1. Convergence method – Policy iteration
2. We have an initial set of transition probabilities. Iteratively we update each probability using below equation,



1. We calculate the weighted average of the reward by multiplying each possible reward by its probability and adding them together.
2. In the GridworldTask, each state has 4 possible next states: 1 for each action: up, down, right, or left. Thus, the first term is the same in each epoch.
3. Next, we calculate the discounted reward and add with expected reward.
4. Iterate the above steps, till we get same transition probabilities for two consecutive occurrences suggesting a convergence.
5. I applied this approach as this was discussed in the class.
6. Convergence method – Value iteration
7. We have an initial set of transition probabilities. Iteratively we update each probability using below equation,

Text

Description automatically generated

1. The value of the terminal states will remain 0 because it is the desired state.
2. V(S) for each cell is calculated by taking the maximum of each possible Q(a). If the agent hits the wall, it still gets the reward and stays in the same state.
3. Iterate the above steps, till we get same transition probabilities for two consecutive occurrences suggesting a convergence.
4. I applied this approach as this was discussed in the class.