Deep Q-Learning

Exercise 6.2: Deep Q-Learning

In this exercise, you will implement a Deep Q-Learning Agent.

Make sure that you have completed the setup requirements as described in the Set Up Lab Environments section.

Now, run jupyter notebook and open the "Ex6.2 Deep Q-Learning.ipynb" notebook under **Module 6** folder.

1. Examine the notebook. We have given you boiler plate and helper code for the implementation of the Deep Q-Learning agent.
2. Let's start with the **init**() function. Here you initialize the agent.
   There're several more things happening here than the previous implementation of q-learning agents:
   - Instead of using a dictionary to store the state-action values, or using thetas to maintain weight variables, here we are using a neural network.
   - The toolkit we are using is Chainer, but you can use other deep learning frameworks as well.
   - An implementation of the model network and the target network is based on the QNetwork class.
   - In addition, an implementation of replay memory is done based on the ReplayMemory class.
   - Other parameters are instantiated in this function using key word arguments.
3. Examine the **act()** function. We are still using the epsilon greedy policy. A simple hack is used to control whether we are using a fixed epsilon or we will use decaying epsilon. More on that later.
   Take note on how we obtain the max(Q(s,a)) using the model network.

DQLearningAgent in the SimpleRoomsEnv Environment (10 episodes)

Let's set up an experiment with your DQLearningAgent and the SimpleRoomsEnv environment.

1. Use the default values for alpha, epsilon, and gamma for your DQLearningAgent.
2. Once you've set up your experiment, run the the experiment for **10** episodes with the interactive set to **True**.
3. Run this experiment several times.

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, does the agent manage to reach the goal within 10 episodes?

- (●) No ✔
- ( ) Yes

Submit

You have used 1 of 2 attempts

DQLearningAgent in the SimpleRoomsEnv Environment (50 episodes)

1. Let's set up another experiment with the same parameters, use the default values for alpha, epsilon, and gamma for your DQLearningAgent.
2. But now, set the interactive to **False**, and run the experiment for **50** episodes.
3. Run this experiment several times.

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, around how many episodes does the agent need to achieve the goal consistently?

- ○ After the first episode the agent already managed to reach the goal albeit with huge number of steps
- ○ All it takes is two episodes
- ○ Ten to Fifteen episodes
- ○ Around 25 episodes
- ○ Only after 50 episodes
- ◉ Unable to observe even after 50 episodes ✔

Submit You have used 1 of 2 attempts

Turns out, with the default parameters of epsilon = 0.01, we will not be able to "solve" this environment within 50 episodes, unlike in our previous labs.

Why?

Now try to run the experiment with more episodes.

1. Set up another experiment with the same parameters, use the default values for alpha, epsilon, and gamma for your DQLearningAgent.
2. Set the interactive to **False**, and run the experiment for **200** episodes.
3. Run this experiment several times.

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, does the agent manage to reach the goal within 200 episodes?

- ○ No
- ◉ Yes ✔

Submit | You have used 1 of 2 attempts

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, around how many episodes does the agent need to achieve the goal consistently?

- ○ After around 50 episodes
- ○ After around 75 episodes
- ◉ After around 100 episodes ✔
- ○ After around 200 episodes

Submit | You have used 2 of 2 attempts

One possible reason is that the agent has yet to explore enough to be able to obtain a "good model" for the network.

Let's use our simple hack of decaying epsilon and set epsilon = 1 when setting up the experiment. Run the experiment for 200 episodes.

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, does the agent manage to reach the goal within 200 episodes?

- ○ No
- ◉ Yes ✔

Submit You have used 1 of 2 attempts

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, around how many episodes does the agent need to achieve the goal consistently?

- ○ After around 50 episodes
- ◉ After around 75 episodes ✔
- ○ After around 100 episodes
- ○ After around 200 episodes

Submit You have used 2 of 2 attempts

In this particular environment, the agent with decaying epsilon performed better than the one with fixed epsilon of 0.01.

As you've learned in Module 2, exploration vs exploitation is a balancing act and very much depend on case by case basis.