Reinforcement Learning: Policy Updates & Policy Gradients Quiz

Ethan B. Mehta
ethanbmehta@berkeley.edu

Sean Lin
seanlin2000@berkeley.edu

Jaiveer Singh
j.singh@berkeley.edu

0.1 Q1

What are States, Actions, and Rewards?

0.2 Q2

How does the Agent in Reinforcement Learning parallel the Control Law from EECS 16B? In what ways are they different?

0.3 Q3

Compute the size of the state space for the game of Monopoly with 2 players, 8 properties in total, 4 corner spaces (non-property locations). Each property can be owned by exactly 1 player or no player. Each player's token can be on one of the properties, or one of the corner spaces.

0.4 Q4

Anant, Jennifer, and Jitendra are talking about Policies. Here are snippets from their conversation:

1. Anant: A Policy is a way of representing how valuable each State is. It maps from State to estimated total Reward incurred from that State through the future.
2. Jennifer: Policies must always be deterministic; it doesn't make sense to have a Policy with any randomness (stochasticity).
3. Jitendra: If I'm dealing with a continuous State space, it makes sense to use a dictionary to store my Policy.

Why are each of their statements wrong?

0.5 Q5

Concisely explain the idea of a Policy Gradient. What formulation of a Policy do we need to apply a Policy Gradient?

0.6 Q6

Why does Deep RL typically use Gradient Ascent, instead of Gradient Descent? (Hint: what is RL seeking to optimize?)

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

0.7 Q7

What is the purpose of the Classification Neural Net in Deep RL methods?

0.8 Q8

Briefly explain each part of A3C's full name: Asynchronous Advantage Actor-Critic.

0.9 Q9

What is one benefit of using OpenAI's Gym environment? Why is it useful for us to use the same environments as other RL researchers and engineers?

0.10 Q10

Briefly describe the state space of the 'CartPole-v0' environment. What is the objective?