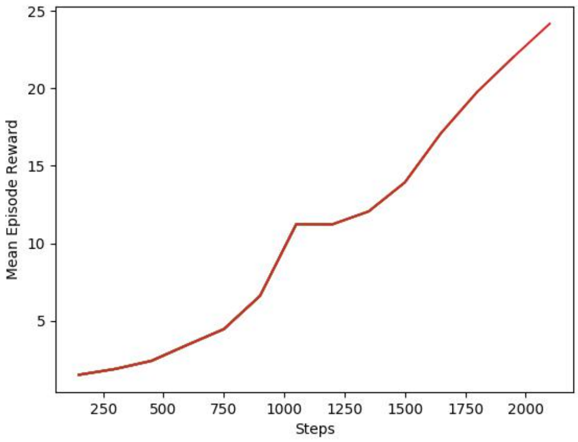
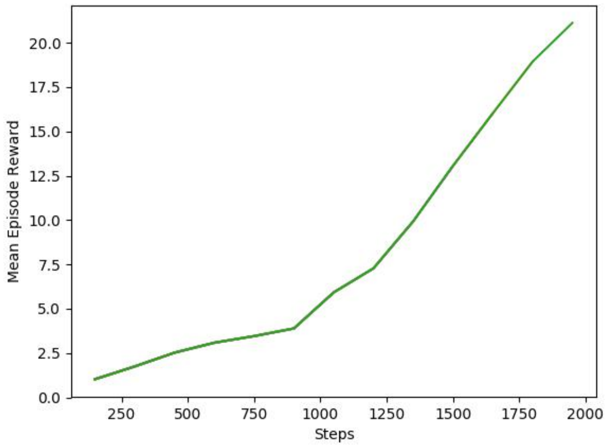
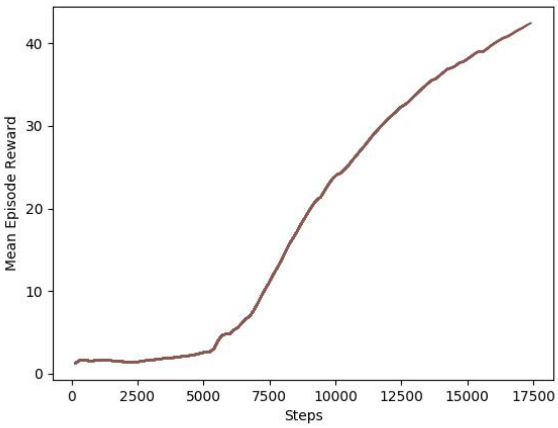
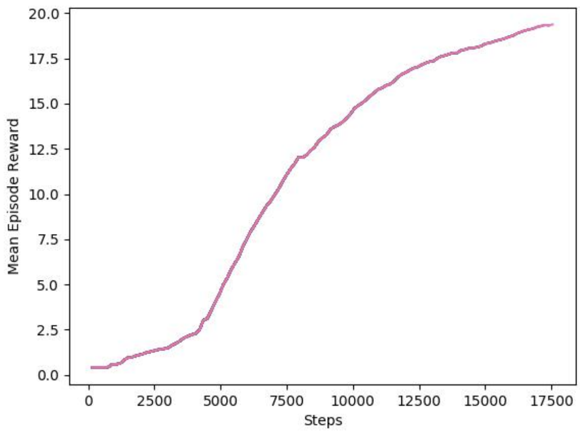
Comparisons of RL for *reacher* with initialized and non-initialized policy

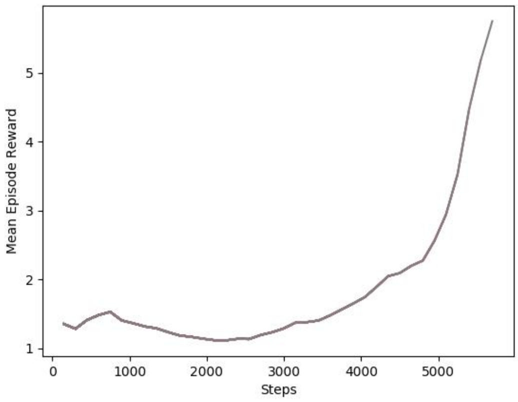
Below are 5 pairs of test (same settings) with different time scales,on the left is with initialization policy from inverse kinematics (just called initialized) and on the right is with general random initialization (just called non-initialized). Generally initialized Reinforcement Learning surely performs better than non-initialized one.

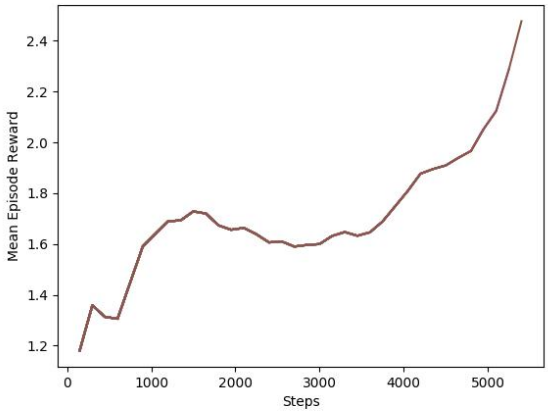


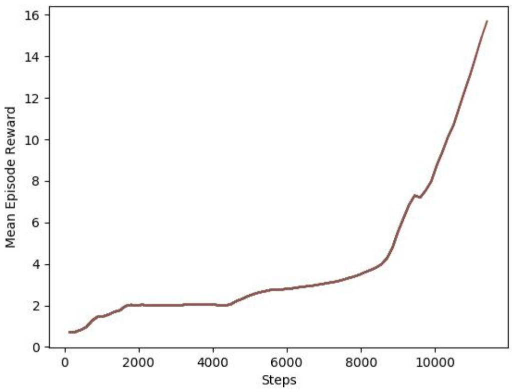


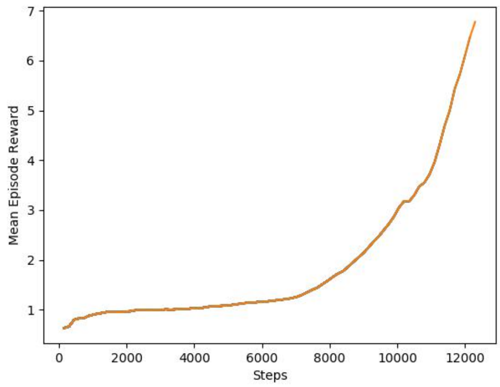


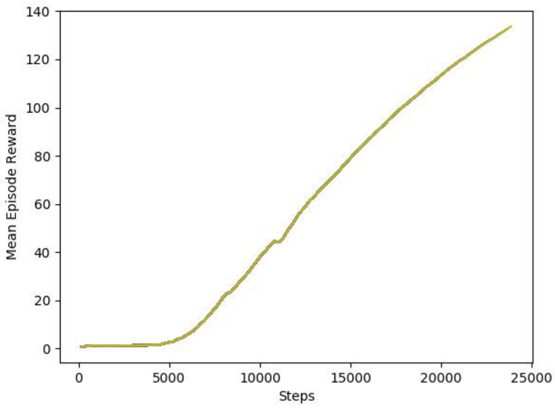


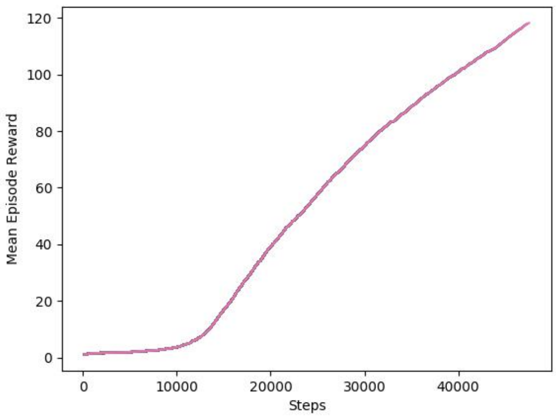












One worth noticing point is that the initialized RL doesn’t show a significant better performance than non-initialized one in the starting stage in training, like first 300 steps. This does not meet my expectation quite much. I guess the the initialized policy only improves the performance a little bit at the beginning, but the effects accumulate after a long run.

I think the limit of improvement with initialized policy for RL mainly lies in the accuracy of supervised learning policy. In my test, the accuracy of supervised learning policy is not very high through my observation, as it needs to generate to all possible goal positions.