

Homework 12, due December 1st, 11:59pm

November 24, 2023

1. The file `map_24x32.csv` from Canvas is a 24×32 map of an environment, with walls marked as negative numbers and a reward marked as 1. The entries with non-negative values are the valid locations (states) for the agent position. The agent's actions are to move from a valid location to a valid adjacent location (left, right, up or down).
 - a) Implement the value iteration algorithm for learning the optimal policy for this map, with $\gamma = 0.9$. Initialize $V(s)$ with the given map values at all valid locations. Run the value iteration for a maximum of 50 epochs of going through all states $s \in S$. Display the initial $V(s)$ and the obtained $V(s)$ every 5 epochs, as grayscale images of size 24×32 . (5 points)
 - b) Display the final learned policy as a table with the actions at each valid location encoded as L,R,U,D as left, right, up and down respectively. (4 points)