Homework #1

(due: Oct. 31)

Show all the work with neat writing.

(P1) In this exercise, we will apply value iteration and policy iteration to solve the fixed maze example (`maze-sample-10x10-v0`) provided by the TA.

(a) Define the dynamic system model, including the state x_k, control input u_k, and the state equation. Also, define a cost function to solve the problem. (Disregard the reward returned by the Maze environment, and compute the cost function you defined instead.)

(b) Use value iteration to find a solution. Start with the initial value $V_0 = 0$. Plot the cost function at each step of the value iteration and discuss whether the value iteration converges.

cost function or vk error?

(c) Use policy iteration to find a solution. Begin with the initial policy $\pi_0 = 0$. Plot the cost function at each step of the policy iteration and discuss whether the policy iteration converges.

do you mean \pi(s)=0\forall s\in \mathcal{S} where 0 is a mapping to one of ["N","S","E","W"]?

(d) Now, repeat (b) and (c) with different initial values V_0 and π_0. Suppose we know the optimal path for the maze problem. Can you devise a way to embed this path into V_0 and π_0?

can you elaborate more by what you mean by 'to embed'?