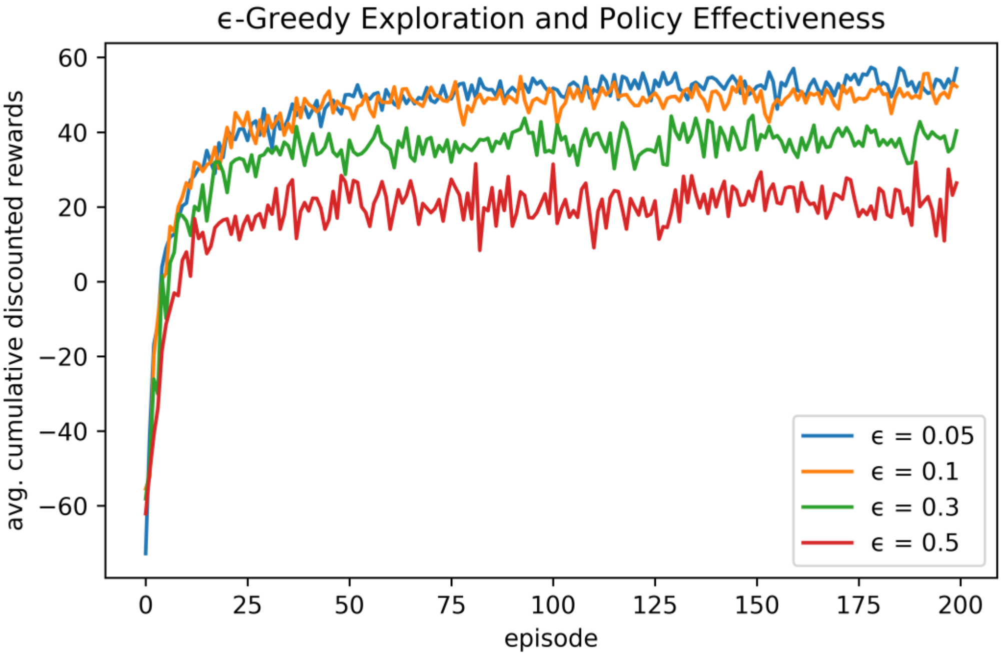
CS9670: Lab 2 Report

The RL (Q-learning) solver code was run on `test_rl_maze.py`. The results for the prompts are given below.

1. *Produce a graph where the x-axis indicates the episode # (from 0 to 200) and the y-axis indicates the average (based on 100 trials) of the cumulative discounted rewards per episode (100 steps). The graph should contain 4 curves corresponding to the exploration probability epsilon=0.05, 0.1, 0.3 and 0.5 . The initial state is 0 and the initial Q-function is 0 for all state-action pairs.*

> We get:



2. *Explain the impact of the exploration probability epsilon on the cumulative discounted rewards per episode earned during training as well as the resulting Q-values and policy.*

> From the graph above, it can be seen that increasing exploration probability (ϵ) reduces the converged policy's effectiveness – estimated by the net discounted reward. This is generally true, although for some intermediate value of ϵ we could have a better result compared to lower exploration levels (ϵ ≠ 0). This can depend on the problem and its parameters. However, this trend is reasonable since higher values of ϵ cause lesser exploitation of actions that are already known to be better over others with some certainty, and it therefore stunts the accrual of more rewards.