




Assignment 6: Reinforcement Learning

 Discover & Learn


Important Information

	<p>School of Engineering Technology and Applied Science</p> <p>Information and Communication Engineering Technology</p> <p>Unsupervised and Reinforcement Learning (COMP257)</p> <ul style="list-style-type: none">• Reinforcement Learning (15%)• Due Date: Friday of Week 13 by 11:59 pm EST (late penalty at 10 points per day)• Upload your assignment here: Assignment 6: Reinforcement Learning (d2l/common/dialogs/quickLink/quickLink.d2l?ou=694716&type=dropbox&rdoc=CENCOL-3439546)
---	---

Instructions

	<ul style="list-style-type: none">• This assignment requires students to work in teams of two. In an odd head count in the class size, there will be one team with three members.• Your team is free to choose any toolkits to solve the problems at hand (e.g., TensorFlow, Sci-Learn, etc.)• All written reports and codes are to be maintained on a repository of your choice such as Github. The course instructor will discuss and exchange with you information to get access to your code.• IMPORTANT NOTES:<ul style="list-style-type: none">◦ 1 point will be deducted for each incident that does not conform to the requirements (e.g., code not properly formatted, comments not relevant to support documentation of code, missing code documentation, etc.).◦ All points will be deducted for submission of <i>nonsensical code</i> (i.e., code that doesn't contribute to the relevancy of the task at hand). This is question-specific.
---	--

Questions

	<ul style="list-style-type: none">• Read the question below carefully.
<p>Question 1</p> <p>[100 points]</p>	<p>Use TF-Agents to train an agent that can successfully land the lunar lander in OpenAI Gym's LunarLander-v2 environment.</p> <ol style="list-style-type: none">1. Create a simple policy network with 4 output neurons (one per possible action) that use the <i>Model</i> class in the <i>keras.models</i>. [15 points]2. Discuss the rationale of the activation functions used in the network. [10 points]3. Discuss the rationale of the loss function used in the network. [10 points]4. Implement a strategy that adjusts the following hyperparameters: (i) the number of iterations, (ii) the number of episodes, (iii) the maximum number of steps, and (iv) the discount factor γ at each step. [50 points]5. Provide the parameters of the worst and best network policy and plot their corresponding <i>Mean reward</i> on the y-axis to <i>Episode</i> on the x-axis. [15 points]