

CS5804 Homework 5 Written Problems

Homework must be submitted electronically following the instructions on the course homepage. Make sure to explain your reasoning or show your derivations. Except for answers that are especially straightforward, you will lose points for unjustified answers, even if they are correct.

1. (2 points) Suppose you are running a learning experiment on a new algorithm for Boolean classification. You have a data set consisting of 100 positive and 100 negative examples. You plan to use leave-one-out cross-validation and compare your algorithm to a baseline function: a simple majority classifier. (A majority classifier is given a set of training data and then always outputs the class that is in the majority in the training set, regardless of the input.) You expect the majority classifier to score 50% on leave-one-out cross-validation, but to your surprise, it scores zero every time. Explain why.
2. (6 points) For each of the following scenarios, argue whether batch supervised learning, online supervised learning, or reinforcement learning is more appropriate setting and representation. There are multiple correct answers, but different reasons for each. Be sure to consider practicality, computational cost, and the goals of the task when coming up with your reasoning. Think about how one would obtain labels for the task, if it's possible at all, or whether they would be obtainable in batches or one at a time.
 - (a) Training a robot dog to play fetch.
 - (b) Personalized spam filters for each user of your popular web-based mail service (e.g., Gmail, Outlook).
 - (c) A song-recognition app for smartphones that listen to audio snippets of songs and identify the recording (e.g., Shazam).
 - (d) A recommendation engine for an online marketplace that recommends purchases to users based on their previous purchases.
 - (e) An automatic medical diagnostic tool that predicts diseases based on vital signs, symptoms, and family history.
 - (f) Auto-complete and spell checkers that adapt to users' individual tendencies, e.g., what words they commonly type and what errors they frequently make and correct.
3. (7 points) For each of the major themes we have studied in this course (search, adversarial search, Markov decision processes, reinforcement learning, logic, probability and Bayesian networks, other forms of learning), describe how you could imagine incorporating these themes into a competitive agent for the Pacman capture-the-flag game. Reason about what challenges may arise in using these concepts, such as efficiency, and suggest solutions to address these challenges. Thoughtful, intelligent, and thorough responses to this question will get full credit. If you discuss these ideas with other students, which you are welcome to do, indicate on your submission whom you discussed with.

(Feel free to be creative on this question. You are not obligated to implement any of these ideas for the final project. However, doing this exercise should also help get you started in brainstorming a plan for your own agent design.)