COMP9444 Assignment 3 Report
z5089358
Fengting YANG

Design:

For q values and target values, there is two fully connected layers, which has 64 units as middle output.

For each step, choose the best one from all possible actions. Then, put state, reward, next state, action and done in replay buffer until full. The buffer size is set to 20000. For the next step, if training process is enough or reach the max performance, stop training.

For optimizer, AdamOptimizer with learning rate 1e-3 is used. Loss function is MSE.

Result:

At ep 100, it can reach avg reward to greater than 190. It normally cost 2 minutes to finish 500 episodes.