Numericals

Multi-arm Bandits

(1) ETC

- Explore each arm 'N' times.
- Select arm with best sample avg. and play that forever after the Exploration rounds.

Example:

Let no. of arms $k=3$. Arms $= \{a, b, c\}$.

Assume $N=3$ (exploration rounds per arm)

we observed $(a, 1), (a, 2), (a, 3)$

$(b, 2), (b, 4), (b, -10)$

$(c, 10), (c, 11), (c, -5)$

be the first 9 rounds data.

Which arm will be played on 10^{th} round?

Ans:-

$$\bar{\mu}(a) = \frac{1+2+3}{3} = 2$$

$$\bar{\mu}(b) = \frac{2+4-10}{3} = \frac{-4}{3}$$

$$\bar{\mu}(c) = \frac{10+11-5}{3} = \frac{16}{3}.$$

Based on these observations, the best arm is "c".

⇒ From round 10, arm 'c' will be played.

---

(2) UCB

At round t, Pick arm

$$\underset{a}{\arg\max} \left\{ \bar{\mu}_{t-1}(a) + \sqrt{\frac{2 \log T}{n_{t-1}(a)}} \right\}$$

Example:

K → no. of arms.

Assume $T = 100$. $K = 2$.

Data observed till round 3.

$(a, +2)$, $(b, +3)$, $(b, +2)$

which arm will be played at round 4?

UCB scores:

Note: base 'e' is used for log.

At $t=4 \Rightarrow t-1=3$.

Arm a:

$$\bar{\mu}_3(a) + \sqrt{\frac{2\log T}{n_3(a)}}$$

$$= \frac{2}{1} + \sqrt{\frac{2\log 100}{1}}$$

$$= 2 + \sqrt{9.21}$$

$$= 2 + 3.0348$$

$$= 5.0348$$

Arm b:

$$\bar{\mu}_3(b) + \sqrt{\frac{2\log T}{n_t(b)}}$$

$$= \frac{3+2}{2} + \sqrt{\frac{2\log 100}{2}}$$

$$= 2.5 + \sqrt{4.605}$$

$$= 2.5 + 2.1459 = 4.6459 //$$

$\Rightarrow$ Arm a played at $t=4$ due to high UCB score.

(3) ε-greedy

- w.p. $1-\varepsilon$ Exploit: play arm with highest sample avg.
- w.p. ε Explore: play any arm randomly.

Example:

Let $K=2$. First three rounds be $(a,2)$, $(b,3)$, $(b,2)$.

Assume $\varepsilon = 0.2$. what is probability of playing arm 'b'?

At $t=4$

$$\overline{\mu}_{t-1}(a) = 2$$

$$\overline{\mu}_{t-1}(b) = \frac{3+2}{2} = 2.5$$

$\Rightarrow$ Best arm according to sample avg: is arm "b".

Prob. of Exploit $= 1 - \varepsilon$

$= 1 - 0.2$

$= 0.8$.

Prob. of Explore $= \varepsilon = 0.2$

Note:

Explore
- During Exploration, any arm is played uniformly at random.
- So, during Exploration both arm a, b have equal (half) chance of playing. Hence

Exploit
- arm 'b' is played during Exploitation since $\bar{\mu}(b) > \bar{\mu}(a)$.

Probability of arm b.

$$= P(\text{Exploit}) + \frac{1}{2} P(\text{Explore})$$

$$= 0.8 + \frac{1}{2}(0.2)$$

$$= 0.9$$

My prob. of arm a

- Arm 'a' has half chance during Exploration.
- Zero chance during Exploitation.

$$P(\text{arm a}) = 0 \times P(\text{Exploit}) + \frac{1}{2} \times P(\text{Explore})$$

$$= 0.1$$