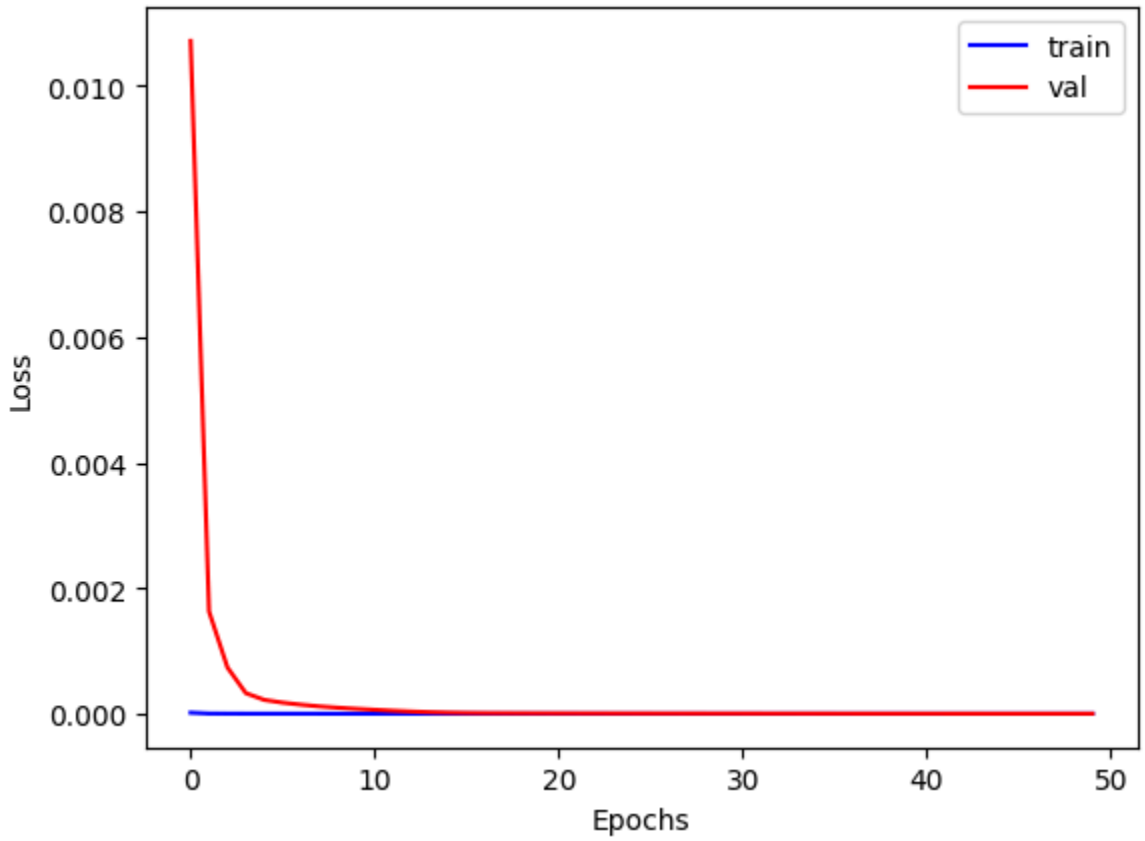
Assignment 5: Model Based RL

Submission Date: 11/26/24

Part 1: StableBaselines3 PPO Expert [9 points]

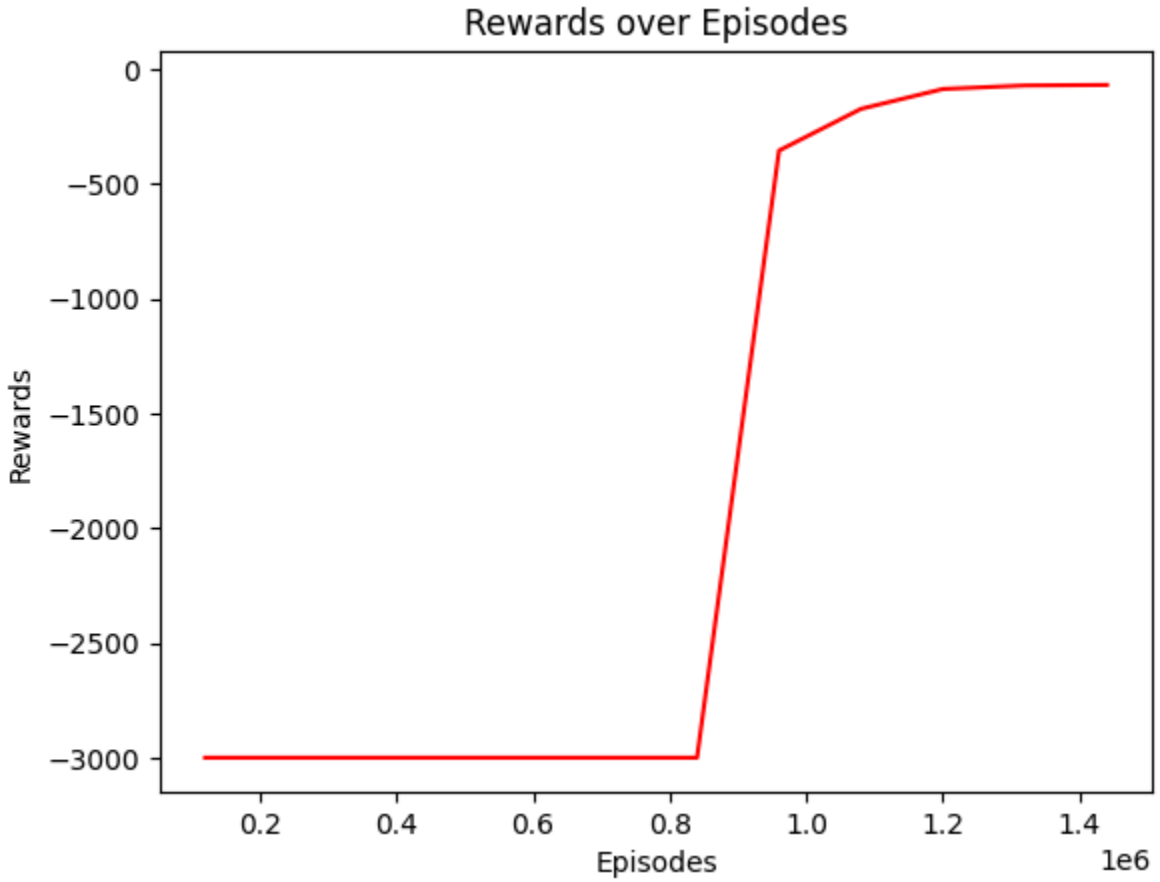
Final reward: -99.55000305175781

Part 2: World Model Learning [26 points]



Final validation loss: 9.22535220982341e-08

Part 3: Learn PPO Agent on Learned Model [9 points]



Average reward, MountainCar-v0: -69.5999984741211
Average reward, WorldModelMountainCar: -198.5500030517578

Since the "strategy" for both models is similar, the model seems to generalize across the different scenarios well. Interestingly, this model actually outperforms the original model, likely due to the difference in optimization strategy and the optimality of the model. This is also probably influenced by a change in the device used and seed midway through the process, since I had the "stuck at -3000" issue.