A Deterministic Streaming Sketch for Ridge Regression

Benwei Shi and Jeff M. Phillips

THE UNIVERSITY OF UTAH®

Ridge Regression

Given : $\mathbf{A} \in \mathbb{R}^{n \times d}, \mathbf{b} \in \mathbb{R}^{n}$ in rows.

Goal : $\mathbf{x}_{\gamma} \equiv \underset{\mathbf{x} \in \mathbb{R}^{d}}{\operatorname{argmin}}(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \gamma \|\mathbf{x}\|^2)$

$$= \underbrace{(\mathbf{A}^{\top}\mathbf{A} + \gamma \mathbf{I})^{-1}}\ \underbrace{\mathbf{A}^{\top}\mathbf{b}}$$

Space:	$O(d^2)$	$O(d)$
Time:	$O(d^3)$	$O(nd)$

Our approach: Use Frequent Directions (FD) (Liberty, 2013) to estimate $\mathbf{A}^{\top}\mathbf{A}$ in stream.

Frequent Directions Ridge Regression

Algorithm FDRR (Based on Frequent Directions)

Input: $\mathbf{A} \in \mathbb{R}^{n\times d}, \mathbf{b} \in \mathbb{R}^{n}, \ell, \gamma$

$\boldsymbol{\Sigma} \leftarrow \mathbf{0}^{\ell\times\ell}, \mathbf{V}^\top \leftarrow \mathbf{0}^{\ell\times d}, \mathbf{c} \leftarrow \mathbf{0}^{d}$ — Initialization

$\mathbf{C} = \boldsymbol{\Sigma}\mathbf{V}^\top$

for batch $\mathbf{A}_\ell \in \mathbf{A}, \mathbf{b}_\ell \in \mathbf{b}$ do — size ℓ batch $\mathbf{A}_\ell$ and $\mathbf{b}_\ell$

$\quad \_, \boldsymbol{\Sigma}', \mathbf{V}'^\top \leftarrow \mathbf{svd}\left(\left[\mathbf{C}^\top; \mathbf{A}_\ell^\top\right]^\top\right)$ — Frequent Directions

$\quad \boldsymbol{\Sigma} \leftarrow \sqrt{\boldsymbol{\Sigma}_\ell'^{2} - \sigma_{\ell+1}^2 \mathbf{I}_\ell}$

$\quad \mathbf{V} \leftarrow \mathbf{V}'_\ell$

$\quad \mathbf{C} = \boldsymbol{\Sigma}\mathbf{V}^\top$

$\quad \mathbf{c} \leftarrow \mathbf{c} + \mathbf{A}_\ell^\top \mathbf{b}_\ell$ — Compute $\mathbf{A}^\top\mathbf{b}$ on the fly

end for

$\mathbf{c}' = \mathbf{V}^\top\mathbf{c}$

$\hat{\mathbf{x}}_\gamma \leftarrow \mathbf{V}(\boldsymbol{\Sigma}^2 + \gamma\mathbf{I})^{-1}\mathbf{c}' + \gamma^{-1}(\mathbf{c} - \mathbf{V}\mathbf{c}')$

return ~~C~~ $\hat{\mathbf{x}}_\gamma$

Return the solution $\hat{\mathbf{x}}_\gamma = (\mathbf{C}^\top\mathbf{C} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\mathbf{b} = (\mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^\top + \gamma\mathbf{I})^{-1}\mathbf{c}$

Recall the RR solution $\mathbf{x}_\gamma = (\mathbf{A}^\top\mathbf{A} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\mathbf{b}$

Frequent Directions Ridge Regression

Running time: $O(nd\ell)$, required space: $O(d\ell)$. Note that $\ell \leq d$.

If

$$\ell \geq \frac{\|\mathbf{A} - \mathbf{A}_k\|_F^2}{\varepsilon\gamma} + k, \qquad \text{or} \qquad \gamma \geq \frac{\|\mathbf{A} - \mathbf{A}_k\|_F^2}{\varepsilon(\ell - k)}$$

Then

- $\left\|\hat{\mathbf{x}}_\gamma - \mathbf{x}_\gamma\right\| \leq \varepsilon \left\|\mathbf{x}_\gamma\right\|$, or the coefficient error $\frac{\|\hat{\mathbf{x}}_\gamma - \mathbf{x}_\gamma\|}{\|\mathbf{x}_\gamma\|} \leq \varepsilon$
- $\left|\hat{\mathbf{x}}_\gamma^\top \mathbf{a} - \mathbf{x}_\gamma^\top \mathbf{a}\right| \leq \varepsilon \left\|\mathbf{x}_\gamma\right\| \|\mathbf{a}\|$ for any $\mathbf{a} \in \mathbb{R}^d$
- $\mathcal{B}^2(\hat{\mathbf{x}}_\gamma) \leq \left(1 + \frac{\varepsilon^2}{\gamma^2} \|\mathbf{A}\|_2^4\right) \mathcal{B}^2(\mathbf{x}_\gamma)$
- $\mathcal{V}(\hat{\mathbf{x}}_\gamma) \leq \left(1 + \frac{1}{\gamma} \|\mathbf{A}\|_2^2\right) \mathcal{V}(\mathbf{x}_\gamma)$

Experiments