Homework 10
Reinforcement Learning, Boosting, and Weighted Majority

CMU 10-601: Machine Learning (Fall 2017)
https://piazza.com/cmu/fall2017/10601b/
OUT: November 27, 2017
DUE: December 4, 2017 11:59 PM

START HERE: Instructions

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Jane explained to me what is asked in Question 2.1"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only. See the Academic Integrity Section on piazza for more information: https://piazza.com/cmu/fall2017/10601b/home
- **Late Submission Policy:** See the late submission policy here: https://piazza.com/cmu/fall2017/10601b/home
- **Submitting your work:**
  - **Canvas:** We will use an online system called Canvas for short answer and multiple choice questions. You can log in with your Andrew ID and password. (As a reminder, never enter your Andrew password into any website unless you have first checked that the URL starts with "https://" and the domain name ends in ".cmu.edu" – but in this case it's OK since both conditions are met). You may only **submit once** on canvas, so be sure of your answers before you submit. However, canvas allows you to work on your answers and then close out of the page and it will save your progress. You will not be granted additional submissions, so please be confident of your solutions when you are submitting your assignment.

Homework 10 will take place entirely on canvas: https://canvas.cmu.edu/courses/2650/assignments.

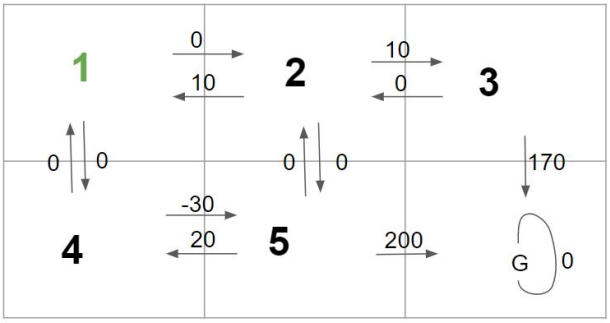
1 Reinforcement Learning

Consider the following Markov Decision Process (MDP), describing a simple robot grid world. Here, the arrows indicate legal actions from each state, and they result in transporting the robot to the adjacent state. The value of the immediate reward $r(s, a)$ which is obtained by taking that action a from that state s is written next to the arrow. Suppose the robot is initially in state 1 and needs to find a path to reach G. Your job is to come up with the an optimal path that the robot can follow. Here, $V^*(s)$ represents value function, and discount factor $\gamma = 0.9$.

Recall that for worlds like this where actions have deterministic outcomes, $V^*(s)$ refers to the value of each state, which can be written:

$$V^*(s) = r(s, a) + \gamma V^*(s')$$

where $r(s, a)$ denotes the immediate reward received when taking action a from state s, and where action a is the optimal action (that is, $a = \pi^*(s)$ is the action recommended by the optimal policy π^*). Here, s' denotes the new state of the robot after taking action a.



Answer the following questions to help the robot achieve its goal :

Q1. **[3.5 pts] What is the $V^*(s)$ value for state 5 ?**

Q2. **[3.5 pts] What is the $V^*(s)$ value for state 3 ?**

Q3. **[3.5 pts] What is the $V^*(s)$ value for box 2 ?**

Q4. **[3.5 pts] What is the $V^*(s)$ value for state 4 ?**

Q5. **[3.5 pts] What is the $V^*(s)$ value for state 1 ?**

Q6. **[3.5 pts] Which one of the following can be an optimal policy for the robot?**

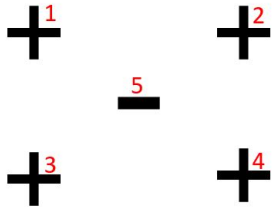
(a) 1-4-5-G

(b) 1-2-5-G

(c) 1-2-5-4-1-2-5-G

2 Boosting

Consider training a boosting classifier using decision stumps on the following data set:



Q1. **[3 pts] Which examples will have their weights increased at the end of the first iteration?**

(a) 1

(b) 2

(c) 3

(d) 4

(e) 5

Q2. **[3 pts] How many iterations will it take to achieve zero training error?**

(a) at least 3

(b) at least 2

(c) at least 1

Q3. **[3 pts] Why do we want to use weak learners when boosting?**

(a) To prevent bias

(b) It doesn't matter whether we choose weak learners or not

(c) To prevent overfitting

3 Weighted Majority Algorithm

Suppose we have results from a study conducted in Amazon Mechanical Turk (a crowdsourcing of data to make predictions). In the worker pool we have all sorts of people. We have experts who have taken Introduction to Machine Learning, some statisticians, and the rest of the people. We have all these people giving predictions for our binary problem. We would want to find a perfect expert among this crowd, but there may not be such perfect expert.Therefore we use the weighted majority algorithm to introduce weights for each person. Let us consider we have ten such workers and a binary classification problem.

Ten Expert Predictions	Ground Truth
$1, 1, 1, 1, 0, 0, 0, 1, 1, 0$	1
$1, 0, 1, 0, 1, 0, 1, 1, 0, 1$	0
$1, 1, 1, 1, 1, 1, 1, 0, 1, 0$	1
$1, 0, 1, 1, 0, 1, 0, 1, 0, 0$	0

Consider we use a learning rate(β) of 0.5 and if a_i denotes the i^{th} pool worker and w_i denotes the weight associated with a_i, q_0 and q_1 are the the cumulative weights for decision 0 and 1. Answer these questions.

Q1. **[3 pts] What will be the value of q_0 on the first example?**

(a) 0.4

(b) 0.6

(c) 3

(d) 4

(e) 6

Q2. **[3 pts] What would the prediction be for the first example?**

(a) 1

(b) 0

(c) Cannot determine with the given data.

Q3. **[3 pts] What are the updated weight w_1 and w_{10} after training on the first example?**

(a) 1 and 1

(b) 1 and 0

(c) 0.5 and 1

(d) 1 and 0.5

Q4. **[3 pts] Train the algorithm for all the given examples and report the value of w_9?**

(a) 0.0625

(b) 0.125

(c) 0.25

(d) 0.5

(e) 1

Q5. **[3 pts] After training the Weighted Majority Algorithm on the above four training examples, the experts are given a fifth example. Their individual predictions are $1, 0, 1, 0, 0, 1, 0, 1, 1, 0$. What is the prediction output by the Weighted Majority Algorithm?**

(a) 1

(b) 0

(c) Either.

Q6. **[3 pts] We happen to know that at least one of the ten experts is perfect (they never make an incorrect prediction), but we don't which of the ten they are. What value of β should we use, to guarantee that at any point during training, the weighted majority algorithm output will be influenced only by experts that have been infallible so far?**

(a) 1

(b) 0

(c) 0.5.