AI ASSIGNMENT #8: Markov decision processes & value iteration

CSCI 3385/5385/6385

Submissions received after **11:59 pm on Friday, November 14** will be considered late.

This assignment is worth 6 points.

Instructions

1. Write a python program that solves the MDP described below using the value iteration algorithm.
    a. Start by using `MDP_only.py` to see how well you understand the algorithm
    b. Once you have given it a try on your own, you are welcome to use the code provided in `value_iteration_starter_code.py` which has two sections of code missing for you to finish (which you can find by searching for “Your code here”):
        i. The utility update portion inside the state-loop of the function `value_iteration()`
        ii. The calculation of state-action utility inside the function `Q()`
    c. Record your solution and add it to your report (more on that below)
2. Change the following and record the solutions you get in your report along with what change led to each solution:
    a. **gamma** - try at least three values (small, medium, large), but don’t set it to exactly 1, as this will cause the utilities to go to infinity and never converge

b. **reward function** - try to see if you can cause a change in optimal policy by adjusting the transition rewards, but without changing their ordering (i.e., the reward for transitioning to Overheated should always be the minimum).

c. **transition model** - try to see if you can cause a change in optimal policy by adjusting the probabilities in the transition model. But be sure that each case's distribution sums to 1.

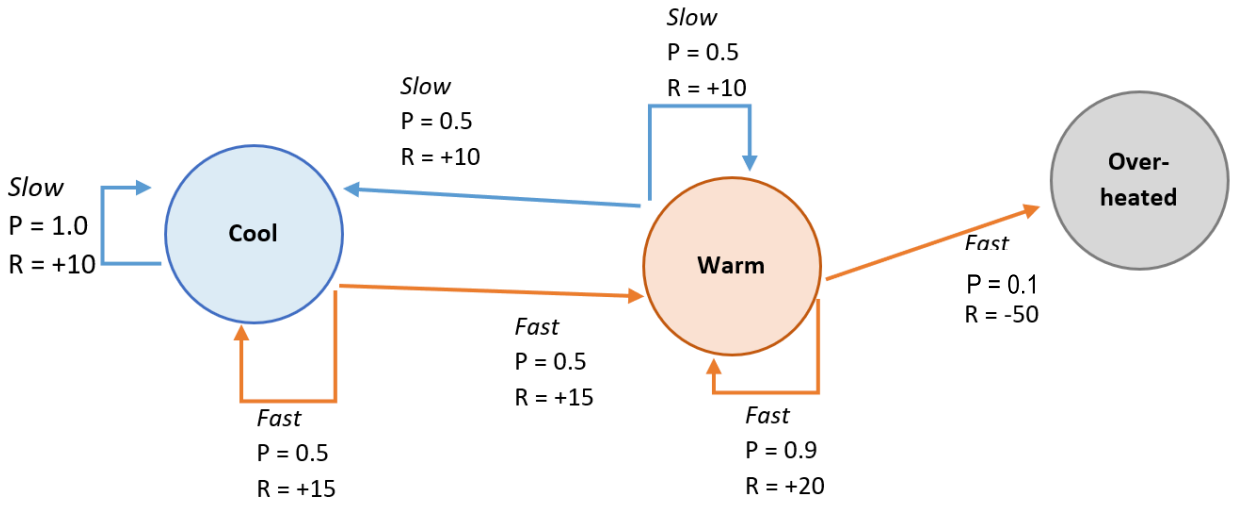
3. Submit the following to Blackboard:

   a. Your program, named `lastname_firstname_a8.py`. Submissions of any other file type will result in a significant penalty to your grade.

   b. A report named `lastname_firstname_a8.{pdf, docx}` that has the solution for the original MDP, solutions for the MDP after doing the changes in step 2, and any thoughts or observations you made while doing this.

4. Double-check your Blackboard submission to make sure you uploaded both files.

The racecar MDP



States: {Cool, Warm, Overheated}

Actions: {Slow, Fast}

Transition probabilities:

P(Cool | Cool, Slow) = 1.0

P(Cool | Cool, Fast) = 0.5, P(Warm | Cool, Fast) = 0.5

P(Cool | Warm, Slow) = 0.5, P(Warm | Warm, Slow) = 0.5

P(Warm | Warm, Fast) = 0.9, P(Overheated | Warm, Fast) = 0.1

Reward function:

R(Cool, Slow, Cool) = +10

R(Cool, Fast, Cool) = +15

R(Cool, Fast, Warm) = +15

R(Warm, Slow, Cool) = +10

R(Warm, Slow, Warm) = +10

R(Warm, Fast, Warm) = +20

R(Warm, Fast, Overheated) = -50