Assignment 3 Malachi Eberly

1.) For this MRP, the value function represents how "good" it is for the agent to travel to a particular island. More specifically, it determines which islands the agent should travel to to maximize its cummulative reward. The treasure islands from the last assignment have a reward of 2 instead of -1 just for traveling to them. The final island has a reward of 15 for reaching, and because the discount factor is still 0.95, it will be important for the agent to reach the terminal island quickly.

The value function can be solved for using either the iterative or close form solution of the Bellman equation. For this assignment, since the MDP is small and because numpy has helpful tools for linear algebra, I decided to use the close form solution:

$$V = (I - \gamma P)^{-1} R$$

where V is a vector of the values, I is the identity matrix, γ is the discount factor, P is the probability matrix, and R is a vector of the rewards for taking the action of travelling to that island. We've eliminated the "dig" action, so the only possible action is to travel from one island to another.

We can set this equation up as such:

$$I = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

$$\gamma = 0.95$$

$$P = \begin{bmatrix}
0 & 0.5 & 0.2 & 0.2 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.2 & 0.2 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.1 & 0 & 0 & 0.8 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0.1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.3 & 0.2 & 0.3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0.6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & 0 \\
0 & 0 & 0 & 0 & 0.3 & 0 & 0.3 & 0 & 0.2 & 0 & 0 & 0.2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0
\end{bmatrix}$$

$$R = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ 2 \\ -1 \\ 2 \\ -1 \\ -1 \\ 2 \\ -1 \\ 15 \end{bmatrix}$$

Key Observations:

- I is the 12x12 identity matrix
- P is the same from the last assignment
- $(I-\gamma P)^{-1}$ is the inverse of $I-\gamma P$
- R is -1 for s_1, the starting island, and 15 for s_{12}, the terminating island.
- s_5, s_7, and s_{10} have a reward of 2 instead of -1 because that's where the treasure was.

When we solve for V, we get:

$$V = \begin{bmatrix} 189.76 \\ 196.68 \\ 200.26 \\ 205.80 \\ 212.52 \\ 211.45 \\ 221.74 \\ 217.69 \\ 244.53 \\ 216.91 \\ 226.22 \\ 300.00 \end{bmatrix}$$

where all values have been rounded to 2 decimal places.

Based on these values, we can pick the optimal set of actions:

① Move from s_1 to s_5 : 212.52

↳ $\max(196.68, 200.26, 205.80, 212.52)$

② Move from s_5 to s_{12} : 300.00

↳ $\max(205.80, 221.74, 300.00)$

This results in a cummulative reward of

$$G_t = R_{t+1} + \gamma R_{t+2} =$$

$$2 + (0.95 \cdot 15) = \boxed{16.25}$$

Therefore, the MDP is solved given this optimal policy.