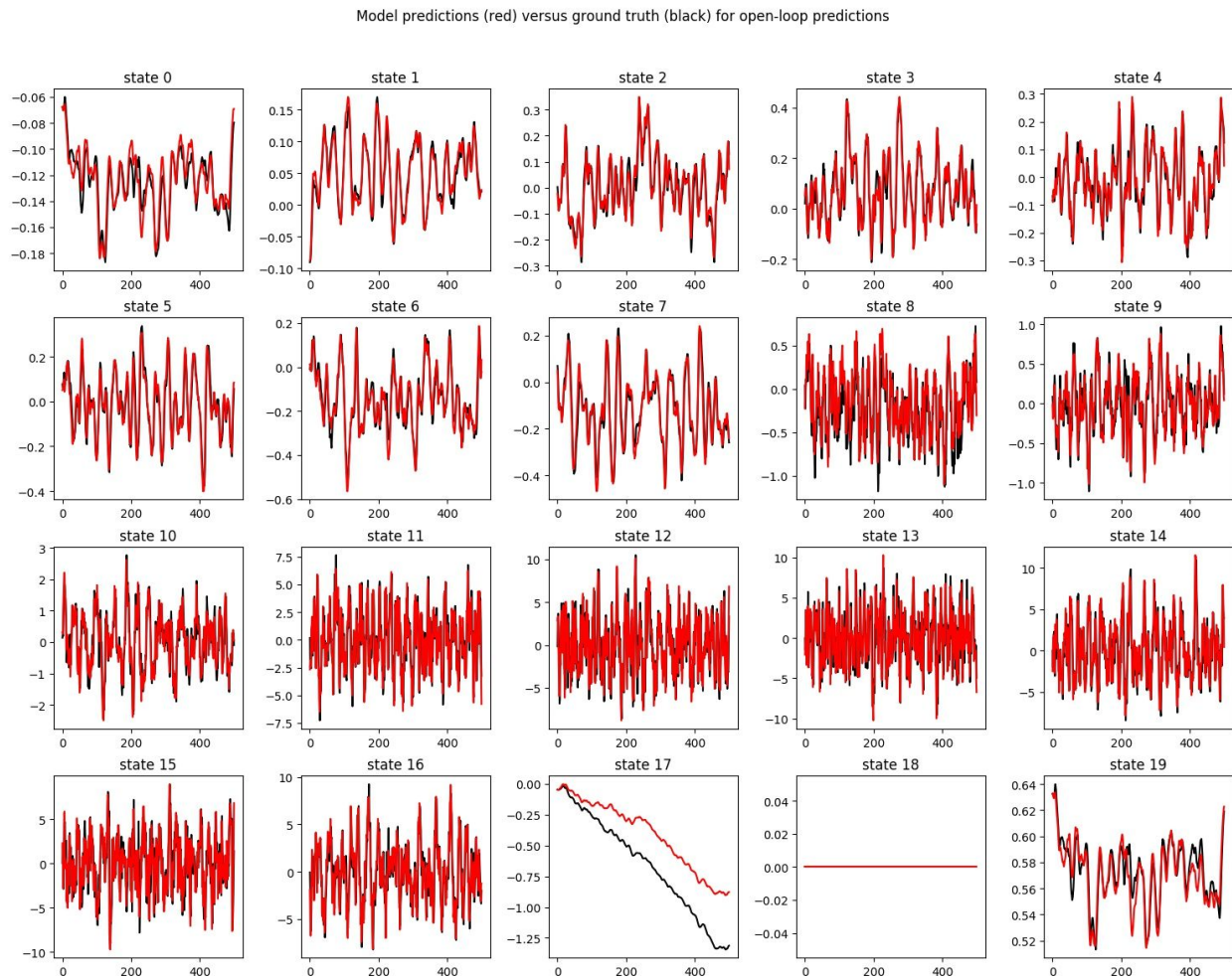


## Problem 1

(a)



(b) Predictions for state 17 is most inaccurate. This is probably because the test trajectory keeps decreasing for this dimension, so the prediction errors are accumulated.

## Problem 2

Random policy

-----

ReturnAvg -145.931

ReturnMax -113.702

ReturnMin -194.955

ReturnStd 24.4834

-----

Trained policy

-----

ReturnAvg -3.95774

ReturnMax 57.4863

ReturnMin -36.0475

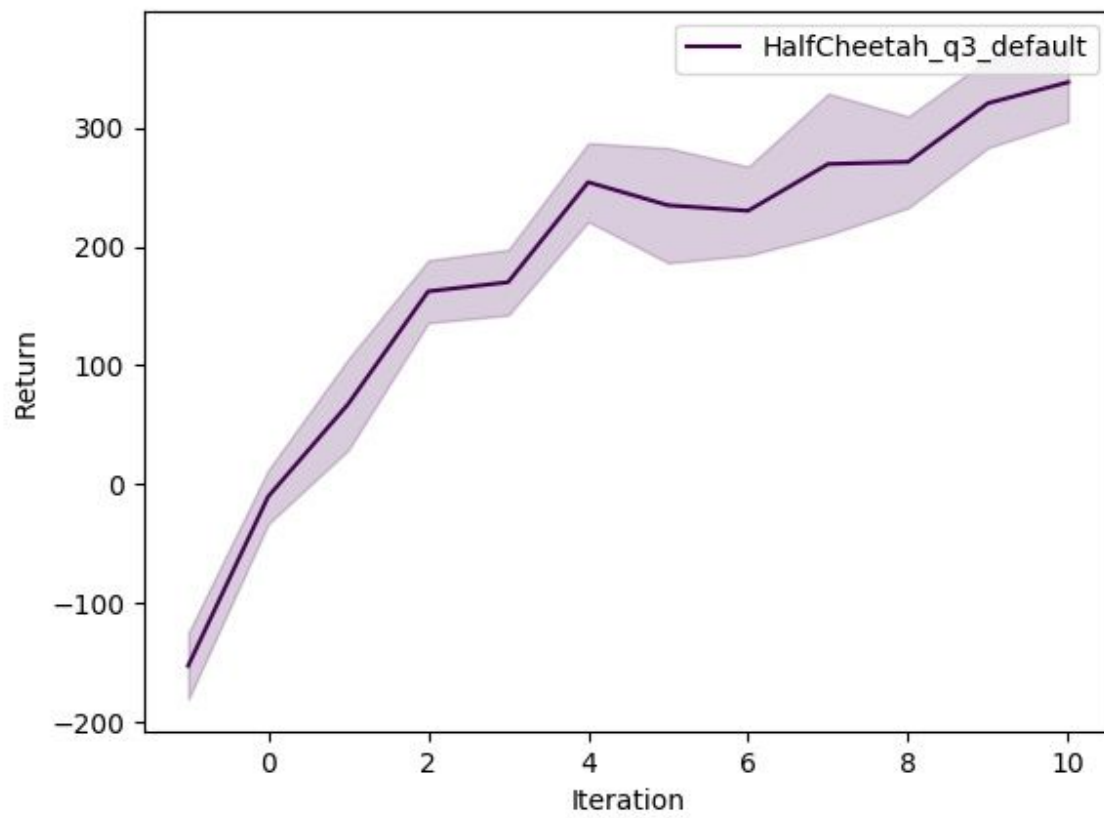
ReturnStd 28.4027

TrainingLossFinal 107.842

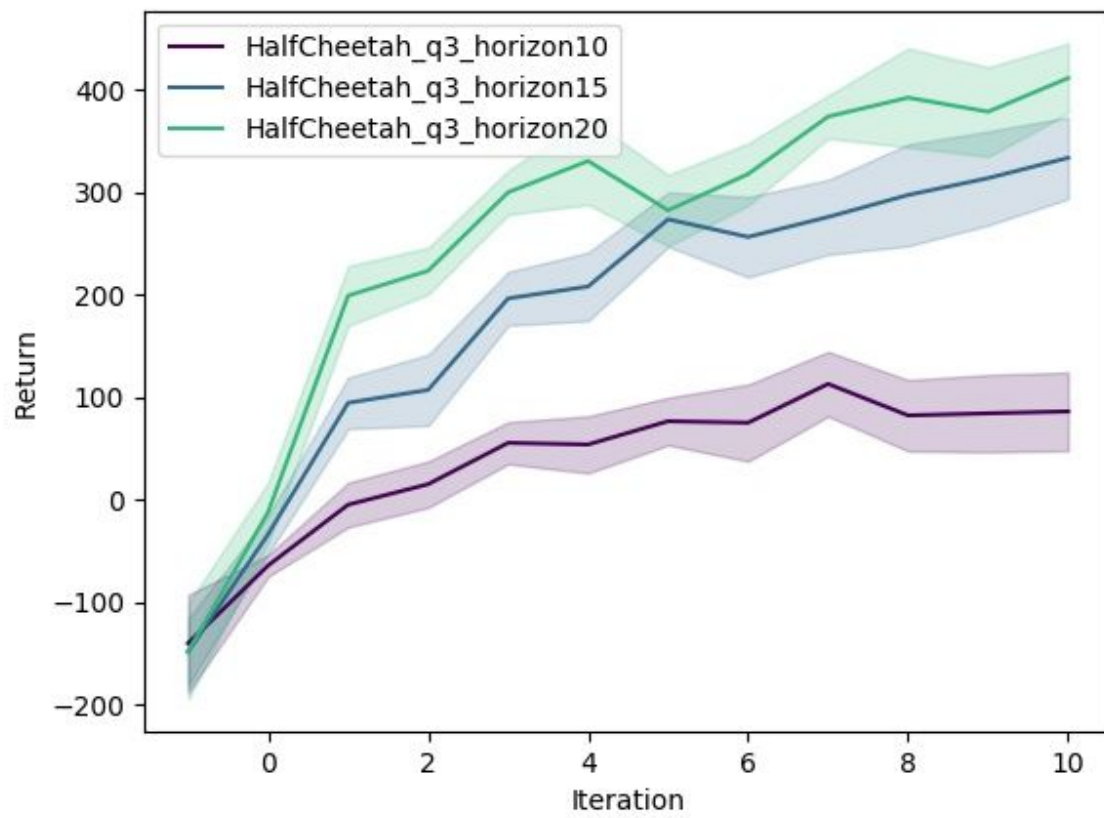
TrainingLossStart 5269.59

-----

### Problem 3a



### Problem 3b



Figure

