باسمه تعالی

یادگیری برخط

پاییز ۱۴۰۳

تمرین سری دوم : بندیت تصادفی با متناهی دسته

---

(۱) نشان دهید پشیمانی تصادفی، $\bar{R}_n = \sum_{t=1}^{n} \Delta_{A_t}$، و شبه‌پشیمانی، $\hat{R}_n = n\mu^* - \sum_{t=1}^{n} X_t$، هر دو برآوردگرهای نااریبی برای پشیمانی هستند. کدام یک از این دو برآوردگر دقت بیش‌تری دارد؟ چرا؟

(۲) نشان دهید در تحلیل پشیمانی الگوریتم ETC، می‌توان از کران وابسته به مساله که به شکل زیر به دست آمد:

$$R_n \leq \min\left\{n\Delta,\, \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0,\, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)\right\}$$

به کران $R_n \leq \Delta + C\sqrt{n}$ رسید که در آن $C > 0$ یک ثابت مستقل از مساله است.

(۳) سعی کنید الگوریتم ETC را طوری تغییر دهید که برای اجرای آن نیازی به دانستن شکاف زیربهینه، Δ، نباشد. برای سادگی فرض کنید تنها دو دسته داریم و توزیع پاداش هر دو دسته یک‌ـ زیرگوسی است. در الگوریتم جدید باید میزان اکتشاف از قبل مشخص نبوده و وابسته به نتایج به دست آمده باشد. با یک منطق بازه‌ی اطمینانی چنین شرطی بسازید و کران زیر را ثابت کنید:

$$R_n \leq \Delta + \frac{C \log n}{\Delta}$$

(۴) در الگوریتم εـ حریصانه اول یک‌بار هر کدام از دسته‌ها را می‌کشیم و بعد از آن در دور t-اُم با احتمال ε دسته‌ی $A_t = \arg\max_i \hat{\mu}(t-1)$ و با احتمال $1 - \varepsilon$ یک دسته‌ی تصادفی با توزیع احتمال یکنواخت انتخاب می‌کنیم. نشان دهید در یک مساله‌ی بندیت kـ دسته‌ای با پاداش‌های یک‌ـ زیرگوسی داریم:

$$\lim_{n\to\infty} \frac{R_n}{n} = \frac{\varepsilon}{k} \sum_{i=1}^{k} \Delta_i$$

(۵) برای الگوریتم $\mathrm{UCB}(\delta)$، با فرض‌های مناسب در صورت لزوم، کرانی برای شبه‌پشیمانی به شکل زیر اثبات کنید:

$$\mathbb{P}\left(\bar{R}_n \geq g(n,\delta)\right) \leq f(n,k)\delta.$$