CS 277 (W24): Control and Reinforcement Learning

Quiz 2: Introduction to Control Learning, Imitation Learning

Due date: Monday, January 22, 2024 (Pacific Time)

Roy Fox
https://royf.org/crs/CS277/W24

Instructions: please solve the quiz in the marked spaces and submit this PDF to Gradescope.

Question 1 Control Learning is interesting because (check all that hold):

- □ Control Learning is a good choice for modeling almost any learning problem.
- □ Control Learning can use data to train an agent to make good sequential decisions.
- □ A control-learning agent can learn from very weak supervision.
- □ A control-learning agent can decide how to collect its training data.

Question 2 Control Learning is hard because (check all that hold):

- □ Learning from weak supervision requires large amounts of data.
- □ Data for Control Learning is often scarce, particularly data that provides stronger supervision.
- □ Even with big training data, a deployed agent may need to make decisions in situations never seen in training.
- □ In most Control Learning settings, it is unclear what it means for an agent to be optimal.

Question 3 Check all settings that exhibit train–test mismatch (a.k.a covariate shift):

- □ Training a dog–cat classifier on photos and using it to classify drawings.
- □ Training a self-driving car on expert driver demonstrations and then taking it for a test drive.
- □ Training a goal-conditioned robot policy to arrange colored blocks in a random sample of goal arrangements and then evaluating it on a new goal.
- □ Training a drone via DAgger by repeating the following until convergence: rolling out the drone's current policy, having an expert provide corrections, and training on this new data.

Question 4 Check all that hold in Imitation Learning:

□ If a demonstrator is good but not perfect, BC can also learn a good (but not perfect) policy.

□ If a demonstrator is good but not perfect, a goal-conditioned policy trained with hindsight BC cannot be good because a trajectory leading to s_t may now be a really bad way to reach s_t.

□ It may be impossible, with any amount of data, to successfully imitate a demonstrator with a different state observability (different sensors) than the learner.

□ Both DAgger and DART can overcome inconsistent demonstrations more easily than BC.