به نام خدا

یادگیری تقویتی　　تمرین اول　　استاد:دکتر هراتی-دکتر غیاثی

در این تمرین قصد داریم مساله ای را که در کلاس مطرح شد در مورد 10 armed bandits را برسی کنیم .

در این تمرین قصد داریم حالت های مختلف ان را به ازای λ های مختلف برسی و شکل ان را رسم کنیم .

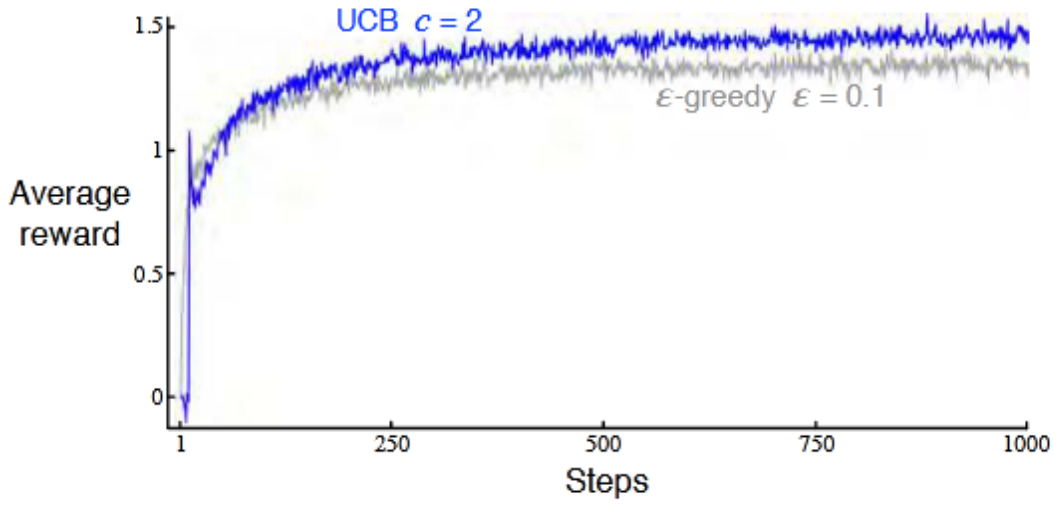
در ابتدا ساده ترین الگوریتم را در نظر می گیریم که به هر اکشن یک میانگین از پاداش های بدست امده را نسبت دهیم و می خواهیم نتایج الگوریتم $\epsilon - Greedy$ را برسی کنیم .مساله 10 armed bandits را که در کلاس مطرح شد را به ازای ϵ های 0.1 و 0.5 و 0.7 و 1 برسی کنید و نمودار میانگین امتیاز به ازای 1000 گام را رسم کنید .

1.با افزایش ϵ نتایج چگونه خواهد شد ؟ چرا ؟

2.الگوریتم $\epsilon - Greedy$ چه مزیتی دارد ؟

3.الگوریتم UCB چگونه می تواند $Exploration$ را برای ما فراهم کند ؟

4. در مساله 10 armed bandits در حالتی که از لگوریتم UCB استفاده کنیم با c=2 نتیجه به شکل زیر خواهد بود.



همانطور که مشاهده می کنید در گام 11 ما یک پیک داریم . توضیح دهید چه عاملی باعث به وجود امدن پیک شده است.

برای انجام تمارین می توانید از کد های اماده لینک زیر استفاده کنید.

https://github.com/ShangtongZhang/reinforcement-learning-an-introduction/blob/master/chapter02/ten_armed_testbed.py

لطفا پاسخ تمرینات را در قالب یک فایل PDF در ویو اپلود کنید .

موفق باشید