Actor Critc with Neural Network Parametrization in OpenAI Gym Pendulum Environment
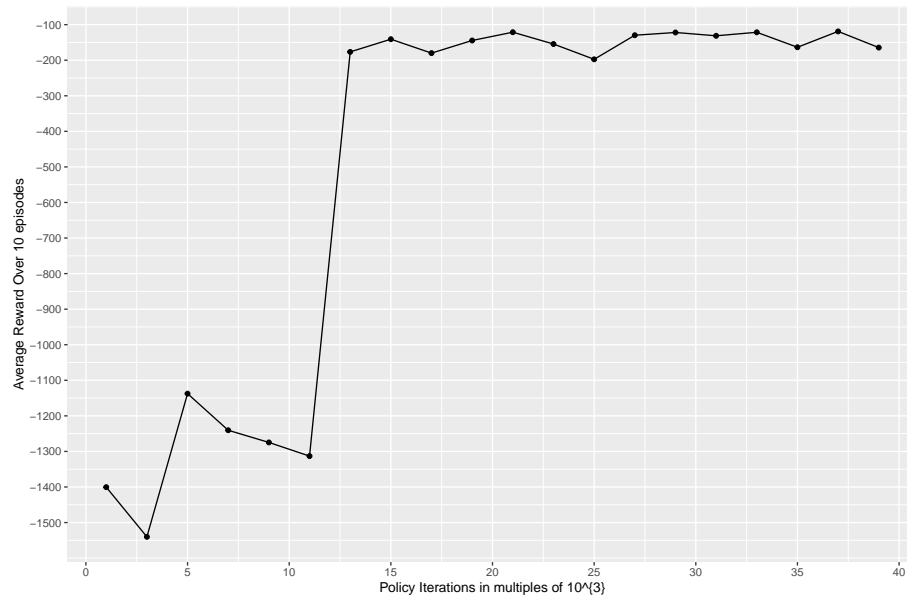

We use a three layer neural network for both the actor and the critic. After every 2000 iterations, the policy is evaluated over 10 episodes and the average reward over the 10 eposiodes is taken.