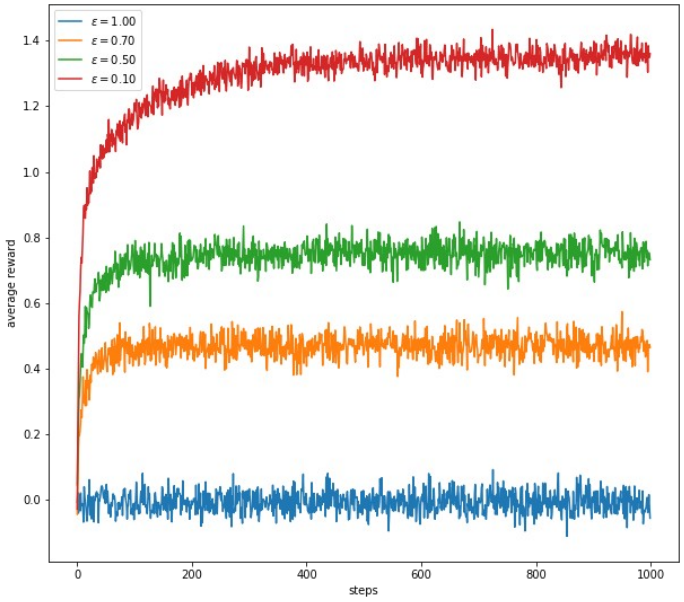
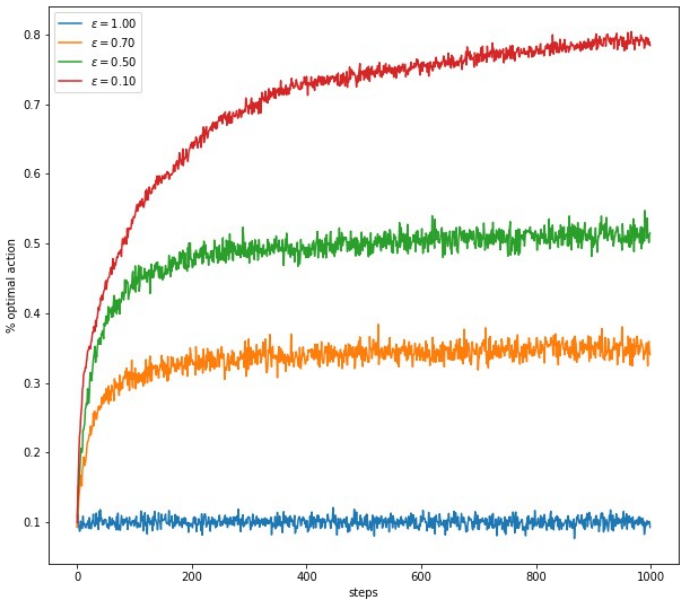
به نام خدا

تمرین اول یادگیری تقویتی سروش ناصری

تمرین اول :





1. مشاهده می شود که با افزایش ابسیلون تعداد اکشن های درست و همچنین امتیاز نهایی بهتر می شود دلیل این موضوع این است که مقدار exploration ما افزایش میابد و این موحب میشود که بتوانیم اکشن درست را بهتر بیابیم .البته اگر این مقدار از یک حدی بیشتر شود باعث می شود که این نتایج کاهش بیاید چون عامل نباید همیشه در حال جست و جو باشد ولی اگر لبسیلورا تا حدی افزایش دهیم موجب بهبود کارایی می شود.

۲.در این الگوریتم ما بخشی از انتخاب های خودرا از میان اکشن هایی که دارای مقدا بیشینه نیستند انتخاب می کنیم تا بتوانیم اثر ان هارا نیز محاسبه کنیم و این باعث exploration می شود و که همین موضوع باعث افزایش کارایی کس شود.

۳.الگوریتم ازرابطه زیر استفاده می کند برای انتخاب اکشن :

$$A_t \doteq \arg\max_a \left[Q_t(a) + c\sqrt{\frac{\ln t}{N_t(a)}} \right]$$

در این رابطه اگر یک اکشن زیاد انتخاب شود با توجه به اینکه تعداد انتخابش در مخرج امده است لذا امتیازش برای انتخاب در مرجله بعد کاهش میابد و این باعث می شود اگر یک اکشن زیاد انتخاب شد از

یک جایی به بعد اکشن های دیگر انتخاب شوند و این موضوع باعث exploration می شود.

.۴
دلیل به وجود امدن این قله در گام ۱۱ این است که تا گام ۱۱ هر اکشن به صورت میانگین یکبار انتخاب می شود و بعد از این ۱۰ گام همه گام اکشن ها انتخاب شده است لذا در گام ۱۱ همه اکشن ها احتمال یکسانی را دارند اما اکشن با بیش ترین مقدار سود انتخاب مس شود و باعث به وجود امدن قله می شود اما در گام ۱۲ چون این گام با بیش ترین سود ۲ با انتخاب شده است طبق فرمول انتخاب نمی شود و یکی دیگر از اکشن ها انتخاب می شود که به صورت میانگین سود کمی دارد و لذا از قله فرود می اییم و این دلیل وجود این قله می باشد.