Homework 7

Problem

Continuous MDP

A continuous MDP is an MDP whose state space is continuous. For example, if we are modeling an agent whose position can be at any point on the earth, one possible model is to consider the state space to be all points (x, y) where x are possible latitudes and y are possible longitudes.

In this problem, your goal will be to find the optimal policy for a continuous MDP in which an agent is attempting to traverse a path modeled by the interval $[0, 1]$. The interval is divided into subintervals that consist of different types of terrain. For example, there could be regular ground in the interval $[0, 1/3)$, icy ground in the interval $[1/3, 2/3)$, and sticky ground in the interval $[2/3, 1]$.

The agent moves in discrete time steps, beginning at $x = 0$. $x = 1$ is an absorbing state. The agent receives a reward of $+10000$ for reaching $x = 1$ and a reward of -1 for each action that does not end at $x = 1$.

The agent has different actions it can take, and these actions have different effectiveness in different terrain types. For example, if the agent can walk, run, or jump, perhaps walking on regular ground is slower than running on regular ground, but walking, running, and jumping are all slower than that on sticky ground. And perhaps there's more variance in the outcome when taking an action on icy ground (because of its slipperiness).

For this problem, we will consider the continuous MDP described below:

We will give you a continuous MDP problem here.

$T(i, j)$ is the transition function for the MDP from a point of terrain type i using action j. It is a Gaussian probability distribution for the amount of movement that occurs,with mean given by $movementMean[i][j]$ and standard deviation given by $movementSD[i][j]$. (A positive value means movement to the right and negative means movement to the left, though the agent will stop at $x = 0$ if it tries to move to the left of $x = 0$, and it will stop at $x = 1$ if it tries to move to the right of $x = 1$.)

Your goal for this problem is to determine the optimal policy for each state in the MDP. Because the MDP is continuous, your policy will consist of a partition of $[0, 1)$ into subintervals (which may or may not be the same as the terrain intervals) along with which action is optimal for points in those subintervals.