Contextual Bandits (Numericals)

- Let there be 2-arms a, b.
- 2 features for each user. (x_1, x_2).
- Model expected reward by linear fn.

$$\bar{\mu}_a(x) = \theta_1^a x_1 + \theta_2^a x_2$$

$$\bar{\mu}_b(x) = \theta_1^b x_1 + \theta_2^b x_2$$

ETC :

Assume $N=2$ Exploration rounds per each arm. If the first 4 rounds of data, is the format

(features, arm, Reward)

are as follows:

$$\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, a, 1 \right)$$

$$\left(\begin{bmatrix} 3 \\ 5 \end{bmatrix}, a, 4 \right)$$

$$\left(\begin{bmatrix} 4 \\ 3 \end{bmatrix}, b, -1 \right)$$

$$\left(\begin{bmatrix} 10 \\ 2 \end{bmatrix}, b, 12 \right)$$

which arm will be played in round 5 to a user with feature vector $\begin{bmatrix} 3 \\ 7 \end{bmatrix}$?

Solution :-

Estimate $\hat{\theta}$ as $\hat{\theta}_a = (D_a^T D_a + I)^{-1} D_a^T b_a$

Rewards during Exploration

features of users in Exploration

Identity matrix

For Arm a

$$D_a = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$$

$\rightarrow$ 1^{st} row = 1^{st} round features

$\rightarrow$ 2^{nd} row = 2^{nd} round features.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$b_a = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$\rightarrow$ reward in first round

$\rightarrow$ reward in second round.

$$\hat{\theta}_a = (D_a^T D_a + I)^{-1} D_a^T b_a$$

$$= \left(\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} 10 & 17 \\ 17 & 29 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 13 \\ 22 \end{bmatrix}$$

$$= \frac{1}{41} \begin{bmatrix} 30 & -17 \\ -17 & 11 \end{bmatrix} \begin{bmatrix} 13 \\ 22 \end{bmatrix}$$

$$\hat{\theta}_a = \begin{bmatrix} 0.3902 \\ 0.5121 \end{bmatrix}.$$

<u>For Arm b</u> :

$$D_b = \begin{bmatrix} 4 & 3 \\ 10 & 2 \end{bmatrix} \qquad b_b = \begin{bmatrix} -1 \\ 12 \end{bmatrix}$$

$$\hat{\theta}_b = \left(D_b^T D_b + I \right)^{-1} D_b^T b_b$$

$$= \left(\begin{bmatrix} 116 & 32 \\ 32 & 13 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 4 & 10 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 12 \end{bmatrix}$$

$$= \frac{1}{815} \begin{bmatrix} -5 & 30 \\ 30 & -17 \end{bmatrix} \begin{bmatrix} 116 \\ 21 \end{bmatrix}$$

$$= \frac{1}{815} \begin{bmatrix} 50 \\ 3123 \end{bmatrix}$$

$$\hat{\theta}_b = \begin{bmatrix} 0.0613 \\ 3.8319 \end{bmatrix}.$$

New user

As per the question, the feature vector of the new user is $\begin{bmatrix} 3 \\ 7 \end{bmatrix} \begin{matrix} \rightarrow x_1 \\ \rightarrow x_2 \end{matrix}$

Based on our $\hat{\theta}_a$, $\hat{\theta}_b$ let us find

$$\hat{\mu}_a(x) = \hat{\theta}_a^1 x_1 + \hat{\theta}_a^2 x_2$$

$$= (0.3902)3 + (0.5121)7$$

$$= 4.7553$$

$$\hat{\mu}_b(x) = \hat{\theta}_b^1 x_1 + \hat{\theta}_b^2 x_2$$

$$= (0.0613)3 + (3.8319)7$$

$$= 27.002$$

Arm b will be played $\because \hat{\mu}_b(x)$ is larger.

LinUCB

Pick $\underset{a}{\operatorname{argmax}} \; x^T \theta^a + \sqrt{x^T (D_a^T D_a + I)^{-1} x}$

Que: Assume the same data as before and calculate UCB scores of arm a, b after 4 rounds.

Sol: —

→ $\hat{\theta}_a$, $\hat{\theta}_b$ should be computed in the same way as done in previous problem.

→ So, $x^T \hat{\theta}^a$, $x^T \hat{\theta}_b$ will also be the same.

→ we have to just calculate "Exploration" $\sqrt{x^T (D_a^T D_a + I)^{-1} x}$ term.

Arm a:

$$\sqrt{\begin{bmatrix} 3 & 7 \end{bmatrix} \frac{1}{14} \begin{bmatrix} 30 & -17 \\ -17 & 11 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix}}$$

$$= \sqrt{6.7857} = \underline{\underline{2.604}}$$

Arm b

$$\sqrt{\begin{bmatrix}3 & 7\end{bmatrix} \frac{1}{815}\begin{bmatrix}-5 & 30\\ 30 & -17\end{bmatrix}\begin{bmatrix}3\\ 7\end{bmatrix}}$$

$$= \sqrt{0.4687}$$

$$= 0.6846$$

UCB scores :

$$UCB_a = \text{Exploit score} + \text{Explore score}$$

$$= \hat{\mu}_a(x) + \sqrt{x^T (D_a^T D_a + I)^{-1} x}$$

$$= 4.7533 + 2.604$$

$$= 7.357$$

Arm b played since UCB_b is larger.

$$UCB_b = \hat{\mu}_b(x) + \sqrt{x^T (D_a^T D_a + I)^{-1} x}$$

$$= 27.002 + 0.684$$

$$= 27.686$$

End