Banana Navigation Project Report

Using a simple deep neural network DQN to approximate the action value function, we train a reinforcement learning agent to navigate a squared world and pick yellow bananas and avoid the blue ones. For every yellow banana the agent get a reward of +1 and a reward -1 for each blue one.

This is an episodic problem and it finishes as the agent achieves an average score of +13 over 100 consecutive episodes.

Hyperparameters used for training are as follow:

1) reply buffer size: 5000
2) Batch size: 64
3) Discount rate (Gamma): 0.99
4) Soft update rule (Tau): 1e-3
5) Learning rate: 1e-4
6) Update the network every: 2