Homework 4

Ted Xiao

March 26, 2017

1 Continuous Policy Gradient

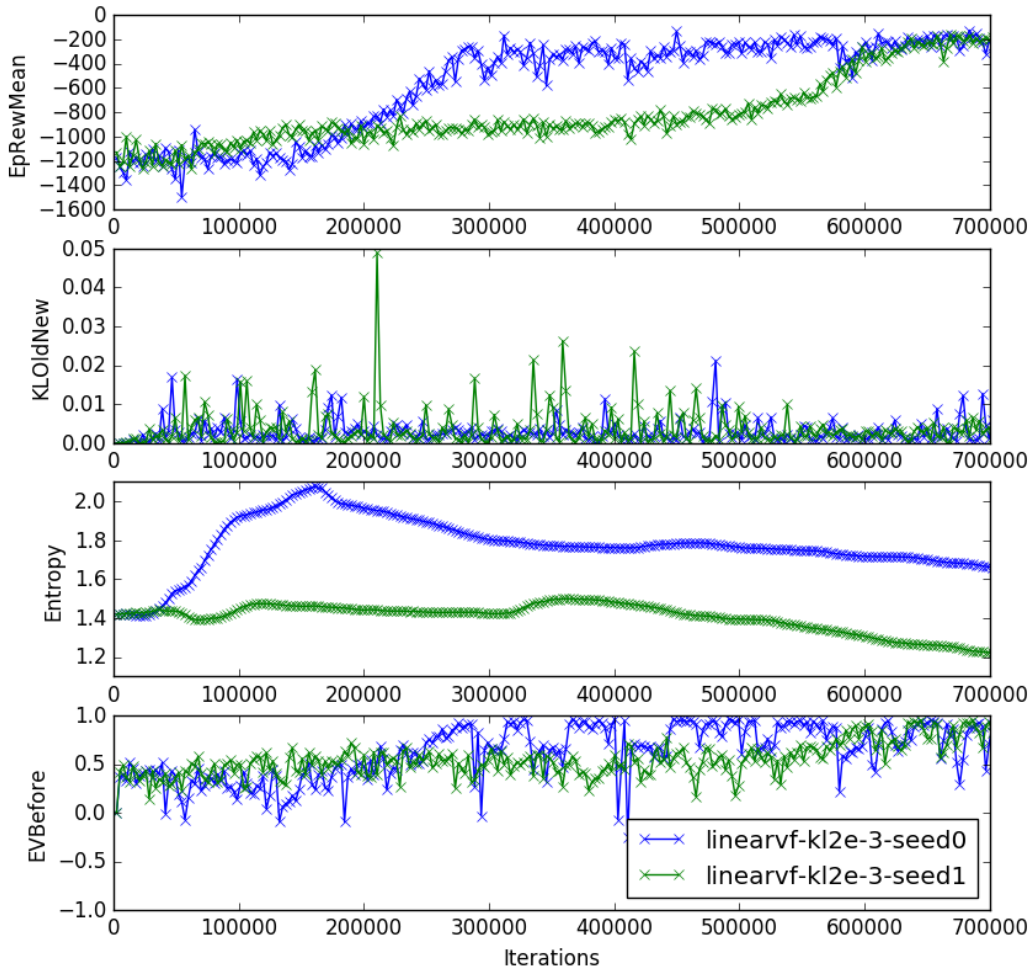

Figure 1: Continuous policy gradients applied to the Pendulum environment. Both seeds converge to rewards of at least -300. I used the default hyperparameters, but cut off the plot as soon as the results converged, which was at around 700000 steps. The default hyperparameters used were: $\gamma = 0.97$, n_iter=300,initial_stepsize=1e-3, desired_kl=2e-3, min_timesteps_per_batch=2500. As seen above, the algorithm learned in around 650000 to 700000 steps.

2 Neural Network Value Function