REINFORCE

Exercise 7.1: REINFORCE

In this exercise, you will implement the REINFORCE algorithm.

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 "Ex7.1 REINFORCE.ipynb" notebook.

Examine the notebook. We have given you boiler plate and helper code for the implementation of the REINFORCE algorithm.

Once you got yourself acquainted with the notebook, go to #TODO1. Here you need to modify the loss function to the desired loss function of the policy.

Lab Question

1/1 point (graded)

What could be an example of a correct loss function for this context?

- ○ loss = -C.reduce_mean(label + output + return_weight, axis=0, name='loss')
- ○ loss = -C.reduce_mean(C.log((label - output) + 1e-4) * return_weight, axis=0, name='loss')
- ○ loss = -C.reduce_mean((C.square(label - output) + 1e-4) * return_weight, axis=0, name='loss')

- ◉ loss = -C.reduce_mean(C.log(C.square(label - output) + 1e-4) * return_weight, axis=0, name='loss') ✔

Submit You have used 1 of 2 attempts

Next, go to #TODO2. Here you need to create a pseudo label, that will encourage the agent to increase the probability of the action it selected.

Lab Question

1/1 point (graded)

What could be an example of a correct pseudo label for this context?

- ○ y = 0
- ○ y = 1
- ○ y = 1 if action == 1 else 0
- ◉ y = 1 if action == 0 else 0 ✔

Submit You have used 1 of 2 attempts

You now have an implementation of the REINFORCE algorithm. Run this notebook several times and use max_number_of_episodes = 500.

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 500 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, what is the variance when the agent reach the goal?

- ○ Between 0 to 500
- ○ Between 500 to 1000
- ○ Between 1000 to 2000
- ◉ Above 2000 ✔

Submit You have used 2 of 2 attempts