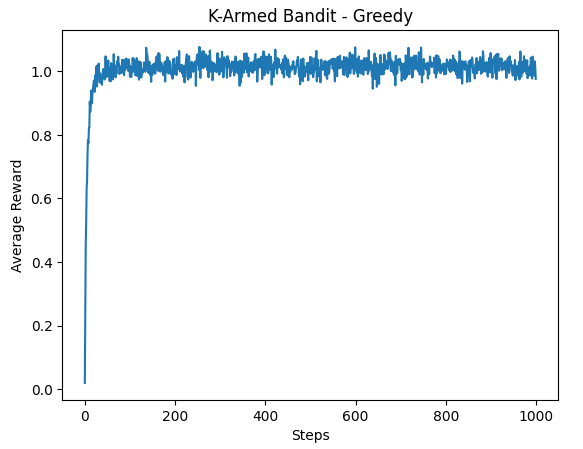
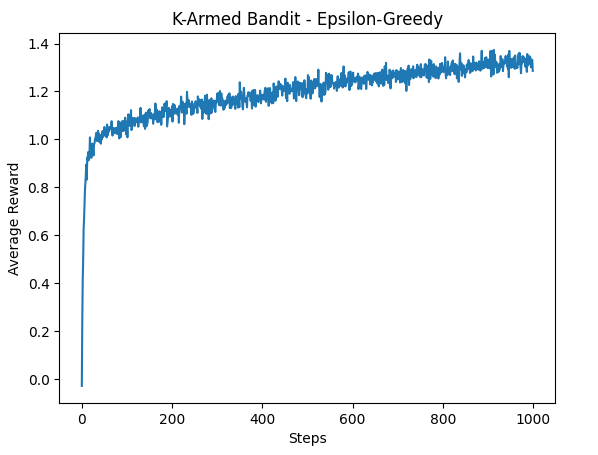


1 pav.k = 10, epsilon =0.1



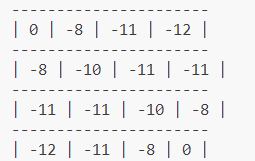
2 pav. k = 10, epsilon =0



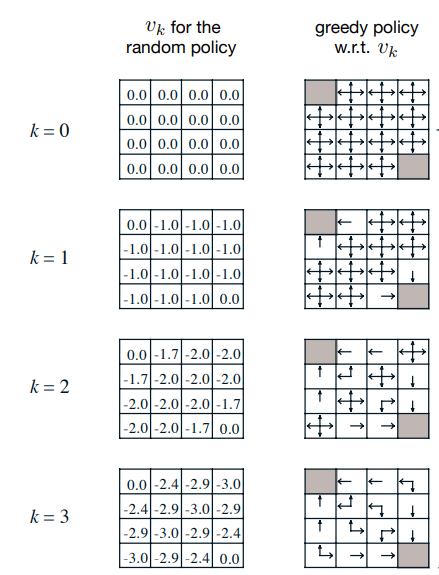
3 pav. k = 10, epsilon =0.01



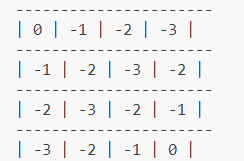
pav. 4 Sample average vs constant step action value changing environment



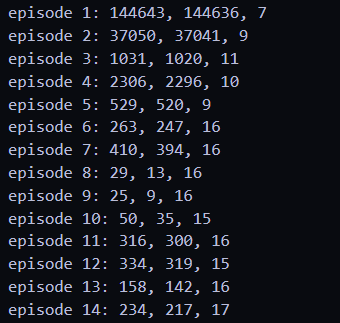
pav. 5 policy ivertinimas

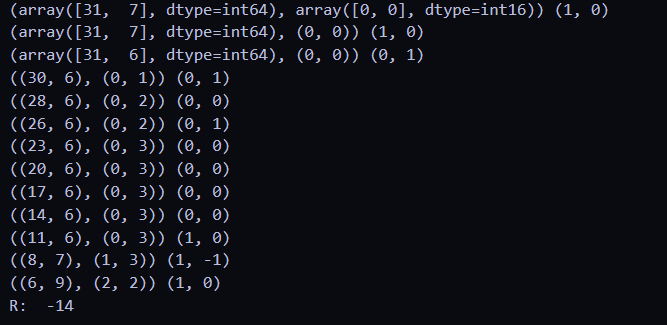
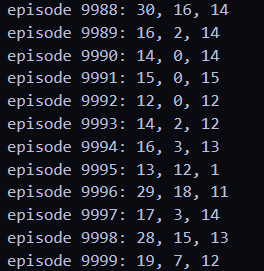


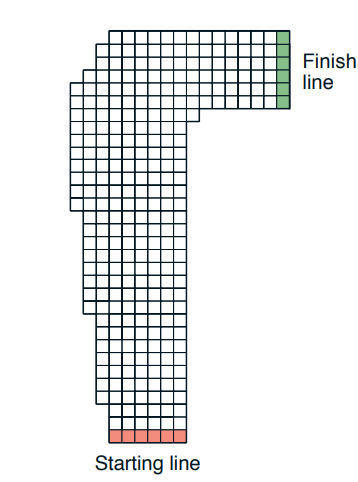
Agentas kuris turi šansą judėti į bet kurią pusę įvertins optimalią policy pagal dabartinės būklės aplinkuj esančias būkles



Jei agentas judėtų optimaliai kiekvienoj stadijoj





Monte carlo metodai, pirmos serijos pries paskutinias, agento judejimas aplinkoj kai laikomasi optimal policy  


pav. 6 Monte carlo environment