CS560 Spring 2020

Assignment 2: Reinforcement Learning`

Due date: April 3, 11:59:59pm.

The goal for this assignment is for you to try out Reinforcement Learning and solving Markov Decision Processes problems. You will be using a modified version of the golf problem that was given on the Midterm and will want to use reinforcement learning to determine the optimal policy for each of the states. You need to implement both Model-Based and Model-Free Active learning.

Here is a revised description of the golf problem from the midterm. There are seven states (***Fairway***, ***Ravine***, ***Close*** *to the Pin*, ***Same*** *Level as the Pin*, ***Left*** *of the Pin*, ***Over*** *the Green* and ***In*** *the Hole*). If you are on the Fairway or in the Ravine then you can either shoot ***At*** *the Pin*, ***Past*** *the Pin* or ***Left*** *of Pin*. If you are over the green you can either ***Chip*** or ***Pitch*** the ball toward the hole. You can only ***Putt*** if you are anywhere on the green

You will be given a file that contains a set of outcomes for each state/action you will be given a probability of ending up in a particular state. For example there will be multiple lines of the form:

Fairway/At/Close/.25

Fairway/At/Same/.35

Fairway/At/Ravine/.15

Fairway/At/Left/.10

Fairway/At/Over/.15

The first line states that there is a 25% chance that you will be Close to the Pin if you shoot At the Pin from the Fairway.

Yes, you could use those values to directly solve this Markov Decision Process problem but that would defeat the purpose of this assignment. Instead you should use these probabilities to randomly select a resulting position during your learning process and then deduce the transition probabilities or utility values. You would then use those resulting values to determine the optimal policy.

Given that you are implementing Active Learning, you should try different values for exploration versus exploitation and see how that may change the resulting policy (if at all). You will also need to decide when you want to stop the learning process, as given the file you could learn forever.

Again your program should be able to be run directly from a shell, and should take a file from standard input that describes those probabilities. For each type of learning you should print out the values that you have calculated along with the resulting policy. So for Model-Based learning you would print out the transition probabilities for each state/action/state triplet and for Model-Free learning you would print out the utility values for each state/action/state triplet.

When you use Model-Based learning to solve the Markov Decision Process you should experiment with different values for the discount value as well as epsilon (the value that controls when to stop the iterations) and be able to describe how those values affected the outcome of the policy.

So in the writeup for your assignment you should answer the following:

1. How did changing the initial value for exploration (starting close to 1 versus starting closer to 0) change the resulting computed policy?
2. How did you decide when to stop learning?
3. How did changing the discount value (starting close to 1 versus starting closer to 0) change the resulting policy?
4. How did changing the value for epsilon change the resulting policy?

Please do not wait until the last couple of days if you are having a problem with this assignment. We are happy to help keep you moving in the right direction but can't do that if you don't ask for help.

Submission Instructions

You should submit this homework by inviting me (ajpalay-unc) to your GitHub repository.

Your submission must include:

1. The source code. The Code should be well commented for us to understand how it works.
2. A README file indicating how to compile and run your code (again the your program must be able to be run directly from the shell).
3. A report in PDF format called HW1.pdf with the names of all group members indicated on top. Along with answering the above questions, the report should describe your implemented solutions in a way that it makes it possible to understand your solution without having to look through your source code.The report needs to include a brief statement of individual contribution, i.e., which group member(s) was/were responsible for which parts of the solution and submitted material.

Late policy: Assignments are due at 11:59pm on the assigned date. Let assignments will not get full credit. The sooner you get them in the less points you will lose.

Academic integrity: All code and written responses for assignments should be original within your group. To protect the integrity of the course, we will be actively checking for code plagiarism (both from current classmates and the internet).