | W 0.0 | W 0.0 | W 0.02 | W 0.02 | W 0.0 | W 0.0 | W 0.03 | W 0.0 | W 0.02 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | C 0.07 | W 0.0 | C 0.0
1 | C 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.03 | C 0.07 | W 0.0 | W 0.01 | W 0.01 | W 0.0 | W 0.0 | C 0.02 | C 0.06 | W 0.0 | W 0.02 | W 0.0 | C 0.02 | W 0.0 | W 0.0 | W 0.01 | W 0.0 | W 0.02 | W 0.01 | W 0.0 | W
0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.01 | C 0.05 | W 0.01 | W 0.0 | W 0.0 | W 0.01 | W 0.02 | W 0.0 | C 0.04 | W 0.0 | C 0.12 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W
0.0 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.01 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | C 0.04 | W 0.0 | C 0.05 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.03 | W
0.0 | W 0.02 | W 0.02 | C 0.04 | W 0.0 | W 0.0 | W 0.01 | C 0.05 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | C 0.07 | C 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | C 0.11 | W 0.0 | W 0.0 | W 0.
0 | W 0.02 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.02 | W 0.0 | W 0.0 | W 0.0 | C 0.08 | W 0.0 | W 0.01 | W 0.0 | W 0.0 | W 0.0 | W 0.03 | W 0.0 | W 0.0 | C
0.05 | W 0.0 | W 0.0 | W 0.0 | W 0.0 | W 0.03 | W 0.0 | W 0.0 | W 0.16 |
```

Forest Management Q Learning optimal policy

Comparing optimal policy of Q Learning for both environments with value iteration and policy iteration of same environment, it is observed that they do not match and varies with substantial degree. This is because Q Learning is very sensitive to hyper parameter tunning and need more exploration to fine tune hyperparameters like discount rate or epsilon values.

Conclusion: Overall project experience was unique and interesting. Although I got good understanding the concept of RL and MDP with this exercise, I'm still not done with learning more on MDP and RL. Experiments are very well design and helps in converging concept very well, but for me I still need more analysis to get converged well with the concept. I intent to study and explore more on different problem environments, possibility to understand more on how hyper parameters to be tunned in Q Learning algorithm, I felt short in observing optimal policy for Q Learning and match up with value iteration and policy iteration, and would more time in experimenting and tunning.

.

References:

- Tim Mitchell: book – Machine Learning
- Dr. Charles Isabell and Dr. Michael Littman: video lectures supervised learning chapter RL1 – SL2, UL1 – UL4
- openAI library and mdptoolbox examples and examples of MDP
- TA's during office hour and piazza forum