Due Date: Dec. 11th, 11:59 pm
Total points available: 100 pts.

Problem 1: Implementation [100 pts.]

In this problem, we try to implement REINFORCE and Actor-Critic in a simple continuous environment InvertedPendulum. Code skeleton is given at here. Please fill in the code and answer the following questions:

1. Comparing the difference between REINFORCE and Actor-Critic. Only need descriptions.

2. When implementing the two algorithms, you can try different structures of neural networks, also, you can use different learning rates, batch sizes, or memory size parameters. In particular, the loss might oscillate and you have to find a good criterium. Choose your parameters with explanations (should include both intuitive and experimental reasons after some attempts), and fix them in the next two problems.

3. Produce a training graph for both methods, where the x-axis indicates the episode and the y-axis indicates the episodic return (discounted cumulative rewards).

4. Comparing the difference in the training process between REINFORCE and Actor-Critic, which algorithm do you think is better?

If you feel hard to complete the task, I recommend you to check REINFORCE and Actor-Critic for reference. Note that they are for discrete environments, and you can refer to PPO Continuous for implementation in continuous environments (the only difference is how to get the action in the actor).