LinUCB Formulation Based on the Sherman-Morrison Matrix Inverse Update

Chris Diehl

May 7, 2015

In the original formulation of the LinUCB algorithm [1] for the contextual bandits problem, a matrix inverse is required following a rank one update during each iteration. In this document, we present an alternative formulation of the algorithm that leverages the Sherman-Morrison matrix inverse update to directly compute the matrix inverse of the perturbed matrix, thereby reducing the computational cost of each iteration. We begin by introducing the original formulation of the LinUCB algorithm as presented in [1].

```
Input: α > 0
for t = 1, 2, 3, ..., T do
    Observe features for all arms a ∈ 𝒜_t : x_{t,a} ∈ ℝ^d
    for a ∈ 𝒜_t do
        if a is new then
            A_a ← I_d
            b_a ← 0_{dx1}
        end
        θ̂_a ← A_a^{-1} b_a
        p_{t,a} ← θ̂_a^T x_{t,a} + α sqrt(x_{t,a}^T A_a^{-1} x_{t,a})
    end
    Choose arm a_t = arg max_{a∈𝒜_t} p_{t,a} with ties broken arbitrarily and
    observe a real-valued payoff r_t
    A_{a_t} ← A_{a_t} + x_{t,a} x_{t,a}^T
    b_{a_t} ← b_{a_t} + r_t x_{t,a}
end
```

Algorithm 1: Original formulation of the LinUCB algorithm

Without loss of generality, we will assume in the following that the feature vector x_t is invariant to the choice of arm. Our objective is to derive an update equation for $\hat{\boldsymbol{\theta}}_{a_t}$ through direct substitution of the updates for $\mathbf{A}_{a_t}$ and $\mathbf{b}_{a_t}$. Direct substitution implies

$$\hat{\boldsymbol{\theta}}_{a_t} \leftarrow \left(\mathbf{A}_{a_t} + \mathrm{x}_t \mathrm{x}_t^\intercal\right)^{-1} \left(\mathbf{b}_{a_t} + r_t \mathrm{x}_t\right). \tag{1}$$

The Sherman-Morrison matrix inverse update equation indicates that

$$(\mathbf{A}_{a_t} + \mathrm{x}_t \mathrm{x}_t^\intercal)^{-1} = \mathbf{A}_{a_t}^{-1} - \frac{\mathbf{A}_{a_t}^{-1} \mathrm{x}_t \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1}}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t}. \tag{2}$$

Applying this result in the context of equation (1) yields

$$\begin{aligned}
(\mathbf{A}_{a_t} + \mathrm{x}_t \mathrm{x}_t^\intercal)^{-1} \mathbf{b}_{a_t} &= \mathbf{A}_{a_t}^{-1} \mathbf{b}_{a_t} - \frac{\mathbf{A}_{a_t}^{-1} \mathrm{x}_t \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathbf{b}_{a_t}}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} && (3) \\
&= \hat{\boldsymbol{\theta}}_{a_t} - \frac{\mathbf{A}_{a_t}^{-1} \mathrm{x}_t \mathrm{x}_t^\intercal \hat{\boldsymbol{\theta}}_{a_t}}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} && (4) \\
&= \hat{\boldsymbol{\theta}}_{a_t} - \frac{\hat{r}_t \mathbf{A}_{a_t}^{-1} \mathrm{x}_t}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} && (5) \\
(\mathbf{A}_{a_t} + \mathrm{x}_t \mathrm{x}_t^\intercal)^{-1} r_t \mathrm{x}_t &= r_t \mathbf{A}_{a_t}^{-1} \mathrm{x}_t - \frac{r_t \mathbf{A}_{a_t}^{-1} \mathrm{x}_t \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} x_t}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} && (6) \\
&= \frac{r_t \mathbf{A}_{a_t}^{-1} \mathrm{x}_t \left(1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} x_t\right)}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} - \frac{r_t \mathbf{A}_{a_t}^{-1} \mathrm{x}_t \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} x_t}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} && (7) \\
&= \frac{r_t \mathbf{A}_{a_t}^{-1} \mathrm{x}_t}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} && (8)
\end{aligned}$$

Integrating these results yields the parameter update rule

$$\hat{\boldsymbol{\theta}}_{a_t} \leftarrow \hat{\boldsymbol{\theta}}_{a_t} + \frac{(r_t - \hat{r}_t) \mathbf{A}_{a_t}^{-1} \mathrm{x}_t}{1 + \mathrm{x}_t^\intercal \mathbf{A}_{a_t}^{-1} \mathrm{x}_t} \tag{9}$$

where $\hat{r}_t = \mathrm{x}_t^\intercal \hat{\boldsymbol{\theta}}_{a_t}$. Reorganizing the computations in algorithm 1 to leverage these update equations, we obtain algorithm 2 which removes the need for a matrix inverse per iteration.

References

[1] Lihong Li, Wei Chu, John Langford, Robert Shapire, A Contextual-Bandit Approach to Personalized News Article Recommendation, In *Proceedings of the International World Wide Web Conference*, pp. 661-670, April 2010.

Input: $\alpha > 0$
for $t = 1, 2, 3, \ldots, T$ **do**
    Observe features for all arms $a \in \mathcal{A}_t : \mathrm{x}_{t,a} \in \mathbb{R}^d$
    for $a \in \mathcal{A}_t$ **do**
        if a is new **then**
            $\mathbf{A}_{t,a}^{-1} \leftarrow \mathbf{I}_d$
            $\hat{\boldsymbol{\theta}}_{t,a} \leftarrow \mathbf{0}_{dx1}$
        end
        $\hat{r}_{t,a} = \hat{\boldsymbol{\theta}}_{t,a}^{\intercal} \mathrm{x}_{t,a}$
        $\mathrm{w}_{t,a} = \mathbf{A}_{t,a}^{-1} \mathrm{x}_{t,a}$
        $v_{t,a} = \mathrm{x}_{t,a}^{\intercal} \mathrm{w}_{t,a}$
        $p_{t,a} \leftarrow \hat{r}_{t,a} + \alpha \sqrt{v_{t,a}}$
    end
    Choose arm $a_t = \arg\max_{a \in \mathcal{A}_t} p_{t,a}$ with ties broken arbitrarily and observe a real-valued payoff r_t
    $\mathbf{A}_{t+1,a_t}^{-1} \leftarrow \mathbf{A}_{t,a_t}^{-1} - \frac{\mathrm{w}_{t,a_t} \mathrm{w}_{t,a_t}^{\intercal}}{1+v_{t,a_t}}$
    $\hat{\boldsymbol{\theta}}_{t+1,a_t} \leftarrow \hat{\boldsymbol{\theta}}_{t,a_t} + \mathrm{w}_{t,a} \frac{r_{t,a_t} - \hat{r}_{t,a_t}}{1+v_{t,a_t}}$
end

Algorithm 2: The matrix inverse-free LinUCB algorithm