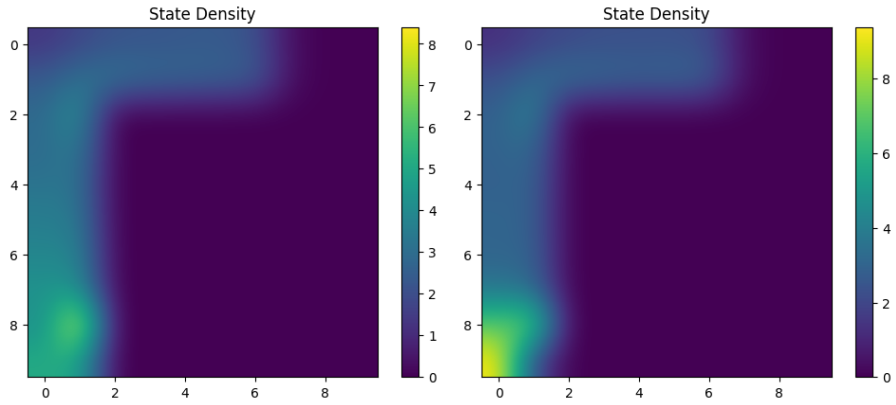
Homework 5

Juanwu Lu (3037432593)

(M.Sc. Civil Engineering, UC Berkeley)

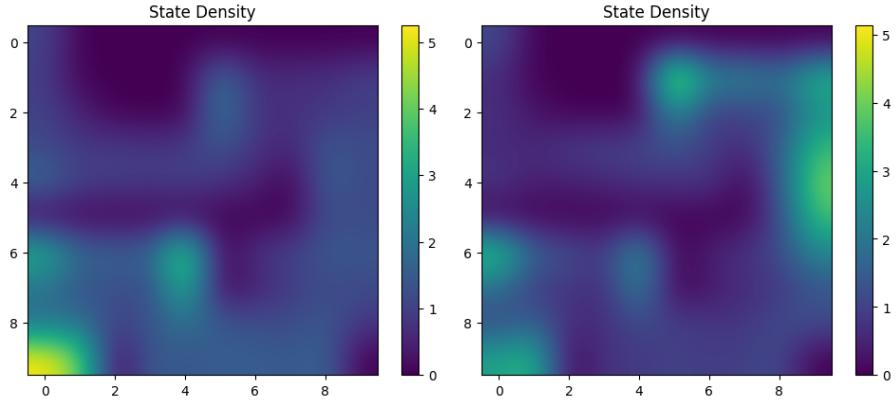
Problem 1 "Unsupervised" RND and exploration performance

Part 1 Results



(a) (b)

Figure 1. Results from PointmassEasy Environment: (a) Random exploration with epsilon-greedy; (b) Exploration with RND. State density of from the two algorithms are similar but RND unexpectedly has a denser density around the lower left corner (origin).



(a) (b)

Figure 2. Results from PointmassMedium Environment: (a) Random exploration with epsilon-greedy; (b) Exploration with RND. State density of from the two algorithms are similar but RND has a more uniformly distributed density than the random exploration one.

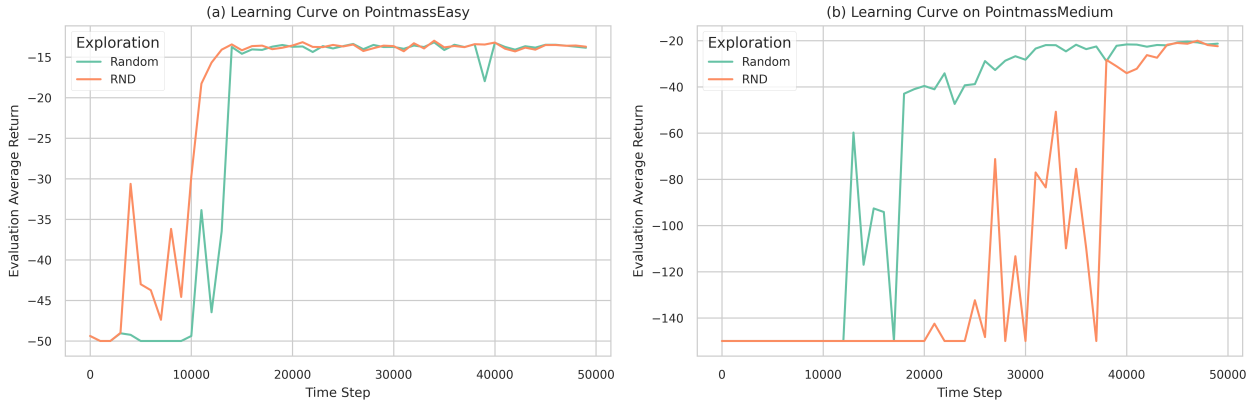


Figure 3. Learning curve from the two environments. RND reaches higher score faster than epsilon-greedy on the PointmassEasy environment, while slower on the PointmassMedium environment.

Part 2 Results

Problem 2 Offline learning on exploration data

Part 1 Compare CQL to DQN on the `medium` environment

Part 2 Ablation study on amount of exploration data

Part 3 Ablation study on the regularizer weight α

Problem 3 ”Supervised” exploration with mixed reward bonuses

Problem 4 Offline Learning with AWAC

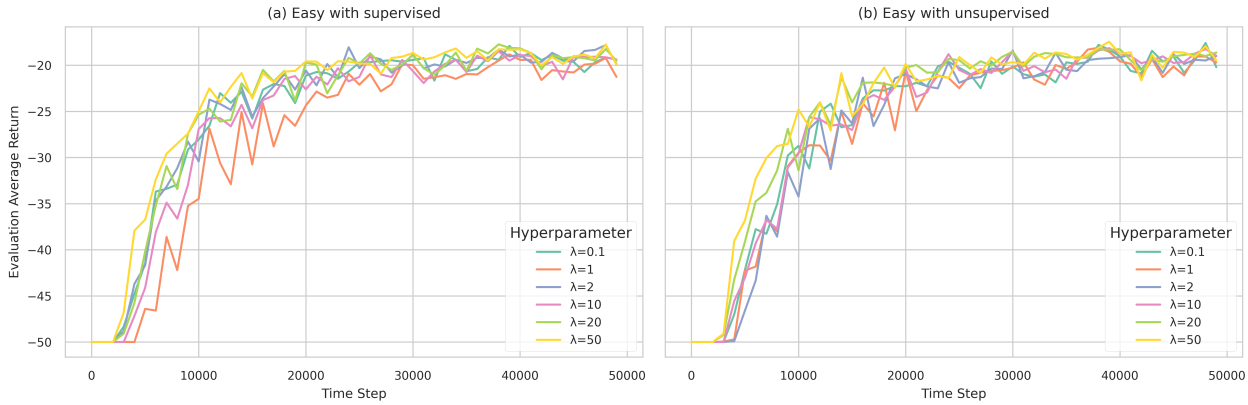


Figure 4. Learning curve from `PointmassEasy` environment: (a) supervised algorithm with different λ settings; (b) unsupervised algorithm with different λ settings. Explorations with supervised and unsupervised perform quite similar in this environment, which I think is partly due to the simplicity of the task.

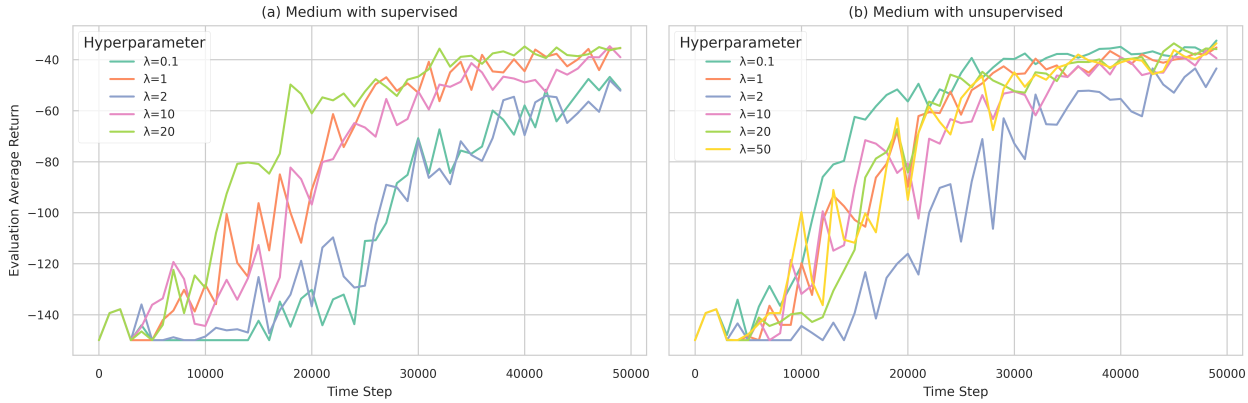


Figure 5. Learning curve from `PointmassMedium` environment: (a) supervised algorithm with different λ settings; (b) unsupervised algorithm with different λ settings. Unexpectedly, explorations with unsupervised perform a bit better overall in this environment with faster reaching a higher return. The best λ setting for supervised and unsupervised RND are $\lambda = 20$ and $\lambda = 0.1$, respectively.

Problem 5 Offline Learning with IQL