



تمرینات برنامه نویسی پایتون سری سوم

نام درس :

یادگیری ماشین

استاد:

دکتر جمیلی

دستیار درس:

هومن پوررستمی

مهلت تحویل تا تاریخ: ۱۴۰۳/۰۳/۱۴

توابع و متدهای مورد نیاز به جهت انجام تمرین:

استفاده از کتابخانه های آماده ممنوع است.

فایل نهایی را به فرمت ipynb در ایلرن بارگذاری کنید.

صورت مسئله:

ما می خواهیم یک عامل ایجاد کنیم که عملی را با بالاترین پاداش مورد انتظار پیدا کند. یکی از راه هایی که یک عامل می تواند عمل کند این است که همیشه اقدامی را با بالاترین ارزش بر اساس برآوردهای فعلی عامل انتخاب کند. این عامل حریص نامیده می شود زیرا حریصانه عملی را انتخاب می کند که فکر می کند بالاترین ارزش را دارد.

۱. تابعی بنویسید که ارزش هر مرحله را به عنوان ورودی دریافت کند و اندیس بیشترین مقدار ارزش را برگرداند. (حداقل ۵ عمل و حداکثر ۱۰ عمل مجاز در هر مرحله امکان پذیر است. همچنین مقادیر تابع ارزش تصادفی باشد)
۲. با توجه به تابع تعریف شده در مرحله قبل درستی عملکرد تابع را بررسی کنید. در این بخش سعی کنید ارزش دو عمل را یکسان در نظر بگیرید و ۱۰۰ بار حلقه را اجرا کنید و ببینید مجموع اندیس های انتخاب شده چقدر خواهد شد. (به کمک assert)
۳. یک کلاس با نام (GreedyAgent) ایجاد کنید. با توجه به عمل های تعریف شده در سوال اول مقادیر تابع ارزش برای هر اهرم جایزه را طبق رابطه زیر برورسانی کنید. در خروجی مقادیر ارزش اهرم ها را بعد از ۵ تکرار نشان دهید. همچنین تعداد عمل انجام شده در هر بازو را نشان دهید.

$$Q_{n+1} = Q_n + \frac{1}{n} (R_n - Q_n)$$

۴. مدل را ۲۰۰ مرتبه اجرا کنید. میانگین ارزش هر اهرم و بهترین اهرم را نشان دهید. مجموع پاداش های موجود را نیز نمایش دهید.
۵. همان مدل قبلی را این بار با سیاست (EpsilonGreedy) پیاده سازی کنید. (مقدار اپسیلون را دلخواه فرض کنید)
۶. نمودار میانگین پاداش را برای دو روش (GreedyAgent, EpsilonGreedy) رسم کنید.
۷. مقادیر اپسیلون را سه مقدار دلخواه در نظر گرفته و نمودار میانگین پاداش را نمایش دهید.