Assignment 4: Model-Based RL and Exploration

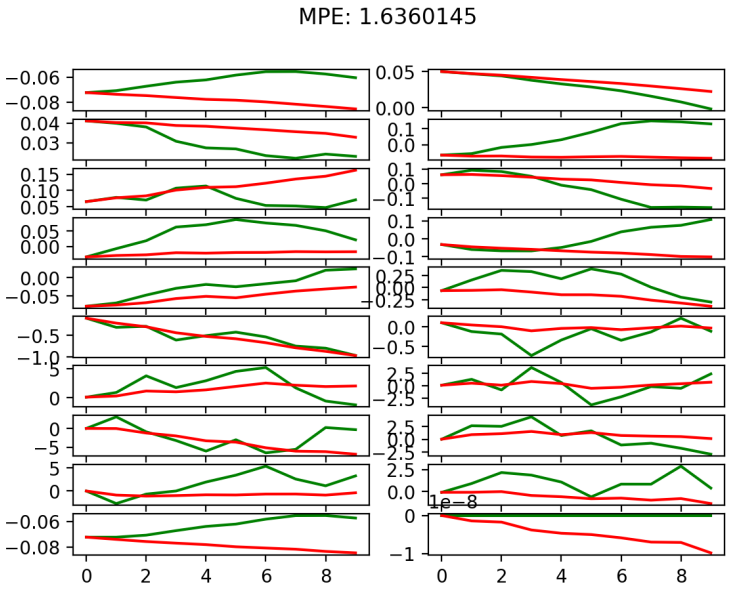
Andrew ID: `ayanovic`
Collaborators: `Write the Andrew IDs of your collaborators here (if any).`
NOTE: Please do **NOT** change the sizes of the answer blocks or plots.

1 Problem 1: Dynamics Model Training – [10 points total]

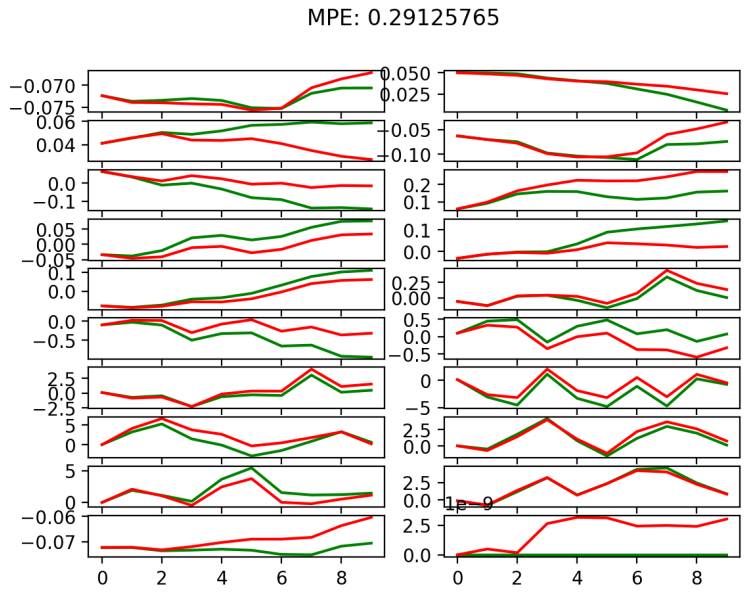
Theory questions

The following plots show the predicted rewards and the true rewards for the first iteration of the model training. The first plot shows the predicted rewards for the model trained with 5 trajectories and a neural network architecture of 2 layers with 250 units each. The second and third plots show the predicted rewards for the models trained with 500 trajectories and neural network architectures of 1 layer with 32 units and 2 layers with 250 units, respectively. Out of the three models, the model trained with 500 trajectories and a neural network architecture of 2 layers with 250 units has the best performance, where the predicted rewards are closest to the true rewards measured by the MPE (Mean Prediction Error).

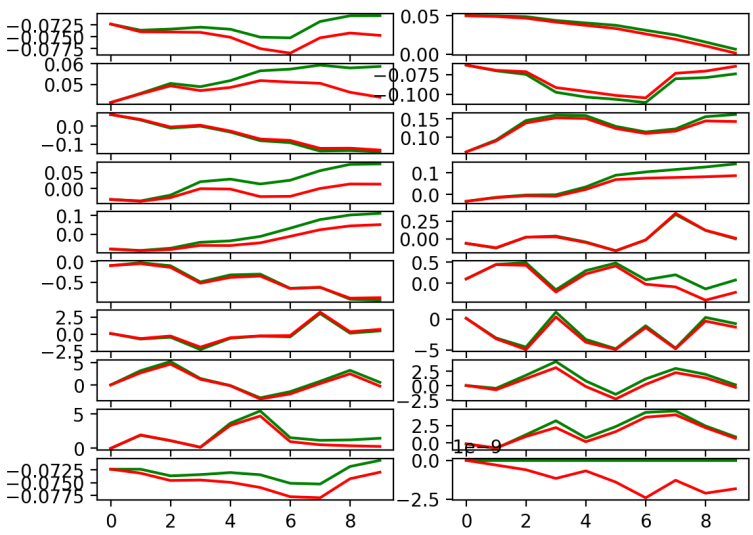
Performance Plot Cheetah Env (1)



Performance Plot Cheetah Env (2)

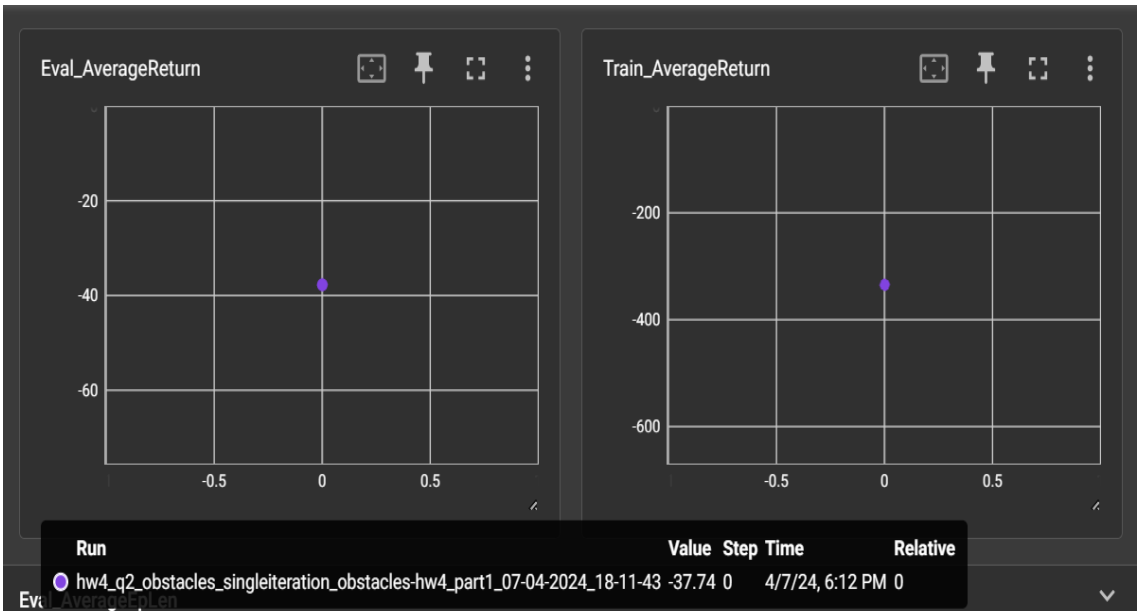


Performance Plot Cheetah Env (3)



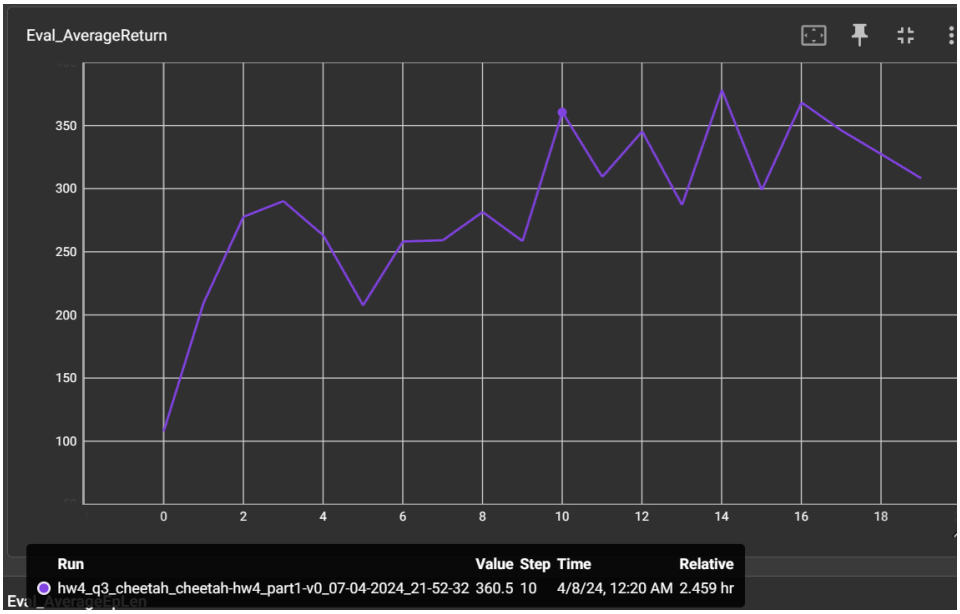
2 Problem 2: Action Selection

Reward Return Plot Obstacles Env

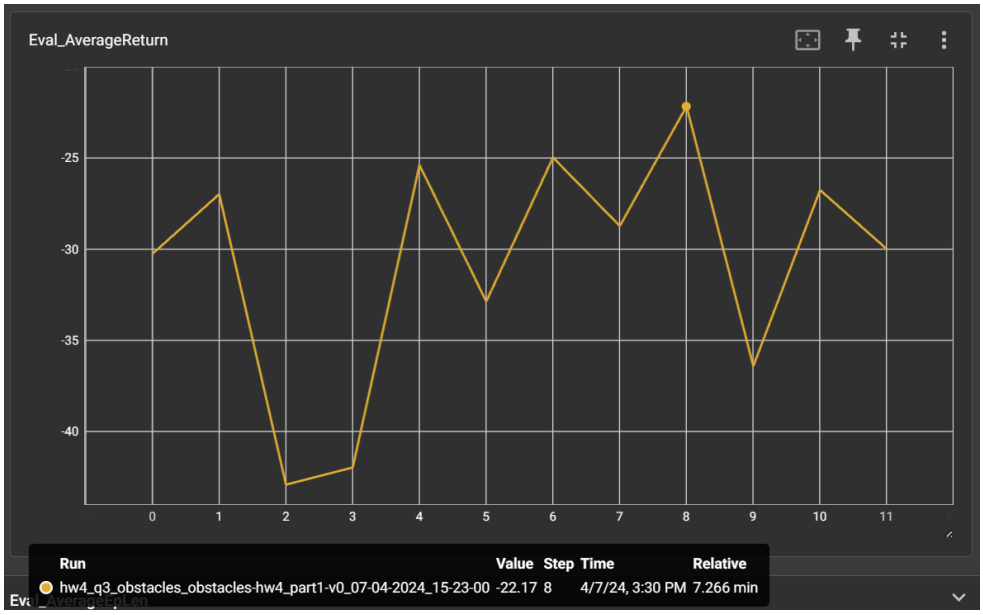


3 Problem 3: Iterative Model Training

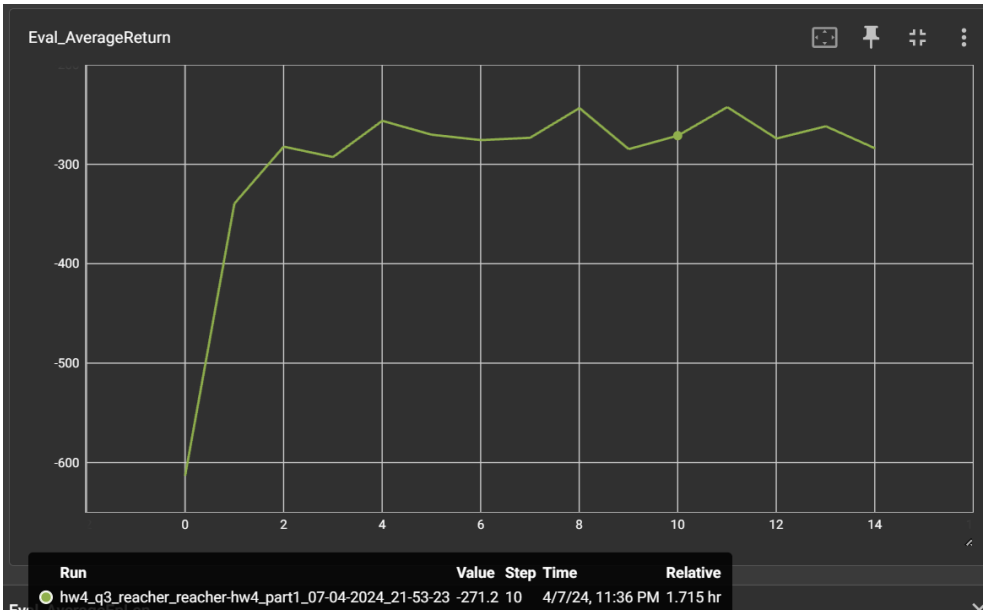
Performance Plot Cheetah Env



Performance Plot Obstacles Env

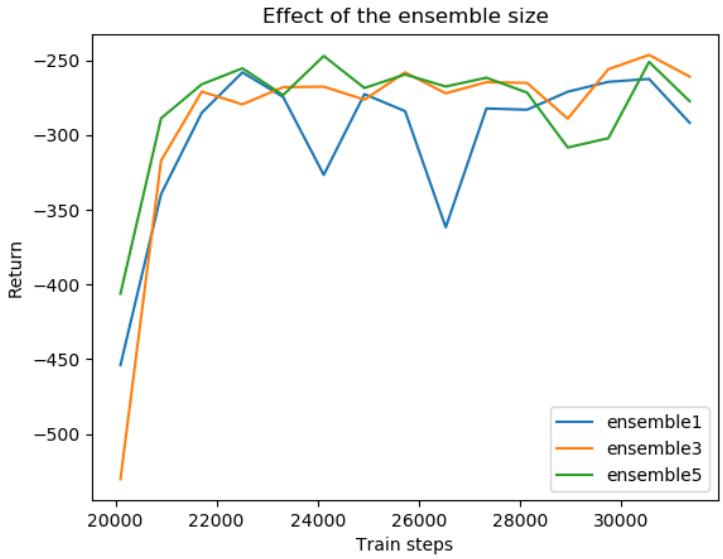


Performance Plot Reacher Env



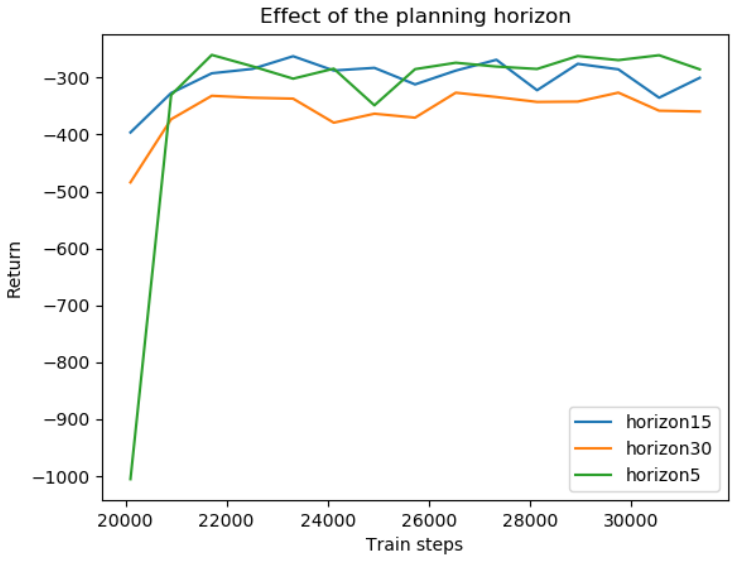
4 Problem 4: Hyper-parameter Comparison

Plot



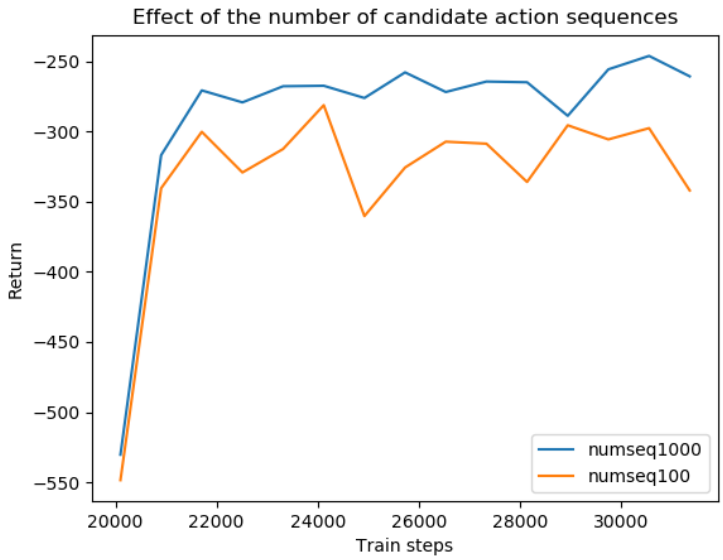
Comments: Model performance tend to be more stable with a larger ensemble size.

Plot



Comments: Model performance tend to be more stable with a larger horizon. However, large horizon does not always lead to better performance. As shown in the plot, the model trained with a horizon of 5 has generally better performance.

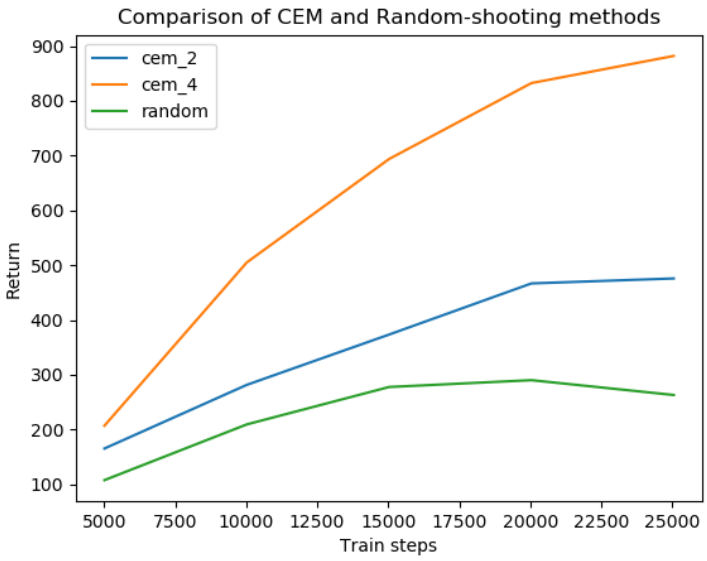
Plot



Comments: Model performance tend to be more stable and yield higher average returns with a larger number of candidate action sequences.

5 Problem 5: Hyper-parameter Comparison (Bonus)

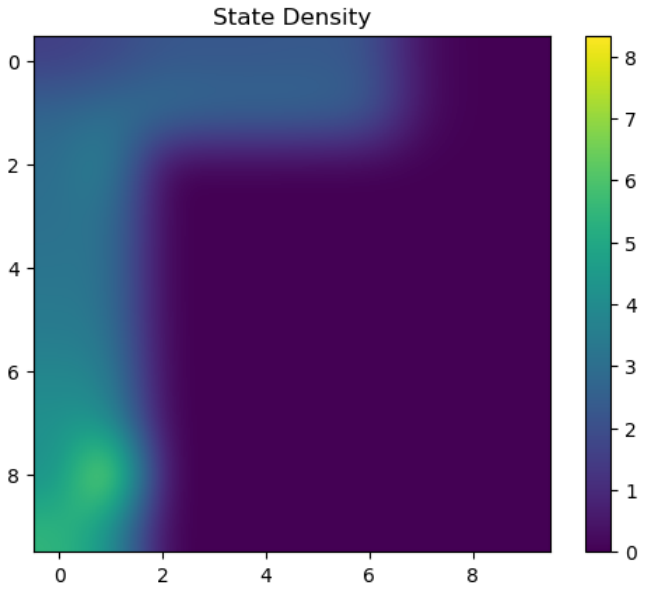
Plot



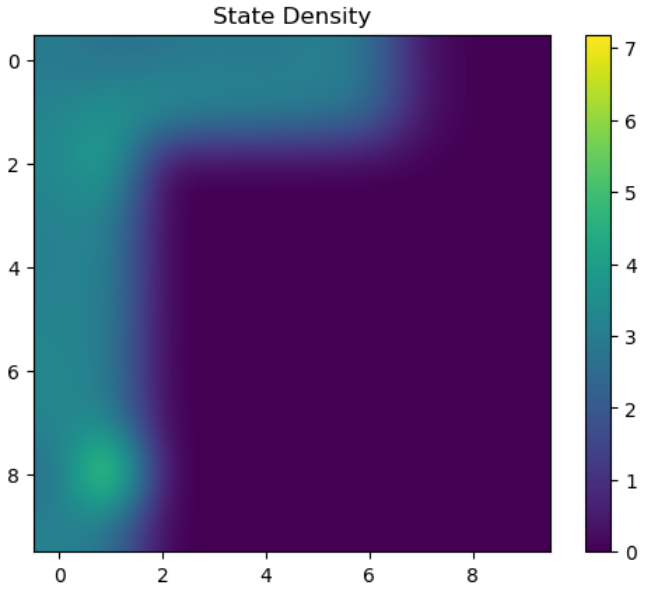
Comments: The models that utilize CEM consistently outperform the models that utilize random action selection. CEM tends to yield higher average returns with more CEM iterations.

6 Problem 6: Exploration (Bonus)

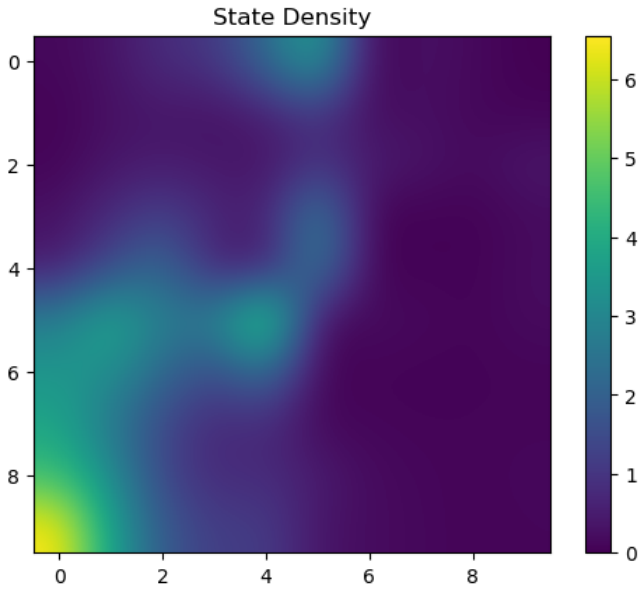
Density Plot Pointmass Easy Random



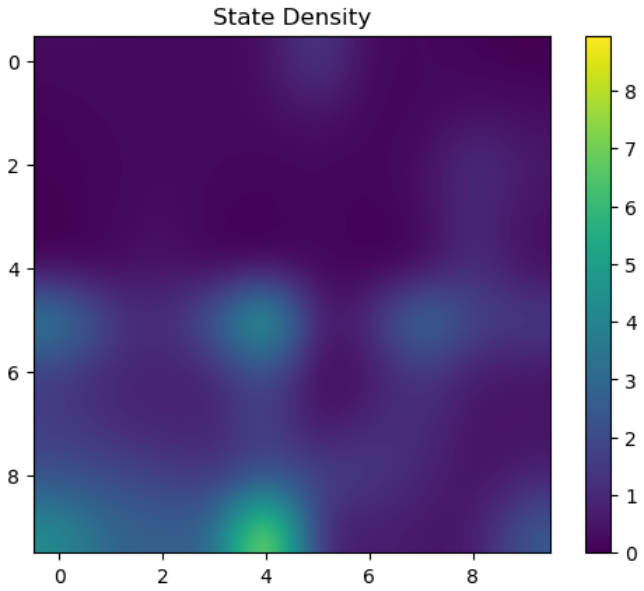
Density Plot Pointmass Easy RND



Density Plot Pointmass Hard Random



Density Plot Pointmass Hard RND



Reward Return Plot for Pointmass Env (Easy and Hard)