DM887 Reinforcement Learning, Fall 2024 Melih Kandemir

Assignment 1: Q-Learning for GridWorld [15 points]

Deadline: 8th March 2024, 23:55 CET

Description:

1. **[4 points]** Design two GridWorld games on the Gym environment implemented under https://github.com/prasenjit52282/GridWorld where one of the games can be played well if an agent learns to sacrifice immediate rewards to achieve a larger delayed reward. Choose grid size to be 20x20 or larger.
2. **[7 points]** Train the learning agent using N-Step Q-Learning for N=1, 2, 3 in tabular form that uses an epsilon-greedy behavior policy with epsilon=0.1, 0.2, 0.3. Repeat the whole training process ten times and compute the learning curve of the agent (x-axis: episode count, y-axis: the observed cumulative cost per episode) for every 5th episode. Plot the mean and standard error (not deviation) of the cumulative cost across the 10 replications for each iteration. The plot should show the mean cumulative cost of each of the nine configurations as a solid curve and standard error as a shaded area surrounding it.
3. **[4 points]** Write a one-paragraph comment about how the choice of epsilon and step count affect the learning curve and why. Also explain how the effect of epsilon and step count differ between the two environments and how this effect is related to the reward structure of the environments.

This is an individual assignment. The submitted deliverables are your own intellectual properties. Group discussions on solutions, brainstorming, and comparing your results with each other are allowed. However, implementation and reporting should be done on individual basis. As the implementation platform, use Python and Numpy.

Please submit the following deliverables via the itsLearning portal (emails do not count as submissions):

- Your one-page long report in PDF format
- All source code required to replicate the numbers and draw the figures you use in your report in a single Python (.py) file that saves the result plot in PNG format with name *“learning-curve.png”*.

until March 8th, 2024 23:55 Copenhagen Time. Late submissions will be graded according to the formula below:

[Valid Points] = max([Earned Points] – [Number of Calendar Days after DL]*3, 0).

For example, [Number of Calendar Days after DL] = 1 if you submit on March 9th at 23:56, and [Number of Calendar Days after DL] = 2 if you submit on March 10th, 23:56 and so forth.