CS 7545: Machine Learning Theory **Fall 2018**

Lecture 1: UCB algorithm

Lecturer: Jacob Abernethy *Scribes: Rui Zhang, Xinshi Chen*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

16.1 UCB Algorithm

Problem Setting

There are K arms.

Arm i has distribution $D_i \in \Delta([0,1])$ with mean μ_i.

At time t, payoff $X_i^t \sim D_i$.

For $t = 1, \cdots, T$:

Algorithm pulls arm $i_t \in [K]$.

Algorithm receives/observes $X_{i_t}^t$.

Definition 16.1 (expected regret) *The* ***expected regret*** *at time* T *is*

$$\mathbb{E}[\text{Regret}_T] := \mathbb{E}_{\text{algo}}\Big[\sum_{t=1}^{T}(\mu_{i^*} - \mu_{i^t})\Big], \quad \textit{where } i^* = \arg\max_{i\in[K]} \mu_i.$$

Definition 16.2 (performance gap) *The* ***performance gap*** *is for* $i = 1, \cdots, K$

$$\Delta_i := \mu_{i^*} - \mu_i.$$

Algorithm 1 UCB

```
for t = 1 to K do
    Pull i_t = t
end for
for t > K do
    N_i^t = sum_{s=1}^{t-1} 1[i_s = i]
    hat{mu}_i^t = (1/N_i^t) sum_{s=1}^{t-1} X_i^s 1[i_s = i]
    i_t = arg max_{i in [K]} [ hat{mu}_i^t + sqrt( log(2/delta) / (2 N_i^t) ) ]
end for
```

Theorem 16.3 *Regret Bound:*

$$\mathbb{E}[\text{Regret}_T(UCB)] = O(\sum_{i\neq i^*} \frac{\log T}{\Delta_i})$$

Proof: By Hoeffding's inequality:

$$\Pr\Big(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| > t\Big) \le \Pr\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu > t\Big) + \Pr\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu < -t)\Big) \le 2\exp(-2nt^2) = \delta.$$

Take $2nt^2 = \log\frac{2}{\delta}$. Then $t = \sqrt{\frac{\log(2/\delta)}{2n}}$. Thus,

$$\Pr\Big(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| > \sqrt{\frac{\log(2/\delta)}{2n}}\Big) \le \delta.$$

WLOG, assume $i^* = 1$. Consider two events at time t.

(A1) $\mu_1 \le \widehat{\mu}_1^t + \sqrt{\dfrac{\log 2/\delta}{2N_1^t}}$.

(A2) $\widehat{\mu}_{i_t}^t \le \mu_{i_t} + \sqrt{\dfrac{\log 2/\delta}{2N_{i_t}^t}}$.

Let $\xi_t = \mathbf{1}[(\text{A1}) \text{ or } (\text{A2}) \text{ fails}]$. Then

$$\Pr(\xi_t = 1) \le \Pr((\text{A1) fails}) + \Pr((\text{A2) fails}) \le 2\delta.$$

If both (A1) and (A2) hold, then

$$\mu_1 \overset{(A1)}{\le} \widehat{\mu}_1^t + \sqrt{\frac{\log(2/\delta)}{2N_1^t}} \overset{alg.}{\le} \widehat{\mu}_{i_t}^t + \sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}} \overset{(A20}{\le} \mu_{i_t} + 2\sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}},$$

and consequently

$$\mu_1 - \mu_{i_t} \le 2\sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}}.(i.e.\Delta_{i_t} \le 2\sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}})$$

Claim: on round t, the regret is bounded by

$$\xi_t(\text{cost paid if A1 or A2 fail}) + 2\sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}}(\text{cost paid if A1 and A2 hold}).$$

Define

$$\Phi(\vec{N}) := \Phi(N_1, \cdots, N_K) = 2\sum_{k=2}^{K}\sum_{n=1}^{N_k}\sqrt{\frac{\log(2/\delta)}{2n}}.$$

$$\begin{aligned}
\mathbb{E}[\text{Reg}_T(\text{UCB})] &:= \mathbb{E}[\sum_{t=1}^{T}\mu_1 - \mu_{i_t}] \le \mathbb{E}[\sum_{t=1}^{T}\Big(\xi_t + 2\sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}}\Big)] \\
&= \mathbb{E}[\sum_{t=1}^{T}\xi_t] + \mathbb{E}[\sum_{t=1}^{T}\Phi(\vec{N}^{t+1}) - \Phi(\vec{N}^t)] \\
&\le 2T\delta + \mathbb{E}[\Phi(\vec{N}^{t+1}) - \Phi(\vec{0})]
\end{aligned}$$

Claim: We only need to consider $N_i^t \leq \frac{2\log(2/\delta)}{\Delta_i^2}$. [1] Only need to look at $\vec{N}^t \leq [\frac{2\log(2/\delta)}{\Delta_i^2}]_{i=1,\cdots,k}$. Denote $N^* = [\frac{\log(2/\delta)}{2\Delta_i^2}]_{i=1,\cdots,k}$. Then

$$\Phi(N^*) = 2\sum_{i=2}^{K}\sum_{n=1}^{\frac{2\log(2/\delta)}{\Delta_i^2}}\sqrt{\frac{\log(2/\delta)}{2n}} \leq \sqrt{\frac{\log(2/\delta)}{2}}\sum_{i=2}^{K}4\sqrt{\frac{2\log(2/\delta)}{\Delta_i^2}} = 4\log(2/\delta)\sum_{i=2}^{K}\frac{1}{\Delta_i}$$

where the first inequality uses $\sum_{n=1}^{x}\sqrt{\frac{1}{n}} \leq \int_1^x\sqrt{\frac{1}{n}} \leq 2\sqrt{x}$. Let $\delta = \frac{1}{2T}$, then

$$\mathbb{E}[\mathrm{Reg}_T(\mathrm{UCB})] \leq 1 + 4\log(4T)\sum_{i=2}^{n}\frac{1}{\Delta_i}$$

■

[1]Otherwise, we have $\hat{\mu}_1^t + \sqrt{\frac{\log(2/\delta)}{2N_i^t}} \overset{(A1)}{>} \mu_1 = \mu_{i_t} + \Delta_{i_t} \overset{N_{i_t}^t > \frac{\log(2/\delta)}{\Delta_{i_t}^2}}{>} \mu_{i_t} + 2\sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}} \overset{(A2)}{\geq} \hat{\mu}_{i_t} + \sqrt{\frac{\log(2/\delta)}{2N_{i_t}^t}}$, which leads to contradiction, as i_t is selected by the algorithm instead of i_1.