

COMP9444 Neural Networks and Deep Learning
Session 2, 2018
Project 3 - Deep Reinforcement Learning
Prajwal Rao | z5176504

Overview

Started with basic model, trained with different hyper parameters and tweaked the learning rates.

Basic Implementation

Created a fully connected network with two hidden layers – 16 and 48 nodes respectively. Used tanh activation function for hidden layers. Chose a discount factor of 0.99 with epsilon in the range of 0.6 to 0.01.

Batching

Created mini batches of size 256 for training. Used this value to determine if enough experience is available for choosing random samples to create a minibatch.

Experience Replay

Used a buffer size of 50000 to store game information as a collection of values: [state, action, reward, next_state, done]

Extras

Modified learning rates and beta values of AdamOptimizer to speed up training.

Summary

Although there were slight variations in how the model performed when it was run on a restarted machine versus successive runs – often decreasing the initial 100 episodes reward, the highest average reward – **200** was obtained with the following hyper parameter combination:

GAMMA	INITIAL_EPSILON	FINAL_EPSILON	EPSILON_DECAY_STEPS	BATCH_SIZE
0.99	0.6	0.01	100	256