6.884 HW1 Seremet Vlad seremetv@

**Problem 1**

**Deliverables:**

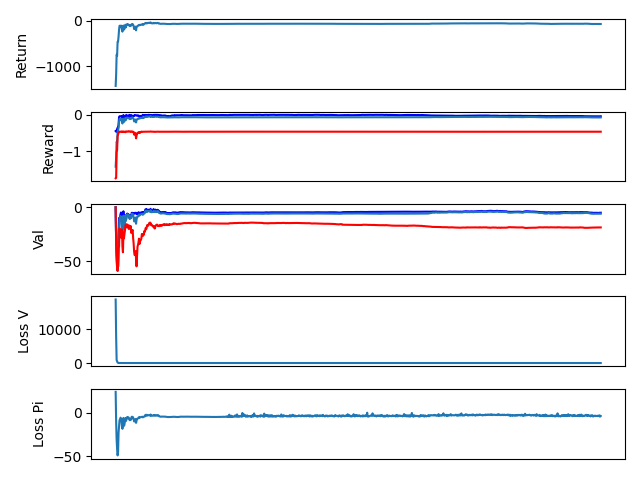
* Source code with a working implementation of PPO (30 pts).

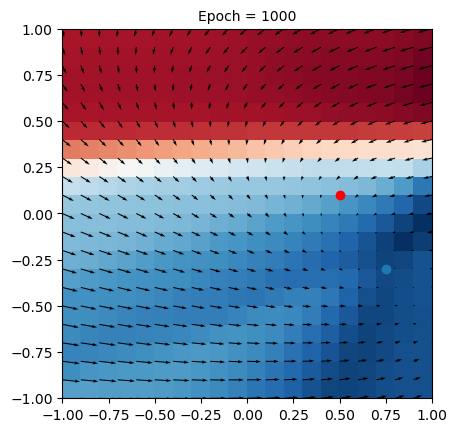
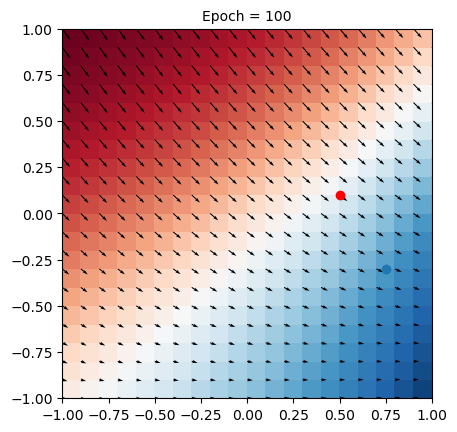
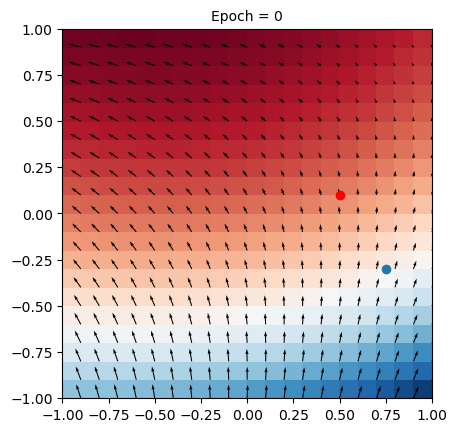
<https://github.com/shervlad/hw1>

* A mathematical description of the reward function (15 pts).

*reward = -distance*, - distance between the gripper and the goal position.

* Training plot showing rewards as a function of time. Report the average performance over 3 random seeds (15 pts).





* Video showing evaluation of policy (5 pts).

*Video can be found in hw1/videos/*

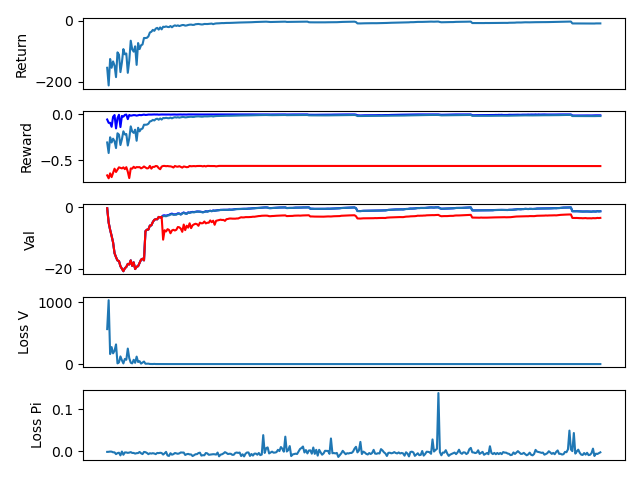
**Problem 2**

**Deliverables**

* Answer to part 1 (10 pts).

*The agent succeeds.*

* Training plot showing reward as a function of time for part 1. Report the average performance over 3 random seeds (15 pts).



* Video showing evaluation of policy for part 1 (5 pts).

*Video can be found at hw1/videos/reacher\_wall.webm*

**Problem 3**

**Deliverables**

* A mathematical description of the reward function (15 pts).

reward = 100 - 50\*(3\*dist + dist2 + ang/5)

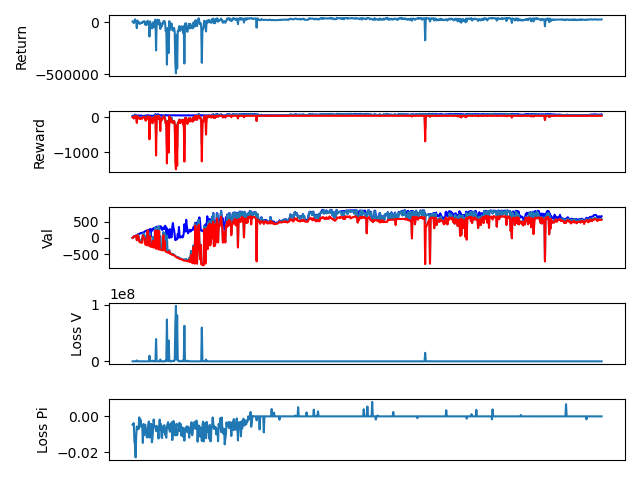
dist = distance between gripper and box

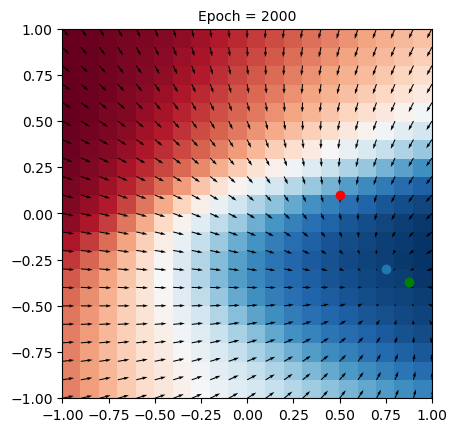
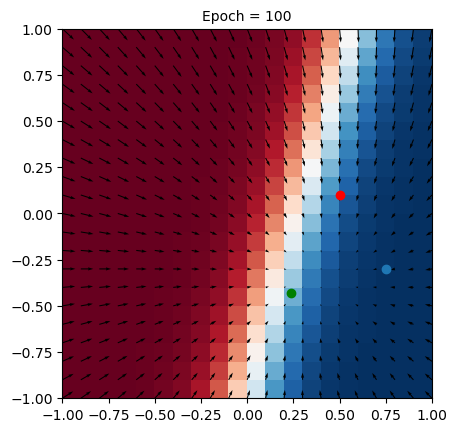
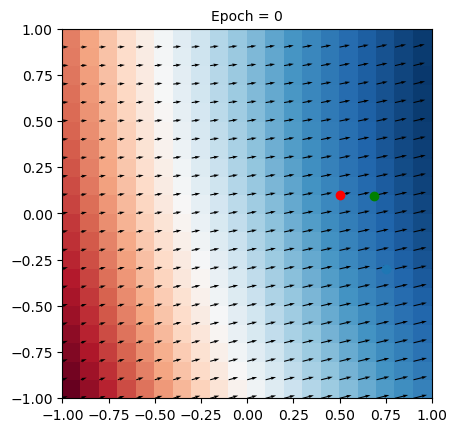
dist2 = distance between box and goal

ang = the angle between the (gripper -> box) vector and (box -> goal) vector.

When this is 0, the gripper, box, and goal are on the same line

* Training plot showing rewards as a function of time. Report the average performance over 3 random seeds (15 pts).





* Video showing evaluation of policy (5 pts).

*Video can be found at hw1/videos/pusher.webm*