Homework #3

COSE490(02)
Reinforcement Learning

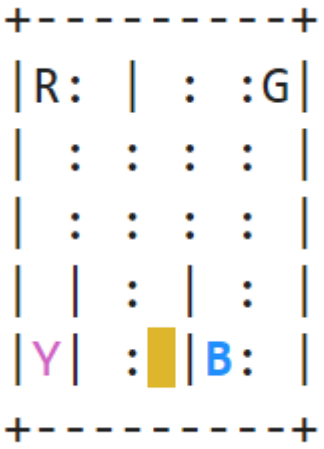
Goal: The goal of the third homework is to practice Deep Q Network (DQN).

(i) In previous homework, you implemented lookup table to save values of state-action pairs. This was possible because state-action space was small.

(ii) In this homework, you train the agent with DQN algorithm which use neural networks to approximate Q function.

Tasks: Use the provided notebook file, and complete the functions

- In this homework, we use taxi environment ("Taxi-v3") provided by Open AI Gym. Find the environment details at
  https://github.com/openai/gym/blob/master/gym/envs/toy_text/taxi.py
  (Detailed descriptions and codes are also provided in the first cell of given notebook.)
  The goal of this environment is to control taxi to pick up a customer and get to destination.



- We consider 5x5 space with total 25 "squares", in which there are four designated "locations" in the grid world indicated by R(ed), G(reen), Y(ellow), and B(lue). When the episode starts, the taxi starts off at a random square and the passenger is at a random location. The taxi drives to the passenger's location, picks up the passenger, drives to the passenger's destination (another one of the four specified locations), and then drops off the passenger. Once the passenger is dropped off, the episode ends.
- Observations: There are 500 discrete states since there are 25 taxi positions, 5 possible locations of the passenger (including the case when the passenger is in the taxi), and 4 destination locations.
- Passenger locations

    - 0: R(ed)
    - 1: G(reen)
    - 2: Y(ellow)
    - 3: B(lue)
    - 4: in taxi
- Destinations:
    - 0: R(ed)
    - 1: G(reen)
    - 2: Y(ellow)
    - 3: B(lue)
- Actions: There are 6 discrete deterministic actions:
    - 0: move south
    - 1: move north
    - 2: move east
    - 3: move west
    - 4: pickup passenger
    - 5: drop off passenger
- Rewards:
    - There is a default per-step reward of -1
    - Delivering the passenger: +20
    - Executing "pickup" and "drop-off" actions without passenger: -10
- State space = (taxi_row, taxi_col, passenger_location, destination)

- **Task 1: Complete the main code and the training function** (marked in the notebook). We recommend you to refer to the sources cited in the head of notebook. If you refer to other sources, you MUST specify those references (in the notebook as "Markdown").
- **Task 2: Optimize several parameters** (marked in the notebook). Even after you successfully complete Task 1, the results may not be satisfactory due to non-optimized parameters. You should delicately tune the parameters to achieve good performance.

Evaluation:

- Working codes and their accessibility (how easy to read)
- Performance of your codes (convergence speed, average reward and variance (after convergence))

Submission:

- After running the program (all the cells), print the results as a PDF file.
- Submit both your code, saved model and the PDF file through BlackBoard (BB).
- Submission is due on June 02 (before the class hour).