Week 9 – Lecture 44

$$\xrightarrow[t=1,2,\ldots]{x_t} \quad \boxed{1}\ \boxed{2}\ \boxed{3}\ \cdots\ \boxed{K}$$

At every round context x_t is revealed and learner has to play an arm.

Reward:

$$r_t = R(x_t, I_t) + \eta_t$$

Regret:

$$R_T(\pi) = \mathbb{E}\left[\sum_{t=1}^{T} \max_a R(X_t, a) - \sum_{t=1}^{T} R(X_t, I_t)\right]$$

$$R : \underset{\substack{\uparrow \\ \text{context set}}}{C} \times [K] \longrightarrow \mathbb{R}$$

$$R(x,a) = \langle \underbrace{\Psi(x,a)}_{\substack{\text{feature map} \\ \text{(known)}}}, \theta^* \rangle$$

θ^* ← unknown, independent of context, arm

Stochastic Contextual MAB (SCB)

Stochastic Linear Bandits (SLB)

In every round a decision set is revealed

$\{D_t\}_{t \geq 1}$ $\qquad D_t \subset \mathbb{R}^d$ (bounded subset)

[here D_t may be uncountably many]

Reward:

observation noise

$$r_t = \langle d_t, \theta^* \rangle + \eta_t$$

d_t played in round t

Maximum expected reward:

$$\sum_{t=1}^{T} \max_{d \in D_t} \langle d, \theta^* \rangle$$

Regret:

$$R_T(\pi) = \mathbb{E}\left[\sum_{t=1}^{T} \max_{d \in D_t} \langle d, \theta^* \rangle - \sum_{t=1}^{T} \langle d_t, \theta^* \rangle \right]$$

Linear setting as trying to optimize a stochastic linear problem through many rounds.

(D_t is finite)

If: $D_t = \{ \Psi(x_t, a) : a \in [K] \}$ $\quad |D_t| < \infty \ \forall t$

then Stochastic Contextual bandits is equivalent to Stochastic Linear Bandits, special case of Stochastic Linear Bandit.

<u>Special Case</u>:

1) If $D_t = \{ \psi(x_t, a) : a \in [K] \}$

$|D_t| < \infty \ \forall t$

then SLB is same as SCB.

2) $D_t = \{ e_1, e_2 \dots e_d \}$

SLB is same as Stochastic d-armed bandit

<u>Solving SLB</u>

$$R_T(\pi) = \sum_{t=1}^{T} \max_{d \in D_t} \langle d, \theta^* \rangle - \sum_{t=1}^{T} \langle d_t, \theta^* \rangle$$