Corrected KernelUCB algorithm

October 8, 2018

Using the following definitions (clears up inconsistencies):

- $k_{x^\star,t} = [k(x^\star, x_1), \dots k(x^\star, x_t)]^\intercal$
- $y_t = [r_1, \dots, r_t]^\intercal$
- $\Phi_t = [\phi(x_1)^\intercal, \dots, \phi(x_t)^\intercal]^\intercal$
- $K_t = \Phi_t^\intercal \Phi_t$
- $\hat{\sigma}_{a,t+1} = \left[\phi(x_{a,t+1})^\intercal (\Phi_t^\intercal \Phi_t + \gamma I)^{-1} \phi(x_{a,t+1})\right]^{\frac{1}{2}}$
- $\hat{\mu}_{a,t+1} = k_{x_{a,t+1},t}^\intercal (K_t + \gamma I)^{-1} y_t$

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

for $t \in 1, \dots, T$ **do**
  Receive contexts $\{x_{1,t}, \dots, x_{N,t}\}$
  if $t = 1$ **then**
    $u_t \leftarrow [1, 0, \dots, 0]^\intercal$
  else
    for $n \in \{1, \dots N\}$ **do**
      $\sigma_{n,t} \leftarrow \left[k(x_{n,t}, x_{n,t}) - k_{x_{n,t},t-1}^\intercal K_{t-1}^{-1} k_{x_{n,t},t-1}\right]^{\frac{1}{2}}$
      $u_{n,t} \leftarrow k_{x_{n,t},t-1}^\intercal K_{t-1}^{-1} y_{t-1} + \frac{\eta}{\sqrt{\gamma}} \sigma_{n,t}$
    end for
  end if
  Choose action $a_t \leftarrow \arg\max u_t$ and receive reward r_t
  Store context for action a_t: $x_t \leftarrow x_{a,t}$
  Update reward history: $y_t \leftarrow [r_1, \dots, r_t]^\intercal$
  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
end for