

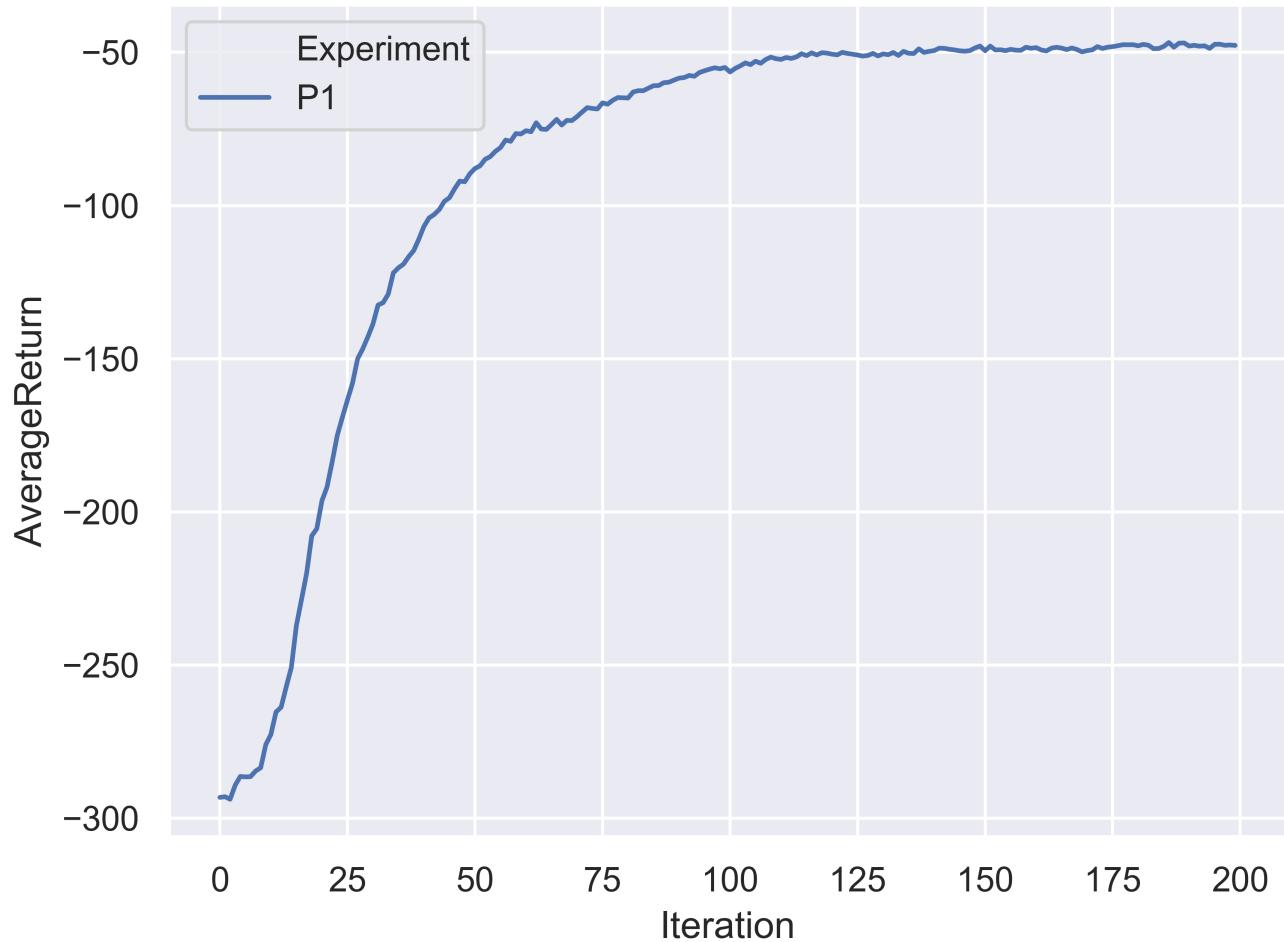
Berkeley DRL HW5

Meta Learning

Anton Makiievskyi

PROBLEM 1:

The policy achieved the reward of -50 when ran with default hyperparameters as was expected



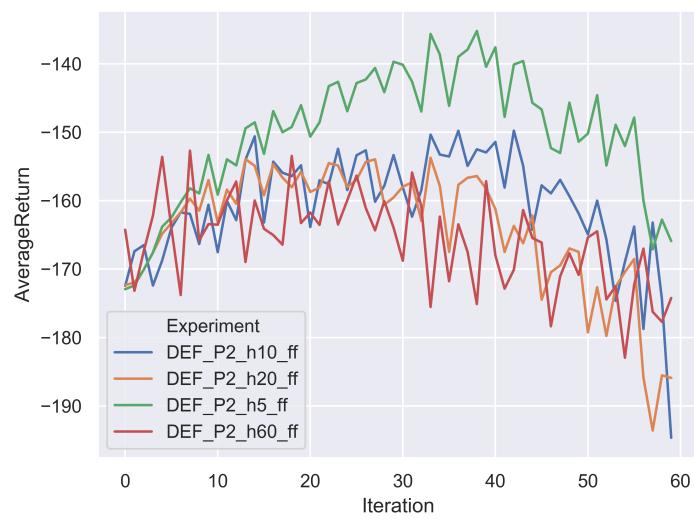
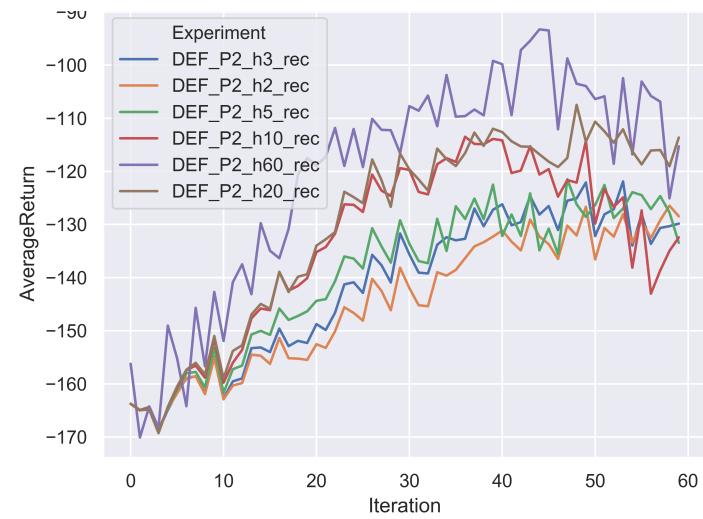
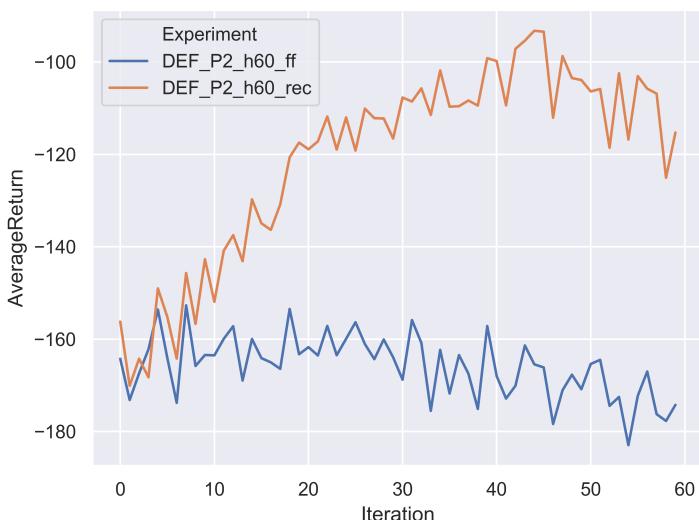
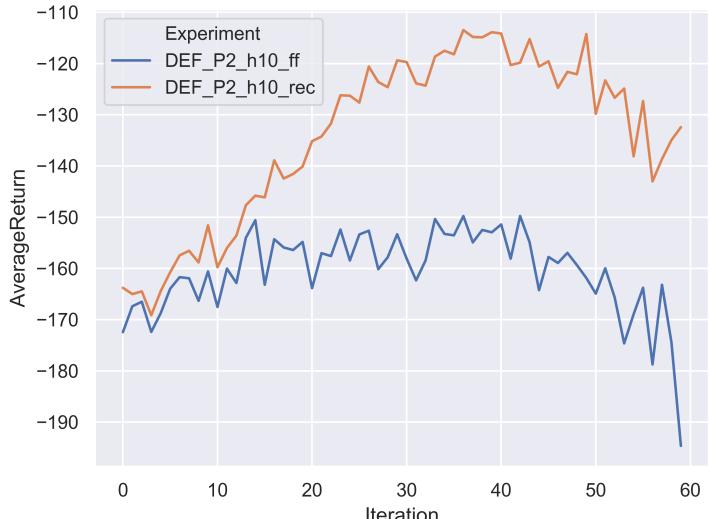
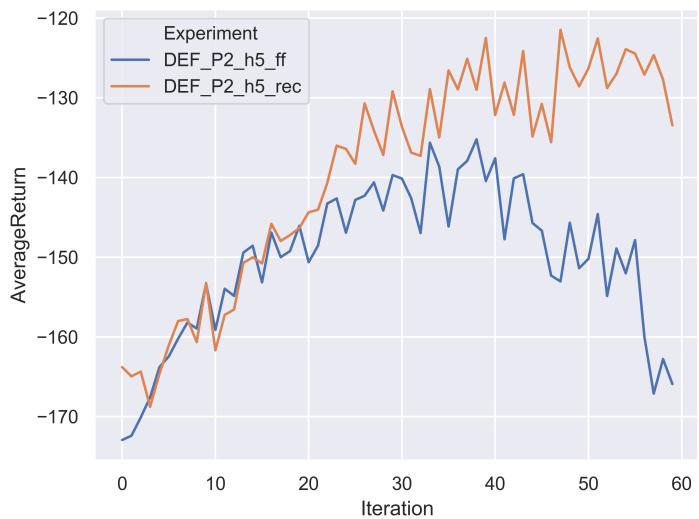
PROBLEM 2:

Below is the comparison of the feed forward network to the recurrent network on the meta learning task, where the policy is supposed to inquire the goal from the history of rewards.

The parameters were chosen such that every agent has ~28k trainable parameters.

The experiment was run for history lengths of [5, 10, 20, 60] and additional for history length [2,3] for just the recurrent case.

Agent with recurrent architecture showed much better performance in all cases showing considerable improvement even with the history size of 2.. The feedforward agent barely improved in most cases.



PROBLEM 3:

In the checkerboard split between training and validation sets, the agent generalized very well even in the biggest granularity setting where the training and testing goals are distributed as shown on the picture.

