COMP9444 (18S2) Assignment 3
Submission Deadline: Sunday, 21th October 2018 at 23:59:59
z5136414 Andrew Wirjaputra

Network Structure

The goal of Q-learning is to learn a policy which tells an agent what action to take under any circumstances without explicitly modeling the environment. One major limitation to classical Q-Learning is the number of possible states had to be reduced from ultimately infinity to a very limited number of discrete states. One way of utilizing more of the information from the observations is by combining it with a deep neural network.

The neural network I'm using is a simple 3 layer feed forward network with tanh activation function. After experimenting with the number of hidden nodes in the hidden layers, I found that having a larger number of hidden nodes in the second hidden layer works better on average. The Adam optimizer is chosen as it is generally regarded as being fairly robust to the choice of hyper parameters. However, the learning rate sometimes needs to be changed from the suggested default. A learning rate (Alpha) of 0.001 is chosen to balance the speed of convergence versus run time.

An easy way to speed up training is to collect small batches of experiences then compute the update step on this batch. This has the effect of ensuring the Q-values are updated over a larger number of steps in the environment; however these steps will remain highly correlated. To ensure batches are de-correlated, save the experiences gathered by stepping through the environment into an array, then sample from this array at random to create the batch. This method called experience replay, should significantly improve the robustness and stability of learning.

A sufficiently large replay memory should be chosen in order to hold experiences from larger number of episodes. Assuming an average survival time of 100 steps, using a replay buffer size of 10000 allows it to hold experiences from 100 episodes.