

$\epsilon$ -greedy optimal policy with gamma = 0.95

