598 RL Fall 2020

Day 2

Bandit problem

k-armed bandit

choose an action A ← takes values in $\{1, \dots, k\}$

get a reward R ← takes values in $\mathbb{R}$

want to maximize $E[R]$ ← $\int_{r \in \mathbb{R}} r\, p(r)\, dr$

– or –

$\sum_r r\, p(r)$

in other words, we want:

$$\arg\max_{a \in \{1, \dots, k\}} \underbrace{E[R \mid A = a]}_{Q^*(a)}$$

$$Q(a) = \frac{r_1 + r_2 + \dots + r_n}{n}$$

① choose "a" uniformly at random from $\{1, \dots, k\}$

② choose "a" as $\arg\max_{a \in \{1, \dots, k\}} Q(a)$

EXPLORE { ① choose "a" uniformly at random from $\{1, \dots, k\}$

EXPLOIT { ② choose "a" as $\underset{a \in \{1,\dots,k\}}{\arg\max}\, Q(a)$ } this is "greedy"

"ϵ - greedy" means do the non-greedy thing with probability ϵ

Example:

if $\epsilon = 0.1$, then

90% of time we exploit

10% of time we explore

$$Q_n(a) = \frac{r_1 + \dots + r_n}{n}$$

$$= \frac{r_n}{n} + \frac{r_1 + \dots + r_{n-1}}{n}$$

$$= \frac{r_n}{n} + \frac{(n-1)}{(n-1)} \frac{r_1 + \dots + r_{n-1}}{n}$$

$$= \frac{r_n}{n} + \left(\frac{n-1}{n}\right) \underbrace{\left(\frac{r_1 + \dots + r_{n-1}}{n-1}\right)}_{Q_{n-1}(a)}$$

$$= \frac{r_n}{n} + Q_{n-1}(a) - \left(\frac{1}{n}\right) Q_{n-1}(a)$$

$$= \underline{Q_{n-1}(a)} + \underbrace{\frac{1}{n}}_{\text{step size}} \Big(\underset{\substack{\uparrow \\ \text{TARGET}}}{r_n} - Q_{n-1}(a) \Big)$$