DL_comp4_13_report

Method we use:
In the competition, we tried to use advantage actor critic to develop our model. Both of the actor and critic composed of 4 middle dense layers with units = 512, 256,128,64. Each of them is connected with RELU.

The last dense layer of critic outputs a critic value, and the last dense layer (connected with softmax) of actor outputs the probability of actions.

We collect the states, next states, advantage value, action probability for each episode and update the weight of actor and critic after an episode ends.

What makes the agent work?
We find out that do some preprocess on the input states can reduce the time to train the agent. Without doing preprocess, we need lots of episodes to train.
The preprocess method we try is to insert the state values into buckets, and then the actor will according to the buckets to output the action probability distribution.
Also, the critic will according to the buckets to output the critic value.
For the optimizer, we use Adam, and the learning rate is 1e-4.
For the discount rate, we have tried to use 0.99.

In the public, the agent can have average reward 17.89.