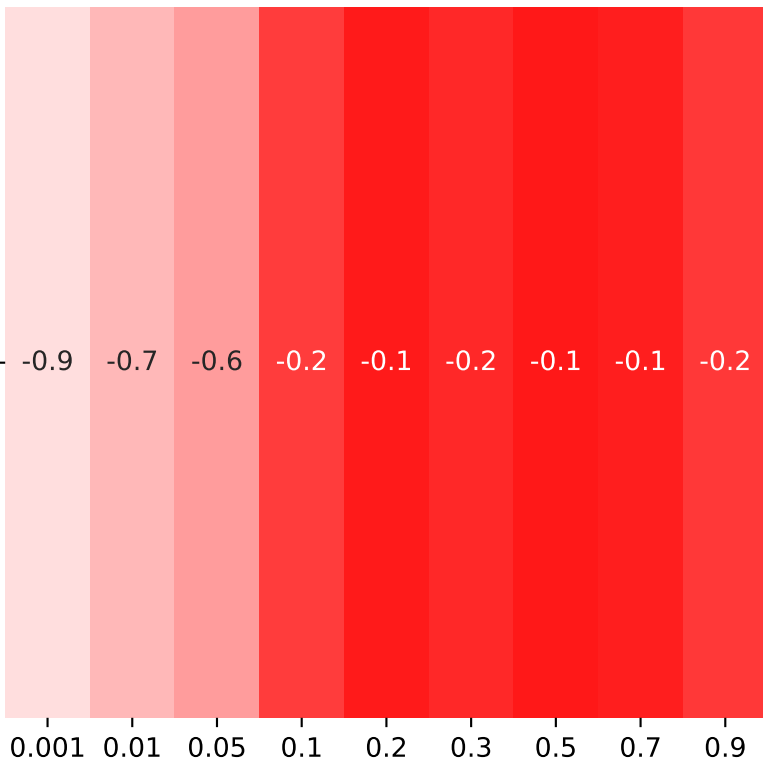


ϵ -greedy optimal policy - last 500 steps

gamma

0.95



epsilon

0

-1

-12