EEE932 Assignment-1 Solution

---

eMasters in Communication Systems, IITK
EE932: Introduction to Reinforcement Learning
Instructor: Prof. Subrahmanya Swamy Peruru
Student Name: Venkateswar Reddy Melachervu
Roll No: 23156022

Question 11: Consider a contextual bandits scenario in which the true mean $\mu(\bar{x}) = \theta_a^T \bar{x}$ of an arm a is a linear function of the context vector $\bar{x}$. Here θ_a and x are $n \times 1$ vectors if n is the number of features in the context vector. Assume that we have two arms a_1 and a_2 and samples *(context, action, rewards)* observed by the agent in the first 6 rounds as follws:

$\left(\begin{bmatrix}1\\3\end{bmatrix}, a_1, r = 17\right), \left(\begin{bmatrix}7\\13\end{bmatrix}, a_2, r = 2\right), \left(\begin{bmatrix}5\\7\end{bmatrix}, a_1, r = 2\right), \left(\begin{bmatrix}5\\3\end{bmatrix}, a_2, r = 1\right), \left(\begin{bmatrix}11\\13\end{bmatrix}, a_1, r = 23\right), \left(\begin{bmatrix}5\\7\end{bmatrix}, a_2, r = 9\right)$

If the context seen in 7th round is $\begin{bmatrix}2\\1\end{bmatrix}$

If LinUCB algorithm is used, what are the UCB scores of arm a_1 and arm a_2 for the above problem w.r.t to the context seen in the 7th round? Upload an attachment showing your solution.

Solution:
For UCB:

- For round 7 pick an arm with $\arg\max_a(\bar{x}^T\theta^a + \sqrt{\bar{x}^T(D_a^T D_a + I)^{-1}\bar{x}})$
- Select the arm that has higher UCB for round 7 with the context seen for it

$$\hat{\theta}_{a_1} = \left(D_{a_1}^T D_{a_1} + I\right)^{-1} D_{a_1}^T b_{a_1}$$

$$D_{a_1} = \begin{bmatrix}1 & 3\\5 & 7\end{bmatrix} \Rightarrow D_{a_1}^T = \begin{bmatrix}1 & 5\\3 & 7\end{bmatrix}, I = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

$$D_{a_1}^T D_{a_1} = \begin{bmatrix}26 & 38\\38 & 58\end{bmatrix} \Rightarrow \left(D_{a_1}^T D_{a_1} + I\right) = \begin{bmatrix}27 & 38\\38 & 59\end{bmatrix}$$

$$\left(D_{a_1}^T D_{a_1} + I\right)^{-1} = \frac{1}{149}\begin{bmatrix}59 & -38\\-38 & 27\end{bmatrix}$$

$$D_{a_1}^T b_{a_1} = \begin{bmatrix}1 & 5\\3 & 7\end{bmatrix}\begin{bmatrix}17\\2\end{bmatrix} = \begin{bmatrix}27\\65\end{bmatrix}$$

$$\boldsymbol{\hat{\theta}_{a_1}} = \frac{1}{149}\begin{bmatrix}59 & -38\\-38 & 27\end{bmatrix}\begin{bmatrix}27\\65\end{bmatrix} = \begin{bmatrix}0.396 & -0.255\\-0.255 & 0.181\end{bmatrix}\begin{bmatrix}27\\65\end{bmatrix} = \begin{bmatrix}\mathbf{-5.8859}\\\mathbf{4.8926}\end{bmatrix}$$

$$\hat{\theta}_{a_2} = \left(D_{a_2}^T D_{a_2} + I\right)^{-1} D_{a_2}^T b_{a_2}$$

$$D_{a_2} = \begin{bmatrix}7 & 13\\5 & 3\end{bmatrix} \Rightarrow D_{a_1}^T = \begin{bmatrix}7 & 5\\13 & 3\end{bmatrix}, I = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

$$D_{a_2}^T D_{a_2} = \begin{bmatrix}74 & 106\\106 & 178\end{bmatrix} \Rightarrow \left(D_{a_2}^T D_{a_2} + I\right) = \begin{bmatrix}75 & 106\\106 & 179\end{bmatrix}$$

$$\left(D_{a_2}^T D_{a_2} + I\right)^{-1} = \frac{1}{2189}\begin{bmatrix}179 & -106\\-106 & 75\end{bmatrix}$$

$$D_{a_2}^T b_{a_2} = \begin{bmatrix}7 & 5\\13 & 3\end{bmatrix}\begin{bmatrix}2\\1\end{bmatrix} = \begin{bmatrix}19\\29\end{bmatrix}$$

$$\boldsymbol{\hat{\theta}_{a_2}} = \frac{1}{2189}\begin{bmatrix}179 & -106\\-106 & 75\end{bmatrix}\begin{bmatrix}19\\29\end{bmatrix} = \begin{bmatrix}0.082 & -0.048\\-0.048 & 0.034\end{bmatrix}\begin{bmatrix}19\\29\end{bmatrix} = \begin{bmatrix}\mathbf{0.1494}\\\mathbf{0.0735}\end{bmatrix}$$

Let's compute $\hat{\theta}_{a_1}^T\bar{x}^7, \hat{\theta}_{a_2}^T\bar{x}^7$

$$\boldsymbol{\mu(a_1) = \hat{\theta}_{a_1}^T\bar{x}^7 = [-5.883 \quad 4.880]\begin{bmatrix}2\\1\end{bmatrix} = -6.8792}$$

$$\boldsymbol{\mu(a_2) = \hat{\theta}_{a_2}^T\bar{x}^7 = [0.166 \quad 0.074]\begin{bmatrix}2\\1\end{bmatrix} = 0.3723}$$

C-501, Salarpuria Serenity, 5th Main, Sector 7, HSR Layout, Bengaluru 560102 KA India
Mobile: +91 97012 22130, Email: vmelachervu@gmail.com, vmela23@iitk.ac.in, Website: www.linkedin.com/in/vmelachervu

Venkateswar Reddy M.
CTO, Brillium Technologies
"...dare to dream; care to win..."

$$e_{a_1} = \sqrt{\bar{x}_7{}^T\left(D_{a_1}^T D_{a_1} + I\right)^{-1}\bar{x}_7} = 0.8631$$

$$e_{a_2} = \sqrt{\bar{x}_7{}^T\left(D_{a_2}^T D_{a_2} + I\right)^{-1}\bar{x}_7} = 0.4095$$

$$\boldsymbol{linUCB_{a_1} = \bar{x}_7{}^T\theta^{a_1} + \sqrt{\bar{x}_7{}^T\left(D_{a_1}^T D_{a_1} + I\right)^{-1}\bar{x}_7} = -6.8792 + 0.8631 = -6.0161}$$

$$\boldsymbol{linUCB_{a_2} = \bar{x}_7{}^T\theta^{a_2} + \sqrt{\bar{x}_7{}^T\left(D_{a_2}^T D_{a_2} + I\right)^{-1}\bar{x}_7} = 0.3723 + 0.4095 = 0.7818}$$

$$\boldsymbol{linUCB_{a_2} > linUCB_{a_1}}$$

∴ In 7th round arm a_2 will be played as per LinUCB

--- End of the Document ---

C-501, Salarpuria Serenity, 5th Main, Sector 7, HSR Layout, Bengaluru 560102 KA India
Mobile: +91 97012 22130, Email: vmelachervu@gmail.com, vmela23@iitk.ac.in, Website: www.linkedin.com/in/vmelachervu