CS 285 Set 4

Erich Liang

Due: 11/3/21

1 Question 1

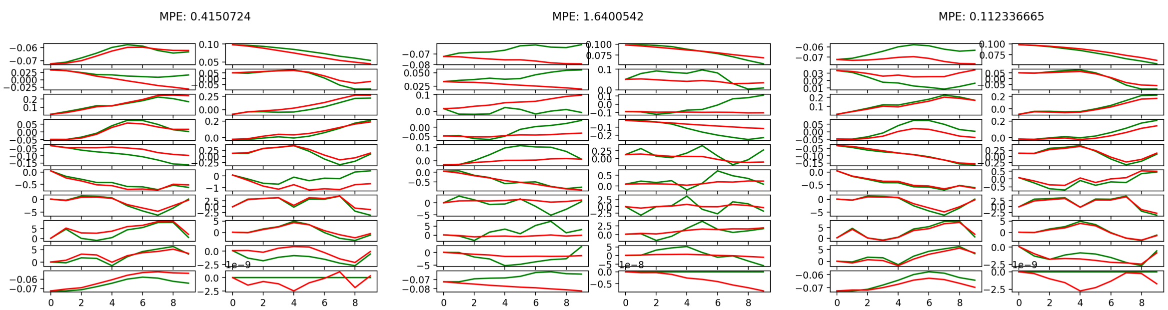

Figure 1: Model prediction results for neural networks that were trained for 1 iteration to learn a dynamics model. Left: results when the neural net had 1 layer of size 32 and was trained for 500 steps. Center: results when the neural net had 2 layers each of size 250 and was trained for 5 steps. Right: results when the neural net had 2 layers each of size 250 and was trained for 500 steps. Between the neural net models, the one that had 2 layers each of size 250 and was trained for 500 steps was the best; this is because its predictions in red seem to best match the ground truth states in green and its MPE value is the lowest. In addition, by looking at how the loss values of each model changed over time, the model with 2 layers each of size 250 and trained for 500 steps converged the fastest to an error below 0.2.

2 Question 2

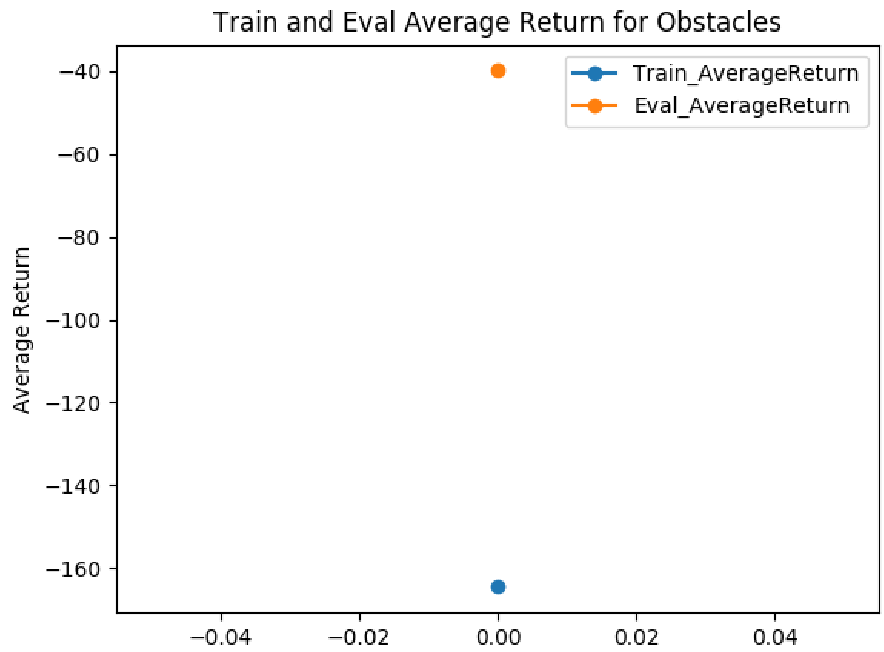

Figure 2: Train and eval average returns for the obstacles problem after one iteration of model based RL.

3 Question 3

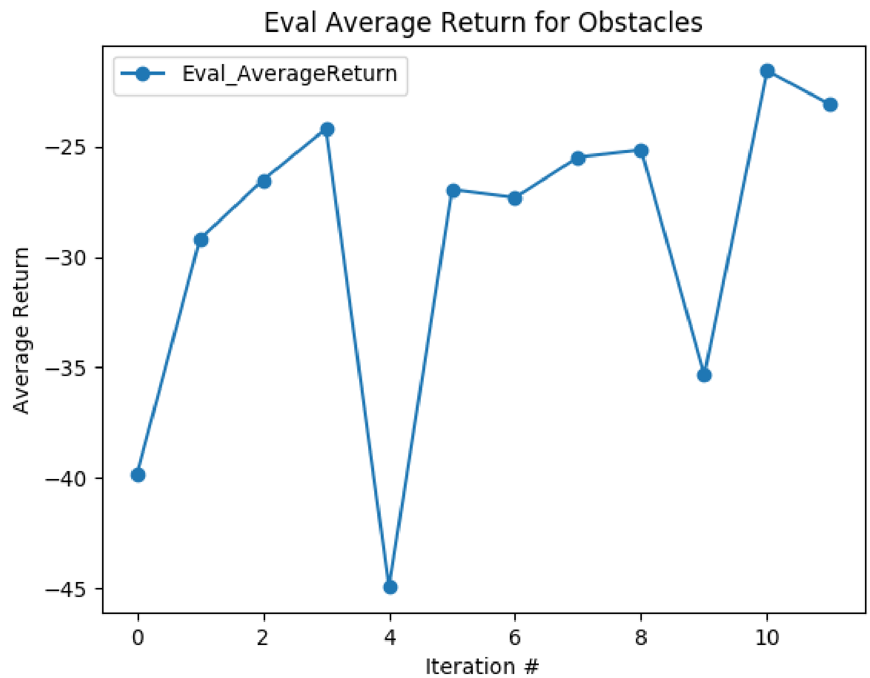

Figure 3: Eval average returns over multiple iterations for the obstacles problem using model based RL.

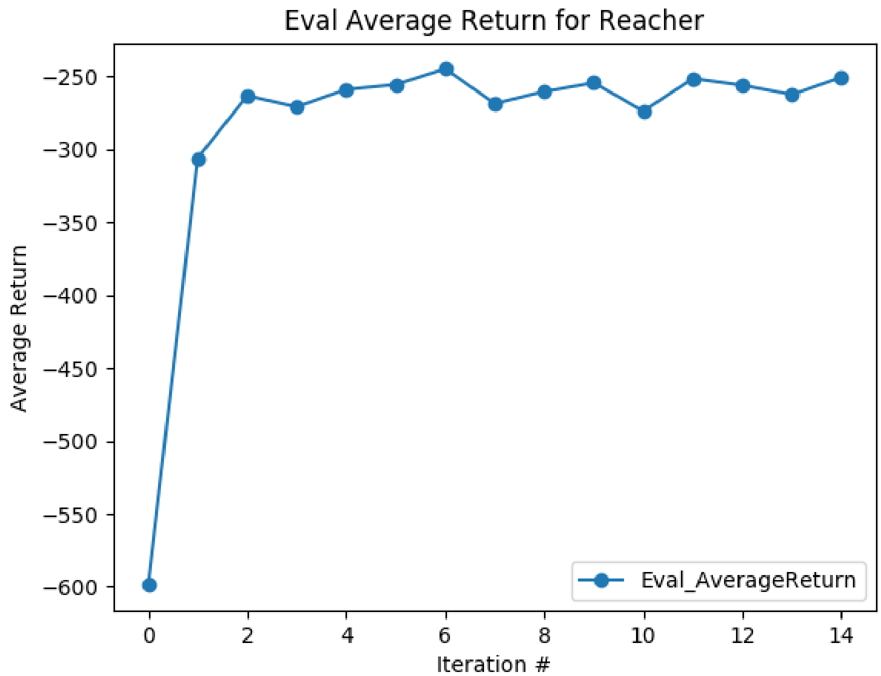

Figure 4: Eval average returns over multiple iterations for the reacher problem using model based RL.

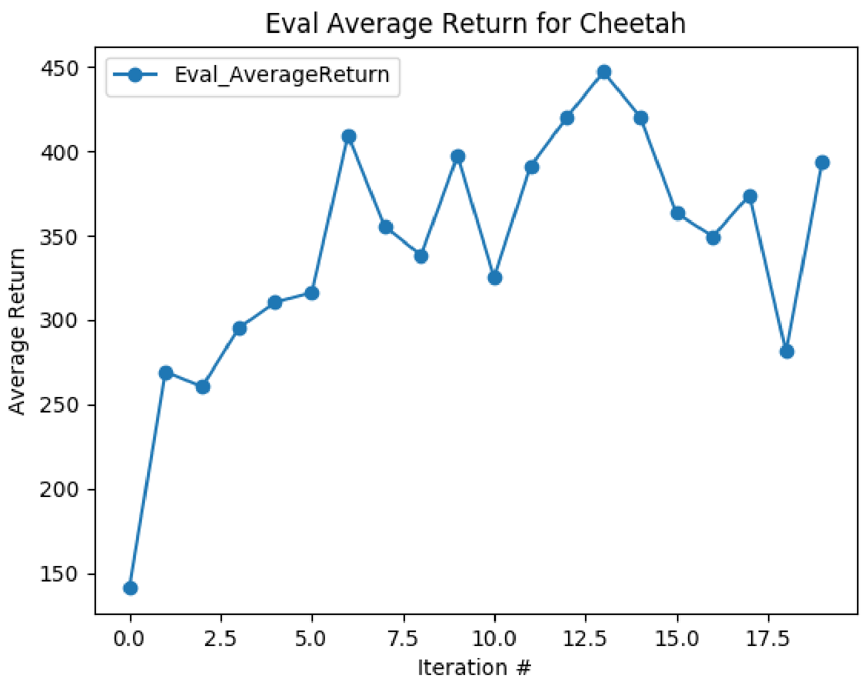

Figure 5: Eval average returns over multiple iterations for the half cheetah problem using model based RL.

4 Question 4

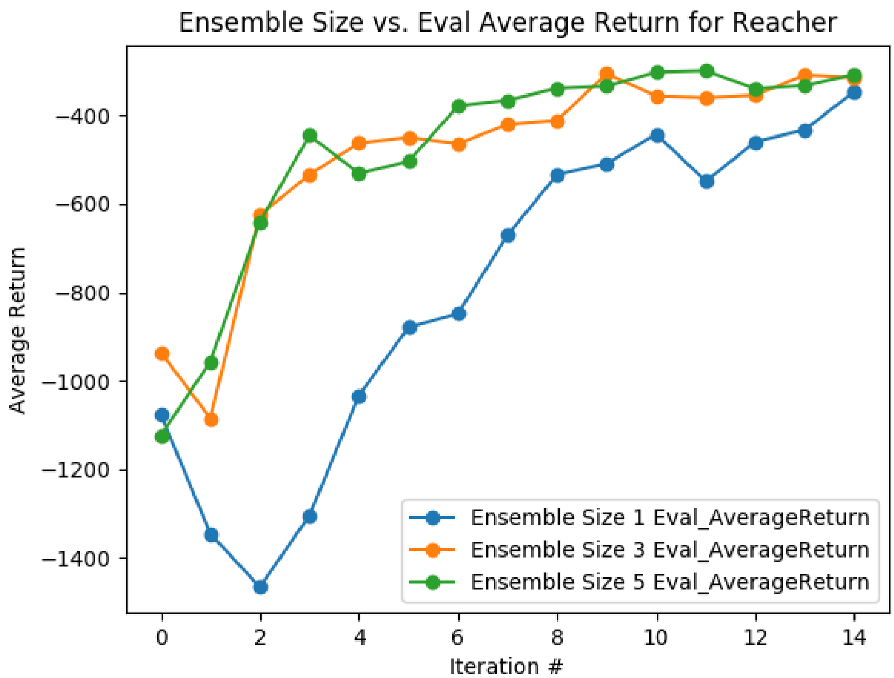

Figure 6: Ensemble size vs. Eval average returns over multiple iterations for the reacher problem using model based RL. From this experiment, we can see that as the ensemble size increases, the algorithm tends to reach higher returns in a fewer number of iterations. There seems to be diminishing returns with the more ensembles added.

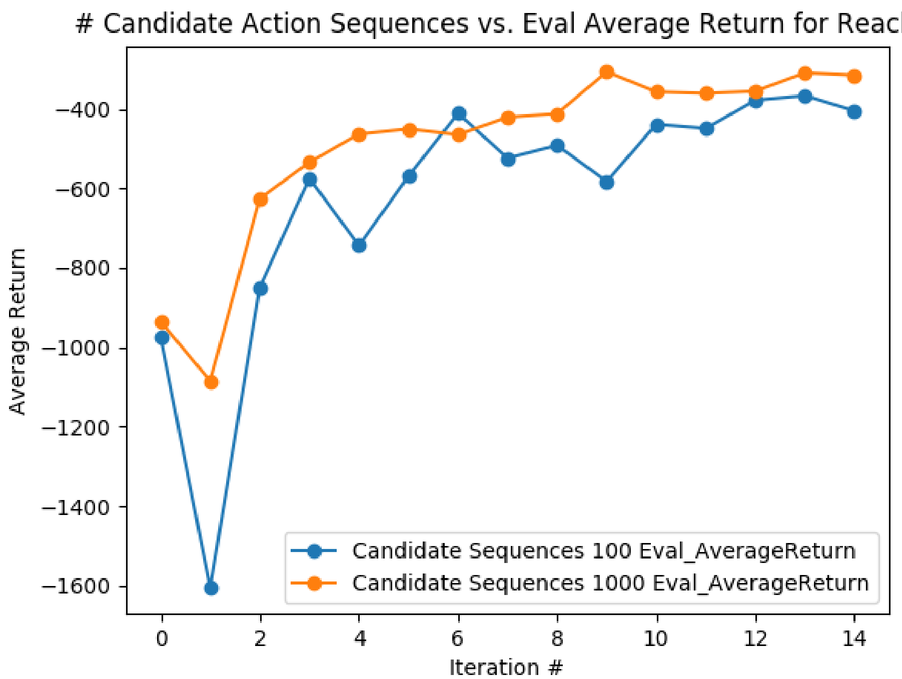

Figure 7: Number of candidate action sequences vs. Eval average returns over multiple iterations for the reacher problem using model based RL. From this experiment, we can see that as the number of candidate action sequences increases, the algorithm tends to reach higher returns in a fewer number of iterations.

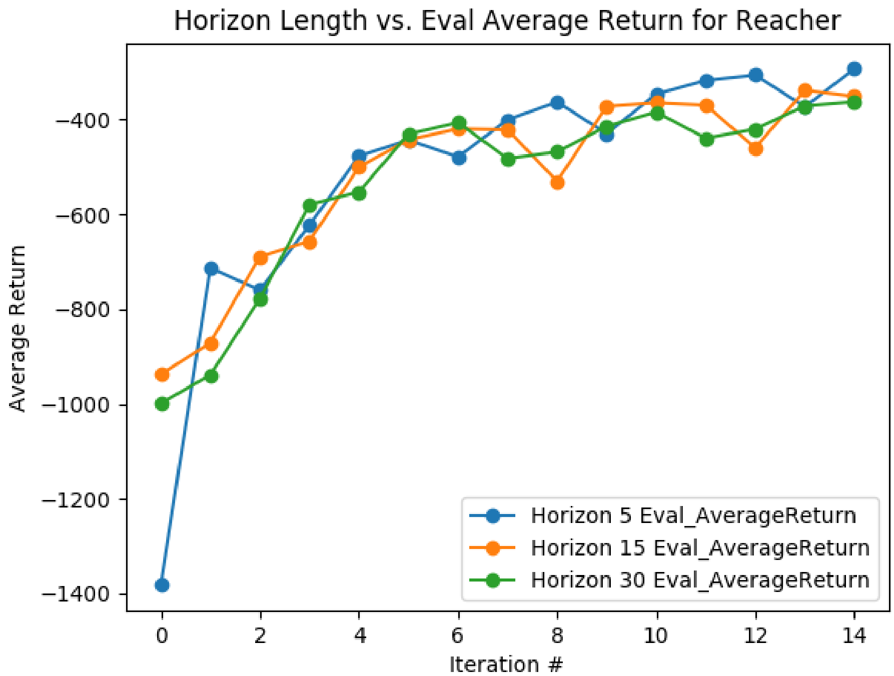

Figure 8: Horizon length vs. Eval average returns over multiple iterations for the reacher problem using model based RL. From this experiment, it seems that as the horizon length is increased, the resulting algorithms achieve lower returns overall. This may be because as we increase the horizon length, our predicted sum of rewards becomes less accurate (due to build up of inaccuracies of our model of the dynamics), which leads to suboptimal actions appearing as the “best” action to take, leading to lower rewards.

5 Question 5

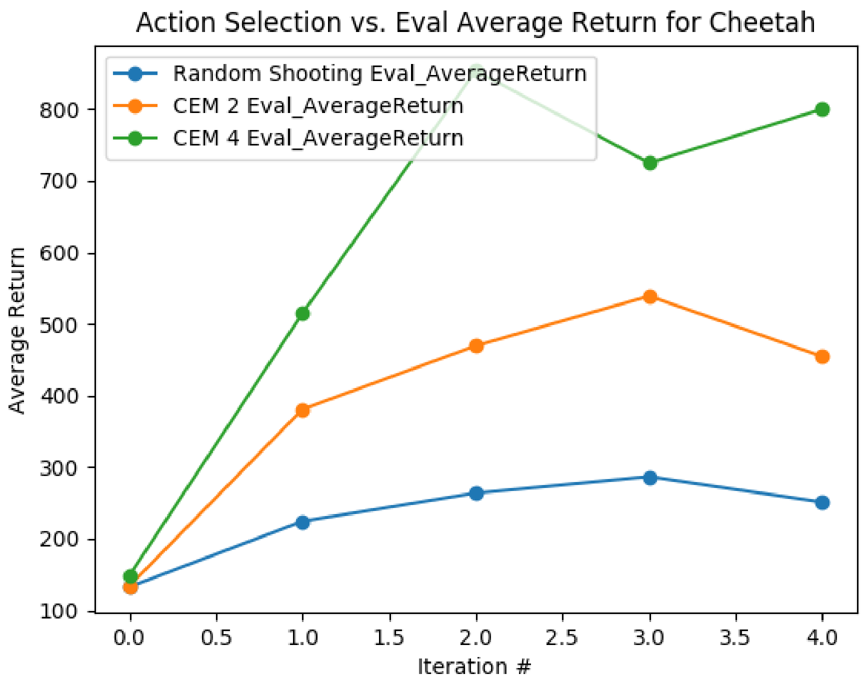

Figure 9: Action selection strategies vs. Eval average returns over multiple iterations for the cheetah problem using model based RL. From this experiment, we can see that using CEM instead of random shooting for action selection results in algorithms that achieve much higher average returns in fewer number of iterations. Additionally between CEM 2 and CEM 4, it using more sampling iterations for CEM (CEM 4) also increases the overall returns seen in most iterations.