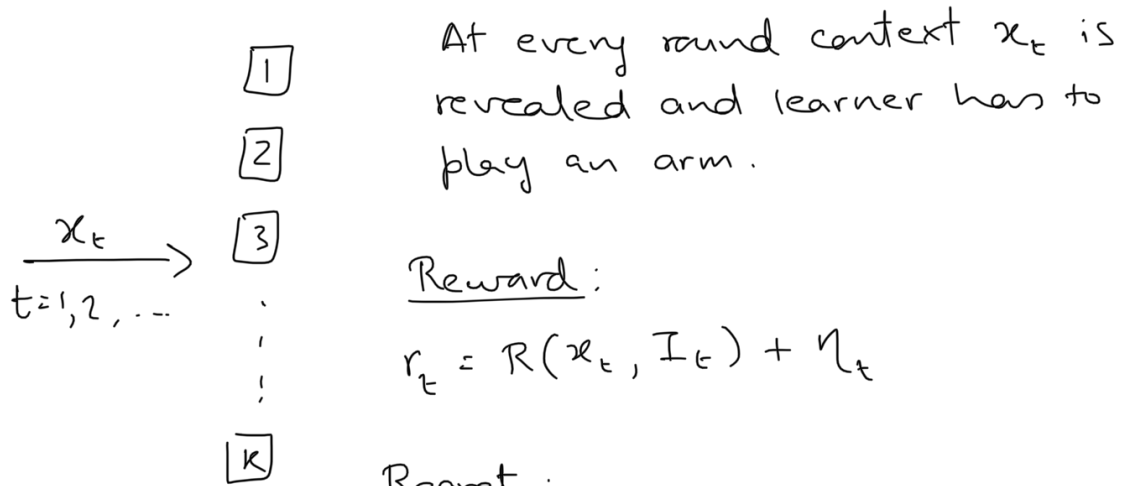


## Week 9 - Lecture 44



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

$$R: C \times [K] \rightarrow \mathbb{R}$$

$\uparrow$   
context set

$$R(x, a) = \langle \underbrace{\Psi(x, a)}_{\substack{\downarrow \\ \text{feature map} \\ \text{(known)}}}, \underbrace{\theta^*}_{\substack{\uparrow \\ \text{unknown} \\ \text{independent of} \\ \text{context, arm}}} \rangle$$

Stochastic Contextual MAB  $\uparrow$  (SCB)

## Stochastic Linear Bandits (SLB)

In every round a decision set is revealed

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

[here  $D_t$  may be uncountably many]

Reward:

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

$\uparrow$   
 $d_t$  played in round  $t$

$\nwarrow$  observation noise

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)

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

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

Special Case:

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

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

then SLB is same as SCB.

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

SLB is same as Stochastic d-armed bandit

---

Solving SLB

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