Corrected KernelUCB algorithm

September 26, 2018

Algorithm 1 KernelUCB with online updates

Input: N the number of actions, T the number of pulls, γ, η regularization and exploration parameters, $k(\cdot,\cdot)$ kernel function

Initialize:

$u_0 \leftarrow [1, 0, \ldots, 0]^\intercal$ ▷ play action 1 in first round

$y_0 \leftarrow [\,]$ ▷ empty vector for reward history

Run:

for $t \in 1, \ldots, T$ **do**

  Receive contexts $\{x_{1,t}, \ldots, x_{N,t}\}$

  Choose action $a \leftarrow \arg\max u_{t-1}$ and receive reward r_{t-1}

  Store context for action a: $x_t \leftarrow x_{a,t}$

  Update reward history: $y_t \leftarrow [r_1, \ldots, r_{t-1}]^\intercal$

  Compute $k_{x_t,t} \leftarrow [k(x_t, x_1), k(x_t, x_2), \ldots, k(x_t, x_t)]^\intercal$

  if $t = 1$ **then** ▷ initialise kernel matrix inverse

    $K_t^{-1} \leftarrow (k(x_t, x_t) + \gamma)^{-1}$

  else ▷ online update of kernel matrix inverse

    $b \leftarrow k_{x_t,t-1}$

    $K_{22} \leftarrow \left(k(x_t, x_t) + \gamma - b^\intercal K_{t-1}^{-1} b\right)^{-1}$

    $K_{11} \leftarrow K_{t-1}^{-1} + K_{22} K_{t-1}^{-1} b b^\intercal K_{t-1}^{-1}$

    $K_{12} \leftarrow -K_{22} K_{t-1}^{-1} b$

    $K_{21} \leftarrow -K_{22} b^\intercal K_{t-1}^{-1}$

    $K_t^{-1} \leftarrow \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$

  end if

  for $n \in \{1, \ldots, N\}$ **do**

    $\sigma_{n,t} \leftarrow \sqrt{k(x_{n,t}, x_{n,t}) - k_{x_t,t}^\intercal K_t^{-1} k_{x_t,t}}$

    $u_{n,t} \leftarrow k_{x_t,t}^\intercal K_t^{-1} y_t + \frac{\eta}{\sqrt{\gamma}} \sigma_{n,t}$

  end for

end for

Unlike the WWW2010 LinUCB which models each arm's rewards separately, KernelUCB paper adopts a pooled model since the paper focuses on $N \gg T$—too many actions to be all pulled over the horizon.

Use the above writeup which clears up the main errors/ambiguities in the original UAI2013 writeup:

- Unclear that x_t refers to the context for the pulled action a at round t, not the contexts for *all* actions at round t.
- Unclear that $k_{x,x'}$ means $k(x, x')$. Made more confusing by the fact that $k_{x,t}$ already means something else (as defined above equation 4).
- Unclear that k_{x_1}, k_{x_2} etc. means $k(x_t, x_1)$, $k(x_t, x_2)$, etc.
- Error in expression for initialisation of kernel matrix inverse: γ should be in the denominator.
- Error in expression for K_{12}: kernel matrix is missing inverse.
- Inconsistent notation: k_{x_t,x_t}, $k(x_{a,t}, x_{a,t})$, k_{x_a,x_a} all mean the same thing.
- Variable a is overloaded (not necessarily incorrect, but could lead to confusion). It's first defined as the action drawn in round t, then reused as an index variable in the for-loop at the end.
- arg max should still be done with random tie breaking.