

# ROB537: HW2

Ammar Kothari

## 1 Introduction

## 2 Bandit Problem

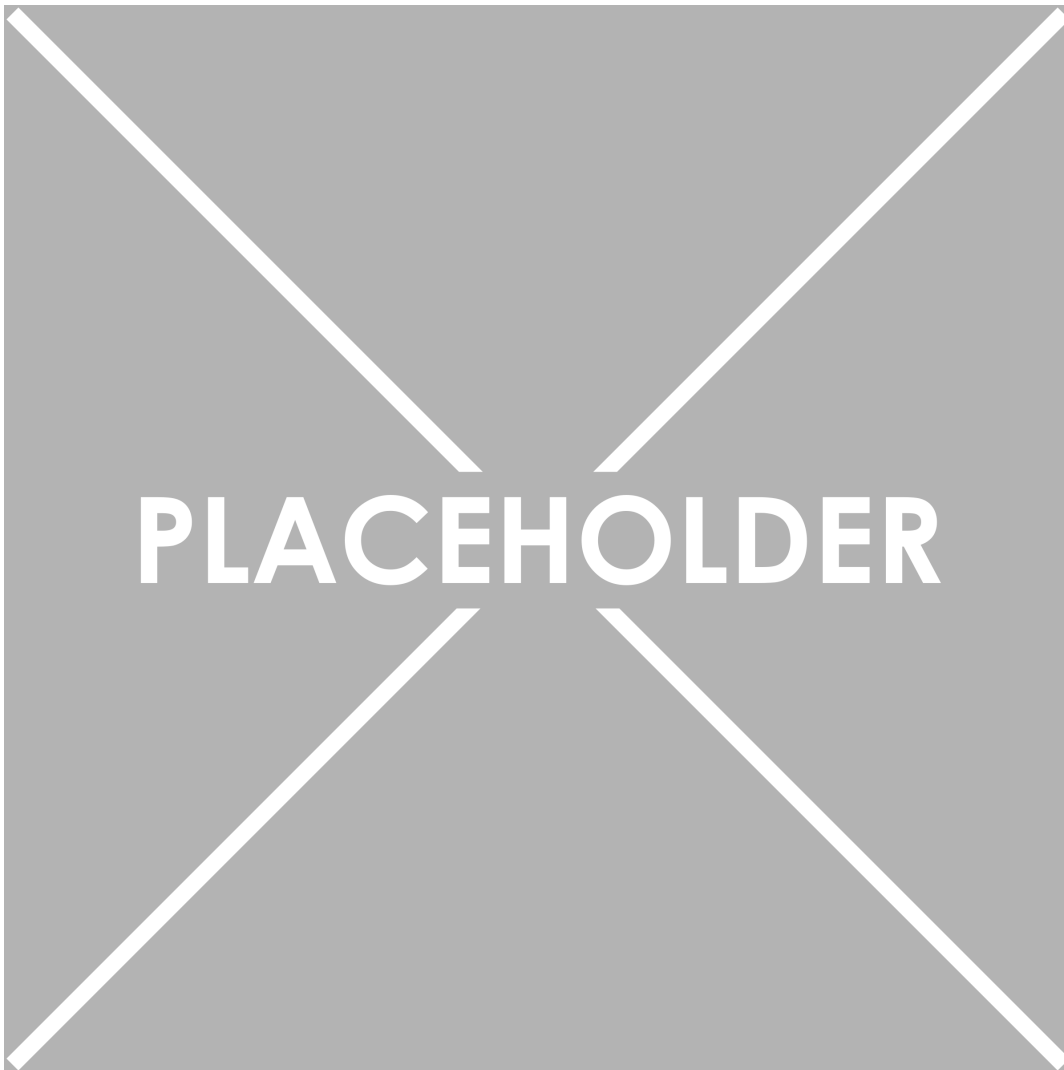


Figure 1: Comparison of performance between greedy and epsilon greedy selection over 10 steps for an action-value learner on a multiarmed bandit problem.

## 3 Conclusion

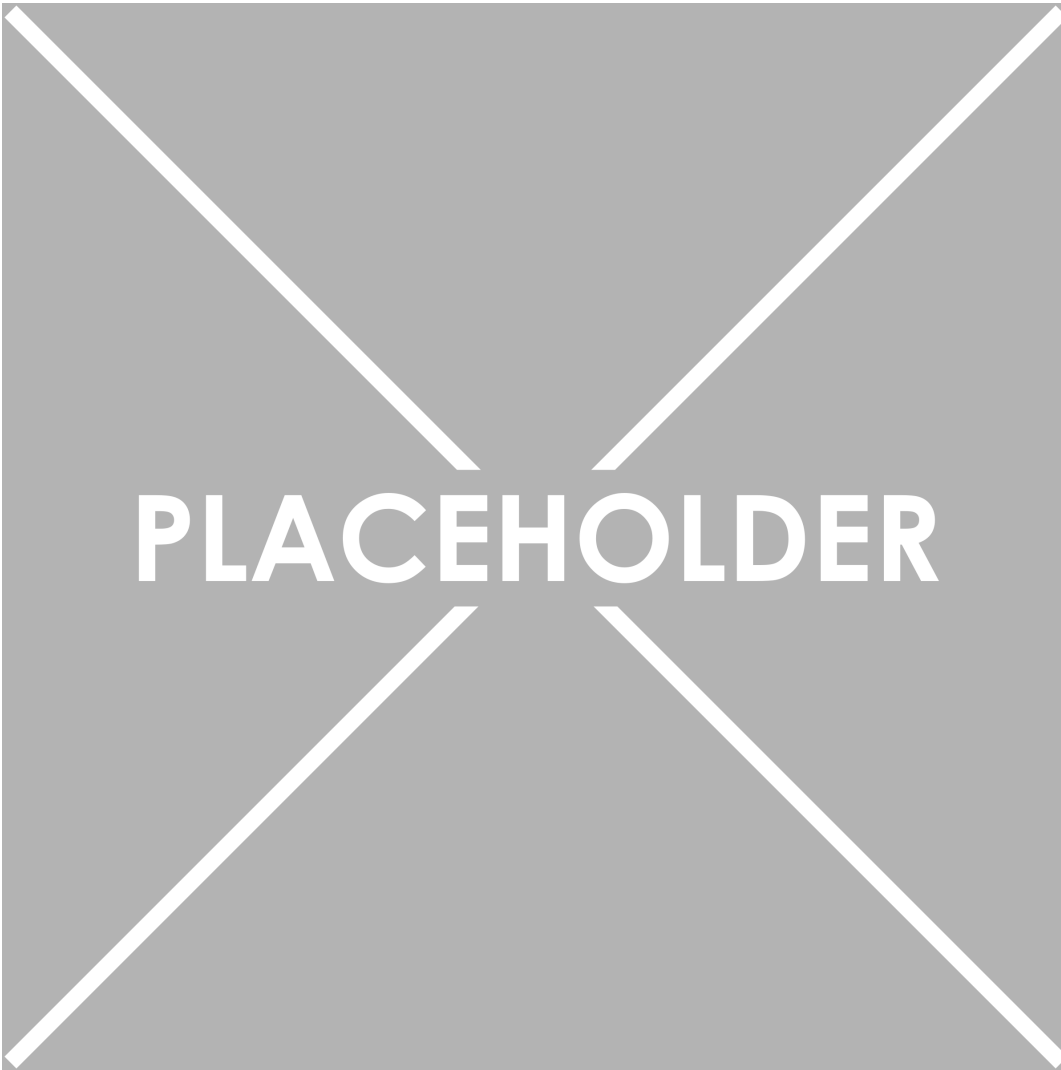


Figure 2: Comparison of performance between greedy and epsilon greedy selection over 100 steps for an action-value learner on a multiarmed bandit problem.

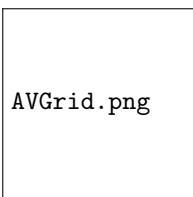


Figure 3: Action value table for epsilon greedy agent for 20 steps.

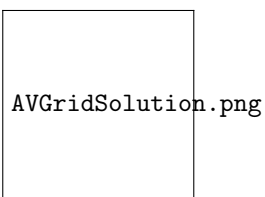


Figure 4: Action value table quiver plot for epsilon greedy agent for 20 steps. Arrows are weighted average of the best action at that state.

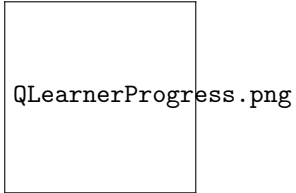


Figure 5: Q learner learning progression.

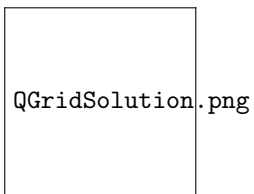


Figure 6: Q table quiver plot for epsilon greedy agent for 20 steps. Arrows are weighted average of the best action at that state.