Batch Online Regression with noise reduction

Ilyass OUBAIK

June 2019

1 Optimization problem

$$C(w) = \frac{||w - w_{t-1}||_B}{2} + \frac{\alpha_B(h)}{r}||\mathbf{Y} - w\mathbf{X}^\top||_B$$

- w : weights at time t
- w_{i-1} : previous weigths at time $t-1$
- B : Batch
- $\alpha_B(h)$: Noise parameter
- r : robustness parameter (hyperparamter)
- Y : N-vector of outputs at time t
- X : (N,p) matrix at time t (Observation x features)

2 Hyperparameters

- Noise parameter:

$$\alpha_B(h) = \frac{1}{1 + h\sigma_B}$$

  - σ_B : noise in N vector(Y) output at time t (ex : $\sigma_B = \frac{1}{N}\sum \sigma_{Bi} = \mathbf{E}_B[\sigma_B]$)
  - h : hyper-parameter

- robustness parameter

  - if $r \to \infty$ then $C(w) \simeq \frac{||w-w_{t-1}||_B}{2}$ then $\min C(w) = w_{t-1}$ (ex for noisy batch ($\sigma_B \gg 0$))
  - if $r \to 0$ then $C(w) \simeq ||Y - w^T X||_B$ then $\min C(w) = (X^\top X)^{-1}X^\top Y$ (ex non noisy batch ($\sigma_B \simeq 0$))

3 Model update

$$w_t = \min C(w)$$

let's solve $\frac{\partial C(w)}{\partial w} = 0$

$$\begin{aligned}
\frac{\partial C(w)}{\partial w} &= \frac{\partial \frac{||w-w_{t-1}||_B}{2}}{\partial w} + \frac{\alpha_B(h)}{r} \frac{\partial ||\mathbf{Y} - w\mathbf{X}^\top||_B}{\partial w} \\
&= w - w_{t-1} + \frac{\alpha_B(h)}{r}(X^\top X w - X^\top Y) \\
&= w - w_{t-1} + \frac{\alpha_B(h)}{r} X^\top X w - \frac{\alpha_B(h)}{r} X^\top Y \\
&= (\mathbf{I} + \frac{\alpha_B(h)}{r} X^\top X) w - w_{t-1} + \frac{\alpha_B(h)}{r} X^\top Y
\end{aligned} \tag{1}$$

Result $\boxed{w = (\mathbf{I} + \frac{\alpha_B(h)}{r} X^\top X)^{-1}(w_{t-1} + \frac{\alpha_B(h)}{r} X^\top Y)}$