CptS 570 Machine Learning, Fall 2019
Homework #4
Due Date: Tue, Nov 30 midnight via Blackboard

NOTE 1: Please use a word processing software (e.g., Microsoft word or Latex) to write your answers and submit a printed copy to me at the beginning of class on Oct 23. The rationale is that it is sometimes hard to read and understand the hand-written answers.

NOTE 2: Please ensure that all the graphs are appropriately labeled (x-axis, y-axis, and each curve). The caption or heading of each graph should be informative and self-contained.

Analytical Part (3 Percent + 1 Percent extra credit)

1. **(Finite-Horizon MDPs.)** Our basic definition of an MDP in class defined the reward function $R(s)$ to be a function of just the state, which we will call a *state reward function*. It is also common to define a reward function to be a function of the state and action, written as $R(s, a)$, which we will call a *state-action reward function*. The meaning is that the agent gets a reward of $R(s, a)$ when they take action a in state s. While this may seem to be a significant difference, it does not fundamentally extend our modeling power, nor does it fundamentally change the algorithms that we have developed.

   a) Describe a real world problem where the corresponding MDP is more naturally modeled using a state-action reward function compared to using a state reward function.

   b) Modify the Finite-horizon value iteration algorithm so that it works for state-action reward functions. Do this by writing out the new update equation that is used in each iteration and explaining the modification from the equation given in class for state rewards.

   c) Any MDP with a state-action reward function can be transformed into an "equivalent" MDP with just a state reward function. Show how any MDP with a state-action reward function $R(s, a)$ can be transformed into a different MDP with state reward function $R(s)$, such that the optimal policies in the new MDP correspond exactly to the optimal policies in the original MDP. That is an optimal policy in the new MDP can be mapped to an optimal policy in the original MDP. *Hint: It will be necessary for the new MDP to introduce new "book keeping" states that are not in the original MDP.*

2. **(k-th Order MDPs.)** A standard MDP is described by a set of states S, a set of actions A, a transition function T, and a reward function R. Where $T(s, a, s')$ gives the probability of transitioning to s' after taking action a in state s, and $R(s)$ gives the immediate reward of being in state s.

   A k-order MDP is described in the same way with one exception. The transition function T depends on the current state s and also the previous $k-1$ states. That is, $T(s_{k-1}, \cdots, s_1, s, a, s')$ $= Pr(s'|a, s, s_1, \cdots, s_{k-1})$ gives the probability of transitioning to state s' given that action a was taken in state s and the previous $k-1$ states were $(s_{k-1}, \cdots, s_1)$.

   Given a k-order MDP $M = (S, A, T, R)$ describe how to construct a standard (First-order) MDP $M' = (S', A', T', R')$ that is equivalent to M. Here equivalent means that a solution to M' can be easily converted into a solution to M. Be sure to describe S', A', T', and R'. Give a brief justification for your construction.

3. Some MDP formulations use a reward function $R(s, a)$ that depends on the action taken in a state or a reward function $R(s, a, s')$ that also depends on the result state s' (we get reward $R(s, a, s')$ when we take action a in state s and then transition to s'). Write the Bellman optimality equation with discount factor β for each of these two formulations.

4. Consider a trivially simple MDP with two states $S = \{s_0, s_1\}$ and a single action $A = \{a\}$. The reward function is $R(s_0) = 0$ and $R(s_1) = 1$. The transition function is $T(s_0, a, s_1) = 1$ and $T(s_1, a, s_1) = 1$. Note that there is only a single policy π for this MDP that takes action a in both states.

a) Using a discount factor $\beta = 1$ (i.e. no discounting), write out the linear equations for evaluating the policy and attempt to solve the linear system. What happens and why?

b) Repeat the previous question using a discount factor of $\beta = 0.9$.

5. Please read the following paper and briefly summarize (at most one page) the key ideas as you understood:

Thomas G. Dietterich (2000). Ensemble Methods in Machine Learning. J. Kittler and F. Roli (Ed.) First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science (pp. 1-15). New York: Springer Verlag.
http://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf

6. Please read the following paper and write a brief summary (at most one page) of the main points.

Matthew Zook, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Pea Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara Knig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, Frank Pasquale: Ten simple rules for responsible big data research. PLoS Computational Biology 13(3) (2017)
https://www.microsoft.com/en-us/research/wp-content/uploads/2017/10/journal.pcbi_.1005399.pdf

7. Please read the following paper and write a brief summary (at most one page) of the main points.

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Franois Crespo, Dan Dennison: Hidden Technical Debt in Machine Learning Systems. NIPS 2015: 2503-2511

8. Please read the following paper and write a brief summary (at most one page) of the main points.

Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, D. Sculley: The ML test score: A rubric for ML production readiness and technical debt reduction. BigData 2017: 1123-1132

9. (Extra credit) Please go through the excellent talk given by Kate Crawford at NIPS-2017 Conference on the topic of "Bias in Data Analysis" and write a brief summary (at most one page) of the main points.

Kate Crawford: The Trouble with Bias. Invited Talk at the NIPS Conference, 2017. Video:
https://www.youtube.com/watch?v=fMym_BKWQzk

10. (Extra credit) Please go through the following program on societal impacts of AI and write a brief summary (at most one page) of the main points.

Video: https://www.pbs.org/wgbh/frontline/film/in-the-age-of-ai/

Grading Rubric

Each question in the students work will be assigned a letter grade of either A,B,C,D, or F by the Instructor and TAs. This five-point (discrete) scale is described as follows:

- **A) Exemplary (=100%).**
  Solution presented solves the problem stated correctly and meets all requirements of the problem.
  Solution is clearly presented.
  Assumptions made are reasonable and are explicitly stated in the solution.
  Solution represents an elegant and effective way to solve the problem and is not overly complicated than is necessary.

- **B) Capable (=75%).**
  Solution is mostly correct, satisfying most of the above criteria under the exemplary category, but contains some minor pitfalls, errors/flaws or limitations.

- **C) Needs Improvement (=50%).**
  Solution demonstrates a viable approach toward solving the problem but contains some major pitfalls, errors/flaws or limitations.

- **D) Unsatisfactory (=25%)**
  Critical elements of the solution are missing or significantly flawed.
  Solution does not demonstrate sufficient understanding of the problem and/or any reasonable directions to solve the problem.

- **F) Not attempted (=0%)**
  No solution provided.

The points on a given homework question will be equal to the percentage assigned (given by the letter grades shown above) multiplied by the maximum number of possible points worth for that question. For example, if a question is worth 6 points and the answer is awarded a B grade, then that implies 4.5 points out of 6.