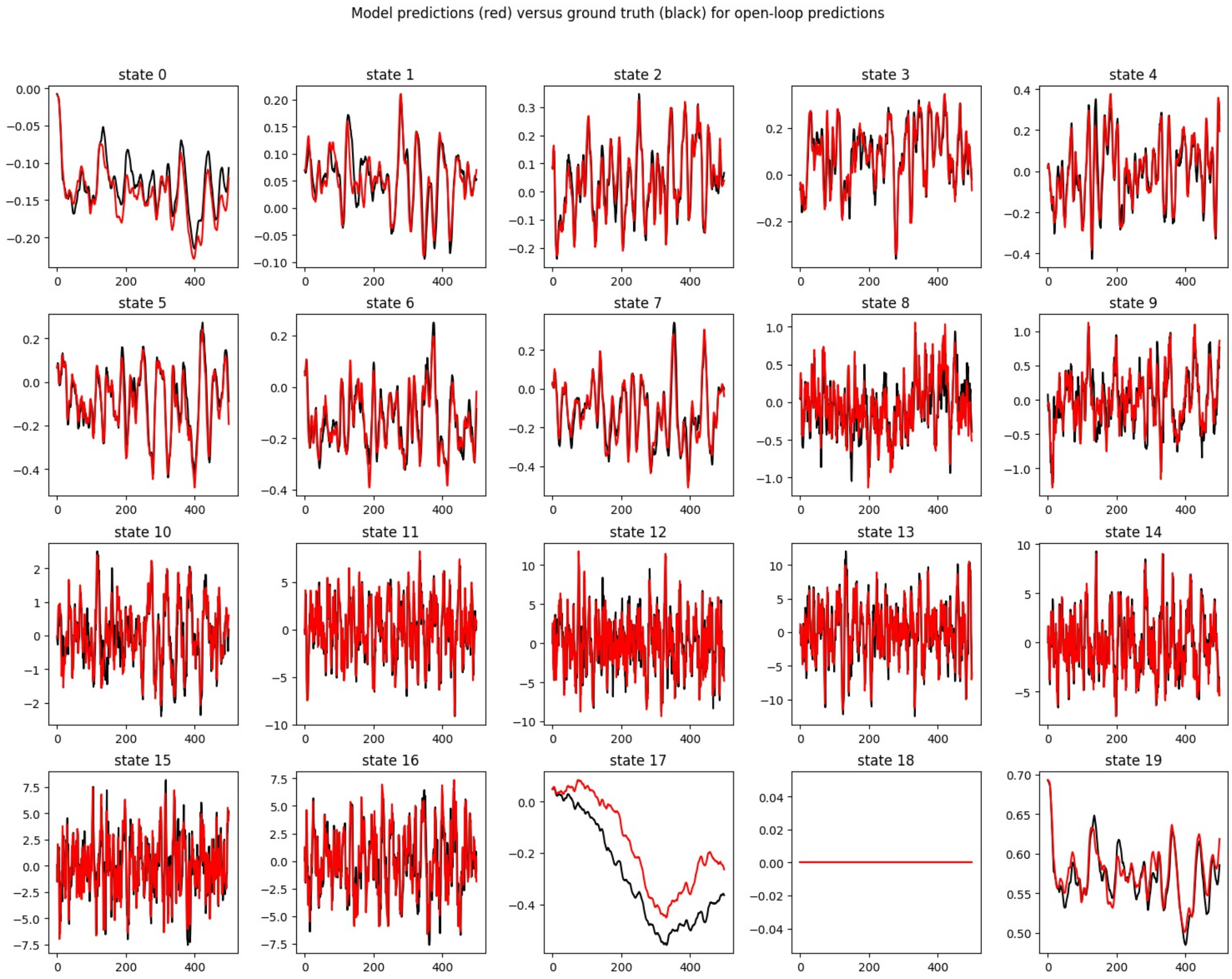
Student: **Ninh DO** SID#**25949105**

CS 294-112 – Homework#4

Problem 1

(a)



(b) State 17 is the most inaccurate.

<u>Possible reasons:</u>

- State 17 has a larger degree of freedom than that of the other states, i.e. State 17 may have more variants given the current state and action, thus it need more training time and data.
- The policy is random and the model is trained on the randomly generated dataset while State 17 has a clear pattern. The natures of the policy and State 17 don’t match.
- The training is not long enough or the generated training dataset is not big enough to capture the pattern of the state 17, i.e. the neural network is not sufficiently mature and should be trained more on a bigger dataset.

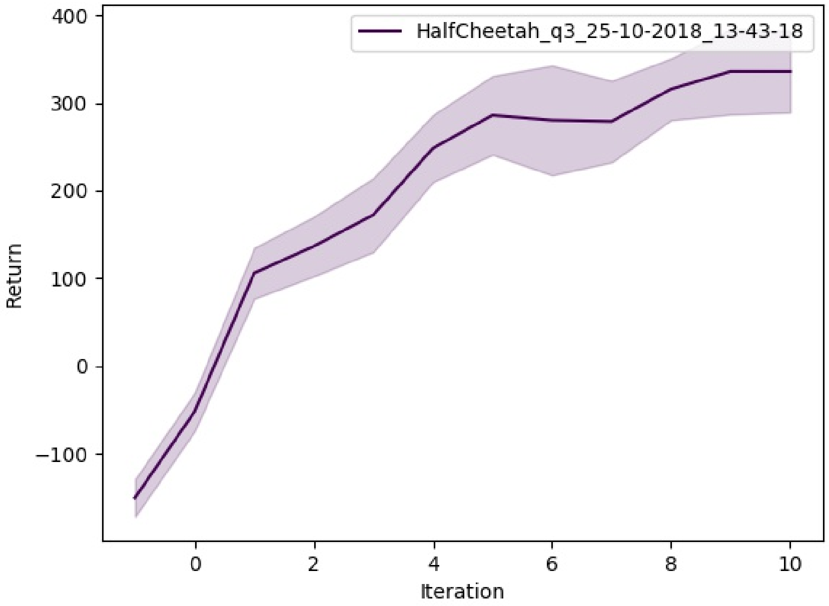
Student: **Ninh DO** SID#**25949105**

Problem 2

```
10-25 14:33:34 HalfCheetah_q2_25-10-2018_14-33-34 INFO     Gathering random dataset
10-25 14:33:34 HalfCheetah_q2_25-10-2018_14-33-34 INFO     Creating policy
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     Random policy
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ---------  ---------
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnAvg  -161.927
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnMax   -89.4764
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnMin  -207.699
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnStd    36.5249
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ---------  ---------
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : total      0.0 (100.0%)
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : other      0.0 (100.0%)
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG
10-25 14:33:38 HalfCheetah_q2_25-10-2018_14-33-34 INFO     Training policy....
10-25 14:33:39 HalfCheetah_q2_25-10-2018_14-33-34 INFO     Evaluating policy...
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     Trained policy
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     -----------------  ----------
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnAvg          48.9224
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnMax          91.9201
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnMin          14.5364
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     ReturnStd          26.7441
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     TrainingLossFinal   0.0295085
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     TrainingLossStart   1.07688
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 INFO     -----------------  ----------
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : total      160.6 (100.0%)
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : get action 157.0 (97.8%)
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : env step   1.7 (1.1%)
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : train policy 1.4 (0.9%)
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG    : other      0.4 (0.3%)
10-25 14:36:18 HalfCheetah_q2_25-10-2018_14-33-34 DEBUG
```

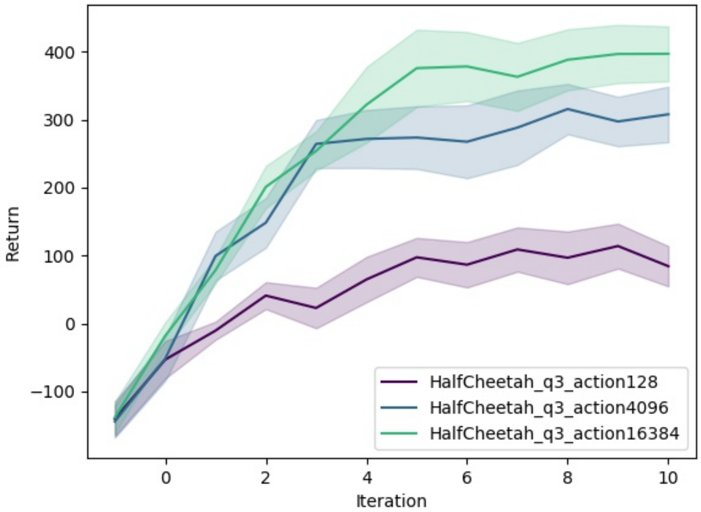
Student: **Ninh DO** SID#**25949105**

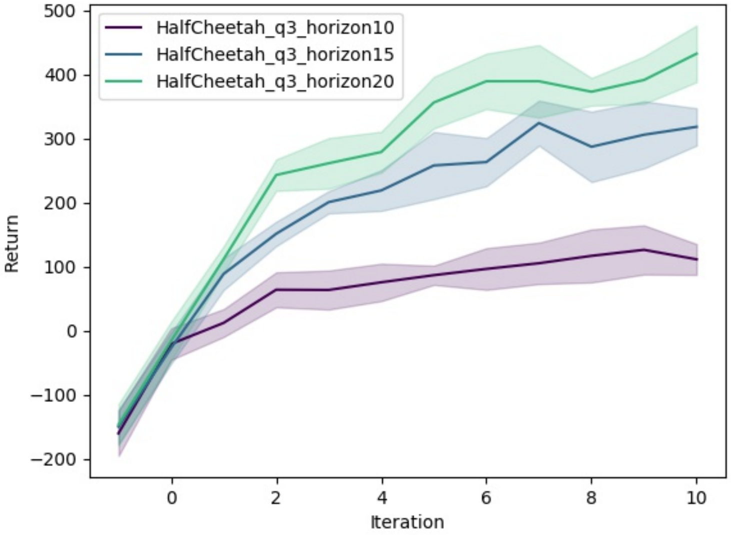
Problem 3a

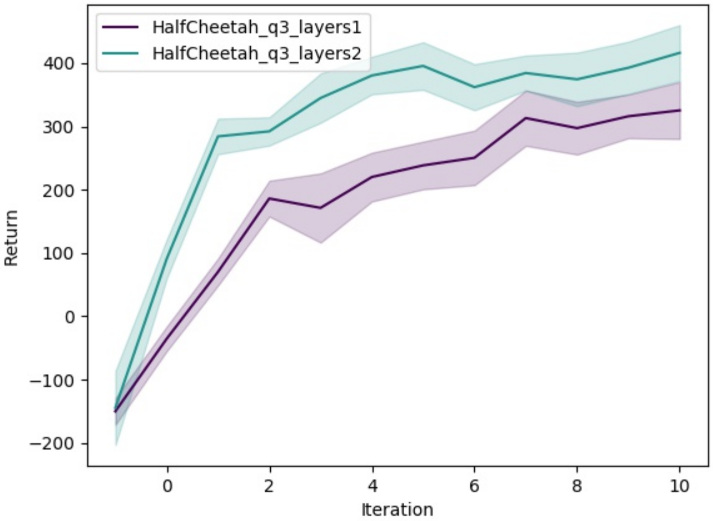


Student: **Ninh DO** SID#**25949105**

Problem 3b





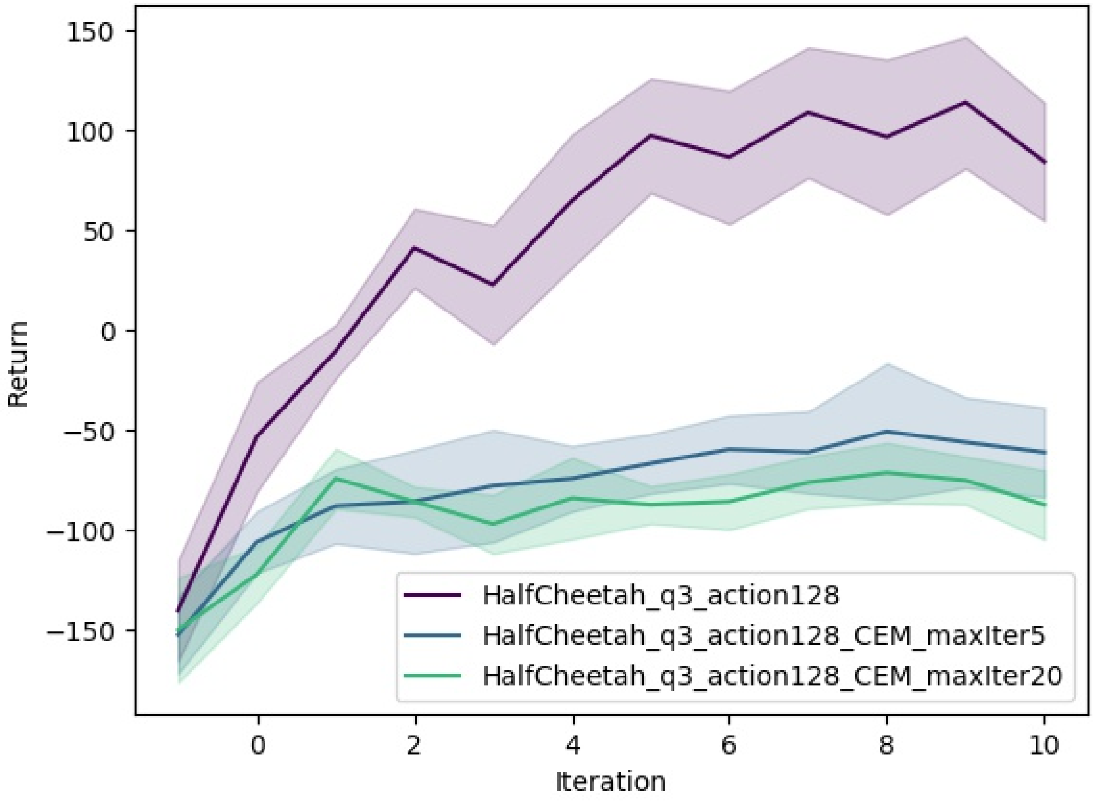


← This plot is still missing the case nn_layers = 3 as I did not have enough time to run it. However, Greg Kahn mentioned on Pizza that this plot is not graded because there is a bug in the skeleton code, i.e. nn_layers was never passed into ModelBasedPolicy (line 42, model_based_rl.py)

Student: **Ninh DO** SID#**25949105**

Extra Credit

I attempted implementing CEM. Unfortunately, the prelim result is pretty poor. The possible explanation is the number of iterations `max_iter` in the CEM algorithm is too small. If `max_iter` increases, the algorithm becomes expensive but may converge. By the way, I present the result as follows.


Return
Iteration
HalfCheetah_q3_action128
HalfCheetah_q3_action128_CEM_maxIter5
HalfCheetah_q3_action128_CEM_maxIter20