CS747 Programming Assignment 2: MDPs

Report

Task 1: Implementing various algorithms to solve the MDP planning problem

- Taking the MDP as input: I didn't need the start and end states so I just ignored them. I also didn't use mdptype.
- Solving Bellman equations: I formulated the system of linear equations as `Ax = b`, where `A` is a matrix (2D `NumPy` array), and `x` and `b` are vectors (1D `NumPy` arrays). Then I used `np.linalg.solve` to get `V^pi`.
- Stopping Condition for Value Iteration: max-norm of the difference of the old and new value functions falling below `1e-10`.
- Howard's Policy Iteration: At every improvable state, I picked the action with the highest action-value function to choose the new policy.

Task 2: Applying MDP planning to solve a maze

- Formulating the maze as an MDP
    - S: {0,1,...n} , one state per unobstructed tile (grid value != 1). Assigning numbers to blocks is done by ordering them lexicographically ased on their grid coordinates (x,y)
    - A: {0,1,2,3} where the numbers correspond to S, E, N, W respectively.
    - T: There are no outgoing transitions from the end state (grid value = 3). From every other state, there are transitions (with probability 1) in those directions where the tile is unobstructed (i.e. the tile corresponds to a state of the MDP) but is not a start state (so grid value = 0 or 3). The action of these transitions depends on the direction.
    - R: Transitions into the end state (grid value = 3) have reward 1. Rest all transitions (into states with grid value 0) have reward 0.
    - Gamma: 0.9
    - Mdptype: My formulation is not episodic. Consider two states which are beside each other in the maze grid. If we choose a policy where the action at the left state is E and the action at the right state is W, then there will be no path to the end state from these two states for this policy. Either way, I don't use the variable mdptype, but this is why I could not take gamma = 1.
- Algorithm of choice: Value Iteration works best (fastest) for me in this task. So I request you to use that for checking.