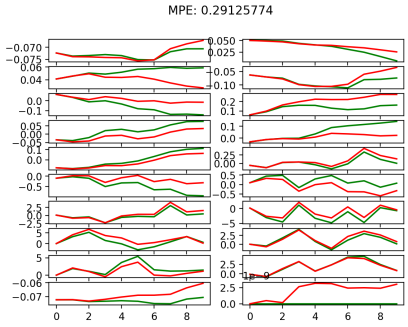
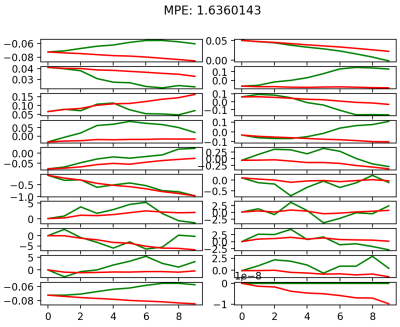
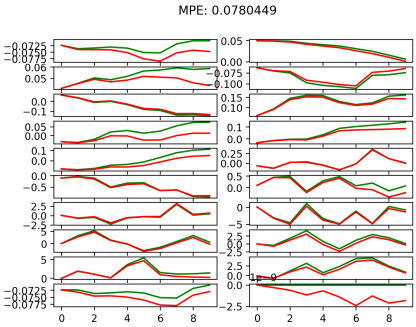
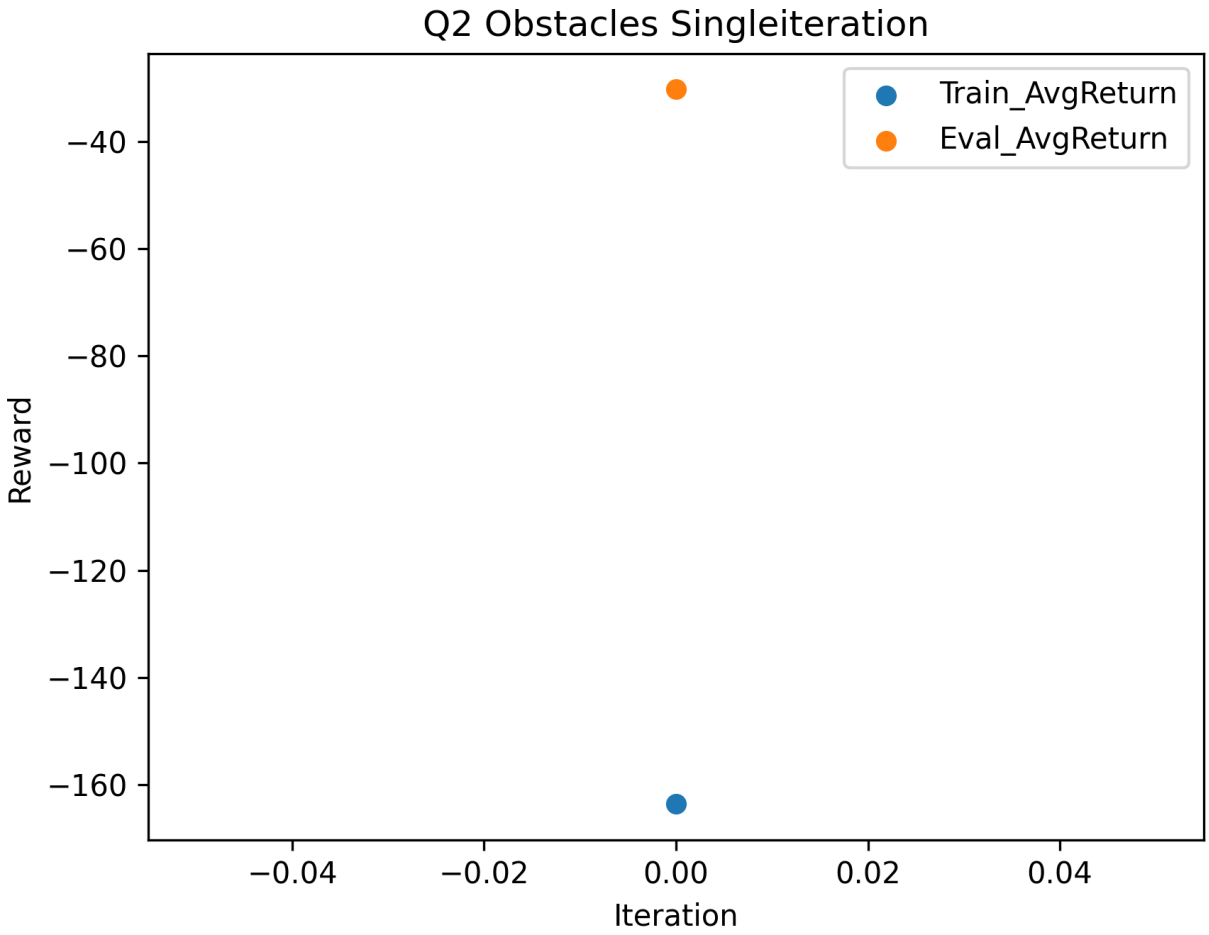
CS285 HW4 Report

Q1 Cheetah

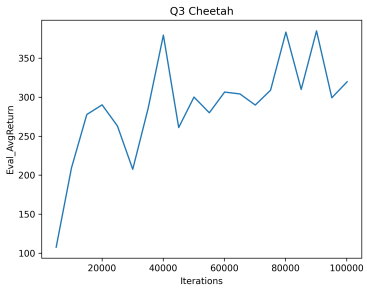
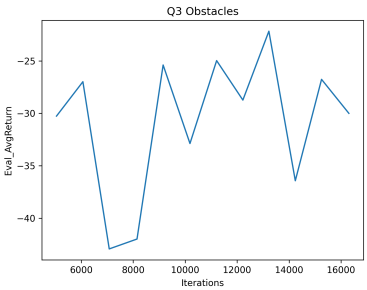
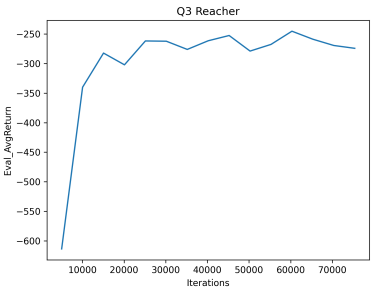
Cheetah-n500-arch1x32	Cheetah-n5-arch2x250	Cheetah-n500-arch2x250
		

Based on MPE comparisons, the Cheetah with 500 (highest) number of agent train steps per iteration and more complicated model performs the best. This makes sense, as a simpler model may not capture the neccessary factors to model the transition (first model), and not making enough train steps will not fully allow the model to learn the patterns (second model).

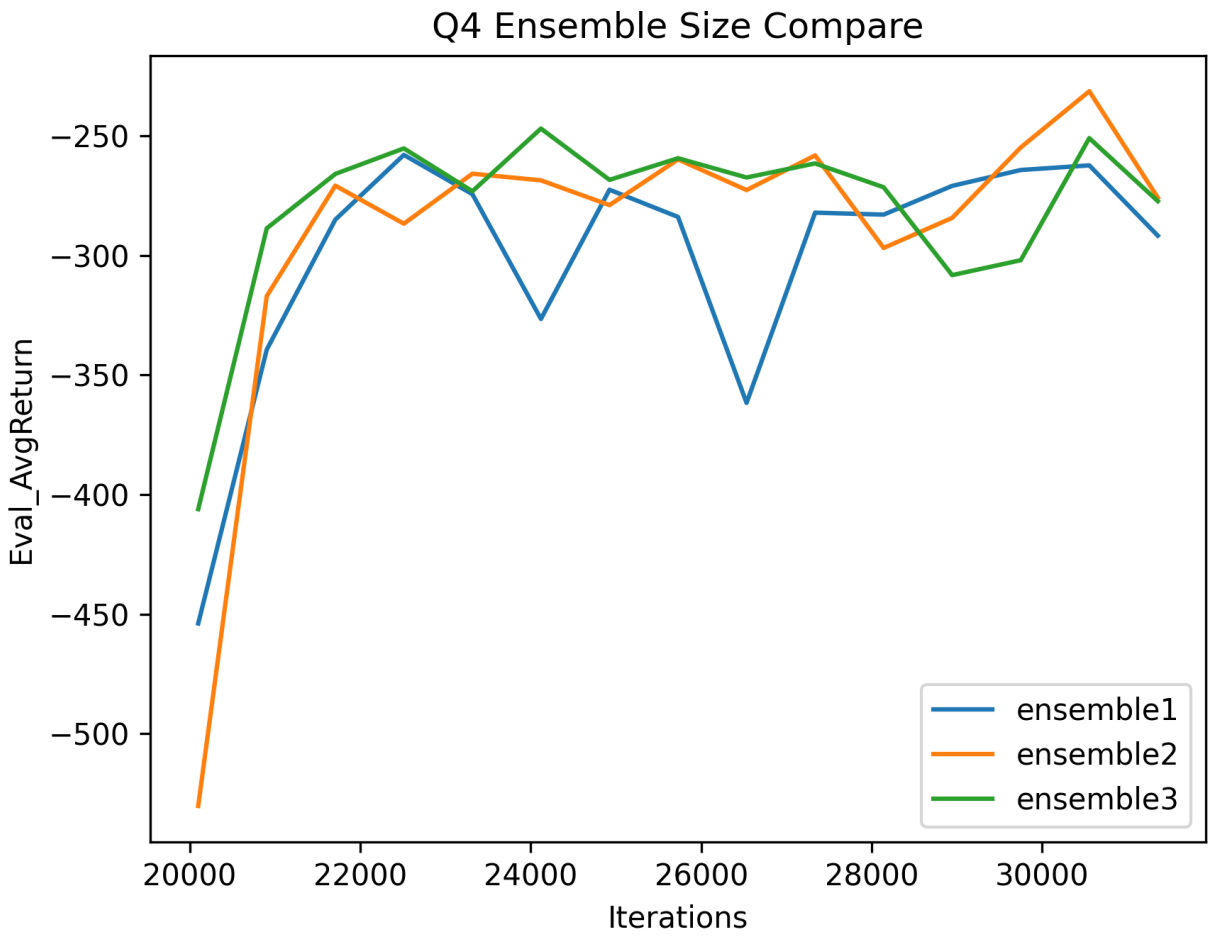
Q2 Action Selection using Random Shooting Check

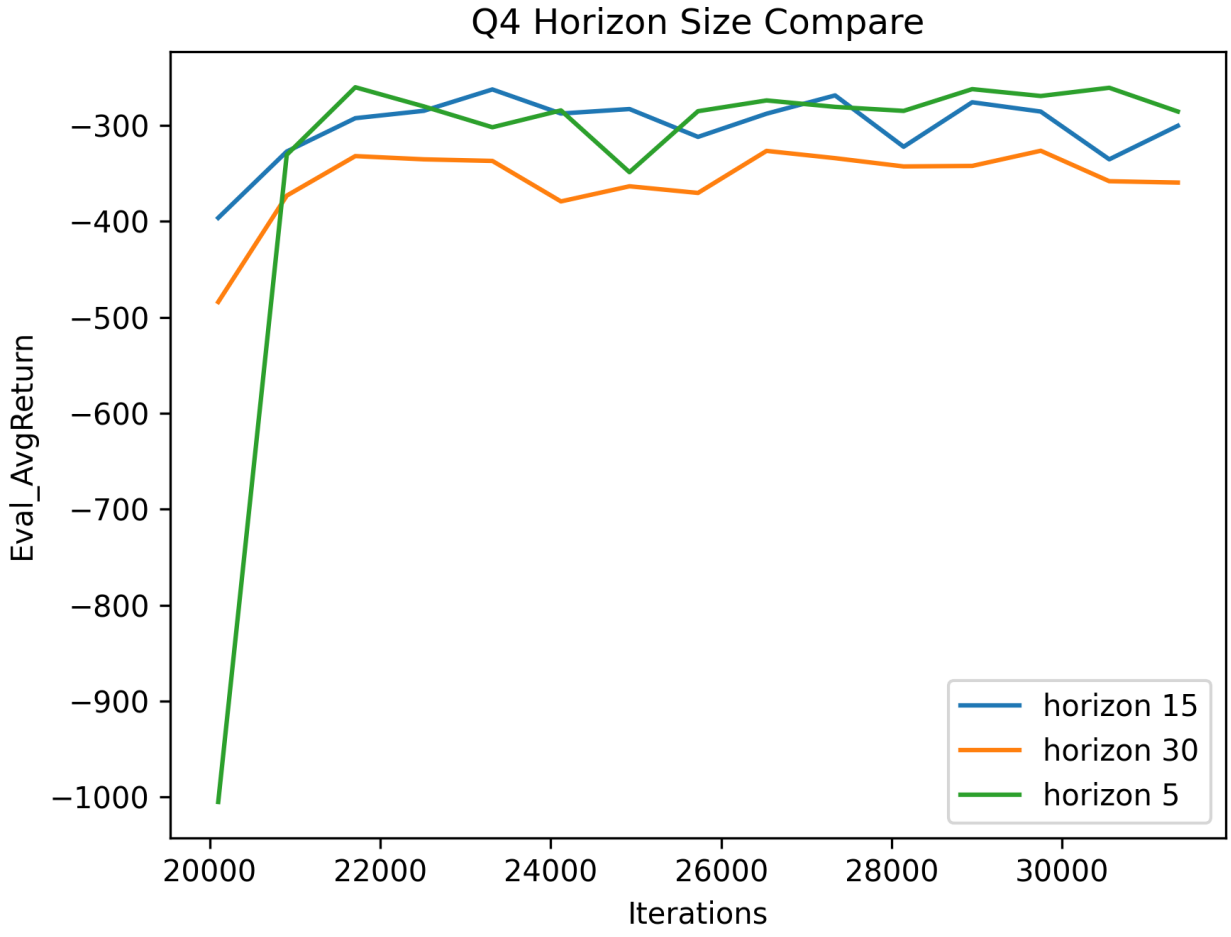


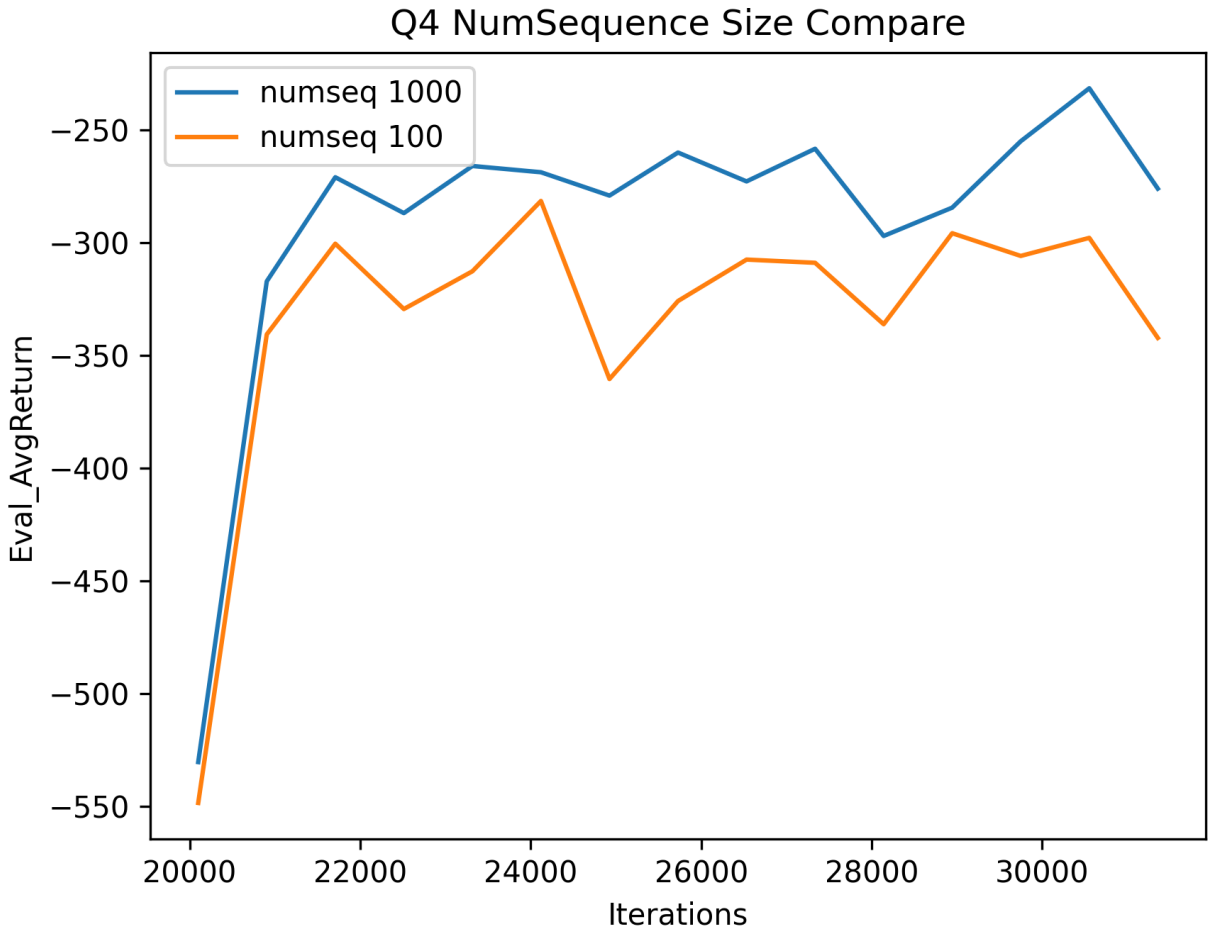
Q3 Action Selection Experiments

Cheetah	Obstacles	Reacher
		

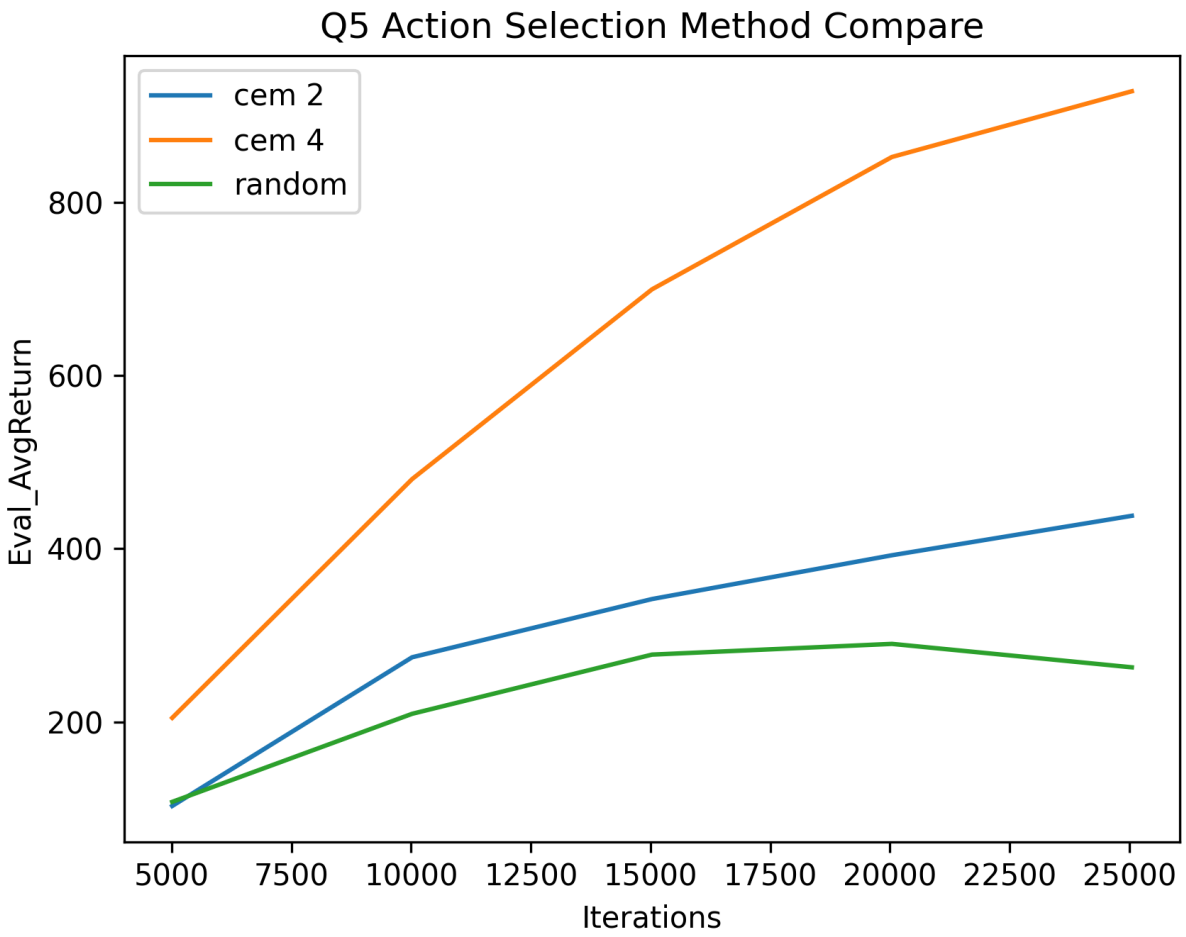
Q4 MBRL Hyperparameters


Q4 Ensemble Size Compare
Eval_AvgReturn
−250
−300
−350
−400
−450
−500
20000
22000
24000
26000
28000
30000
Iterations
ensemble1
ensemble2
ensemble3


Q4 Horizon Size Compare
Eval_AvgReturn
−300
−400
−500
−600
−700
−800
−900
−1000
20000
22000
24000
26000
28000
30000
Iterations
horizon 15
horizon 30
horizon 5

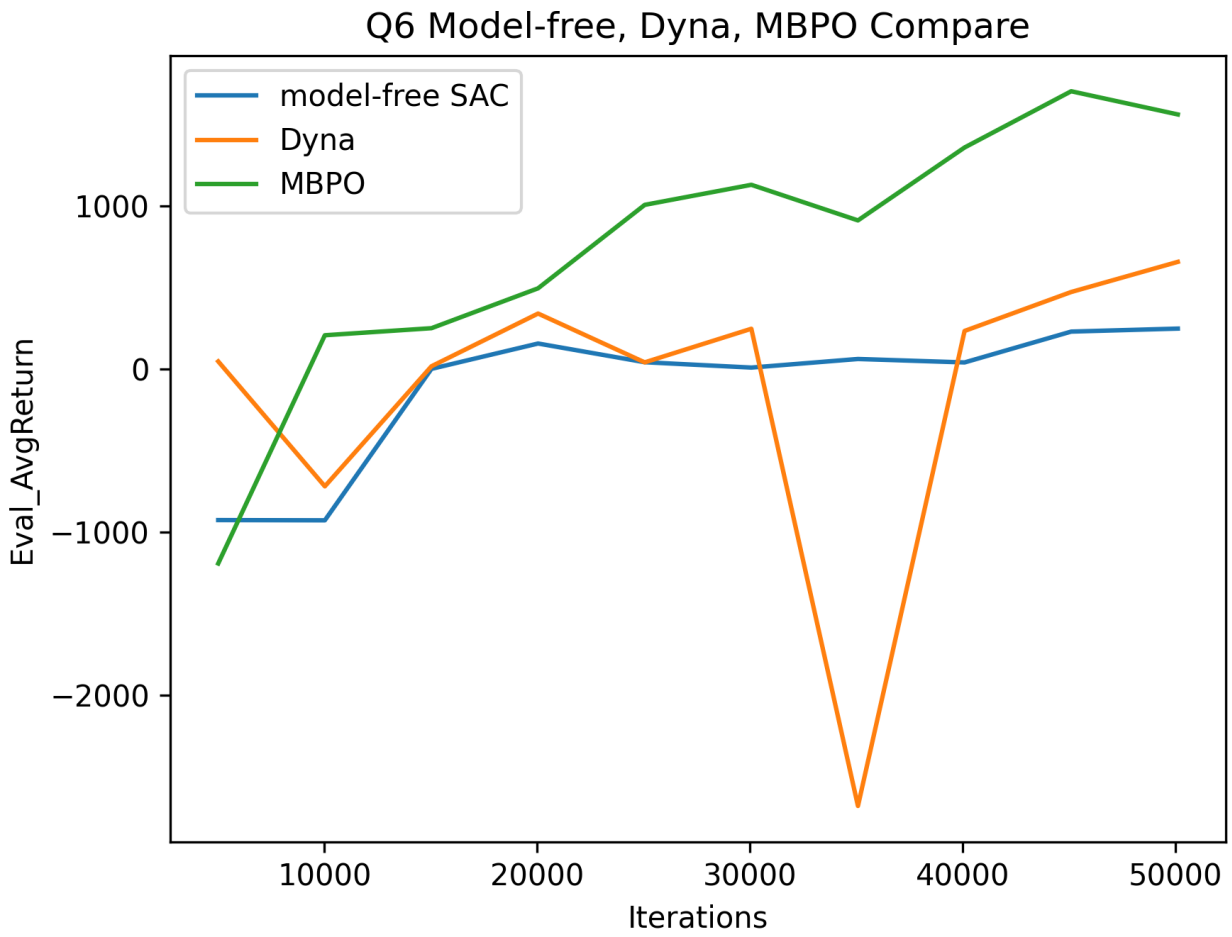


Q5 MBRL using CEM



The higher the cem iteration, the better the performance. It makes sense as more iteration of cross-entropy method allows to select better/more accurate actions, where as random actions perform the worst. It's basically a trade off between time and performance.

Q6 MBPO



Model-based RL definitely helps with the performance. By rolling out a reasonable amount of steps from the model during training, it allows the policy to learn and gain more experience, and thus performs better for the task.