Sketching Algorithms and Lower Bounds for Ridge Regression

Praneeth Kacham and David Woodruff
Carnegie Mellon University

Introduction

- We study fast algorithms for "Ridge Regression".

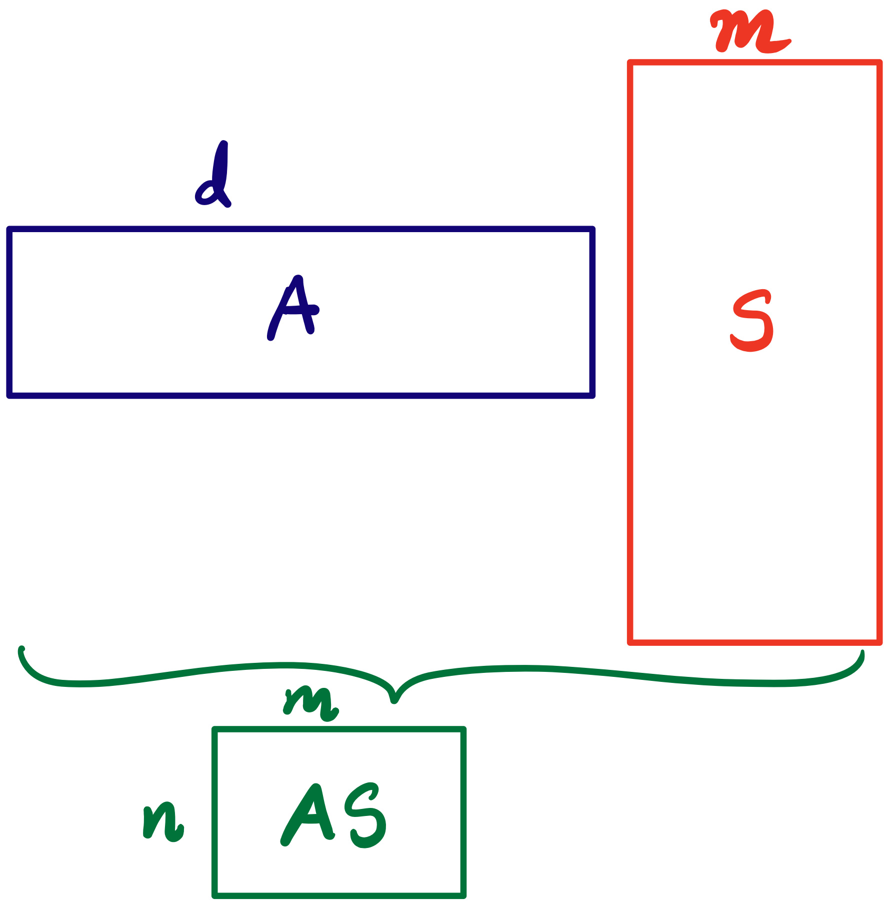
$$\min_x \|Ax - b\|_2^2 + \lambda \|x\|_2^2 \qquad A \in \mathbb{R}^{n \times d}, \quad b \in \mathbb{R}^n$$

- Linear Regression + ℓ_2 regularization
- Underdetermined Case : $n \leq d$
- Optimal $x^* = \underbrace{A^T}_{d \times n} (\overbrace{AA^T + \lambda I_d}^{n \times n})^{-1} \underbrace{b}_{n \times 1}$

  $n^2 d$ time to compute

- Very slow if n & d are large

Sketching



Req. for S:

(i) $m \ll d$

(ii) Should be able to compute AS quickly

(iii) Should of course be useful downstream

Earlier Work for 1-pass algorithms

- Chowdhury et al., used subspace embeddings to sketch the matrix A

- $\forall x$:

  ε- Subspace Embedding

  $$\|x^T A S\|_2 = (1 \pm \varepsilon)\|x^T A\|_2$$

- We know many distributions that satisfy this property:
  - Gaussians, SRHT, CountSketch, OSNAP

- As ε goes down, m and t_{AS} typically increase

- Let $\tilde{x} = A^T(AS(AS)^T + \lambda I)^{-1} b$

- Chowdhury et al., showed that if "S" is a $\sqrt{\varepsilon\lambda/\|A\|_2^2}$ subspace embedding, then

$$\|A\tilde{x} - b\|_2^2 + \lambda\|x\|_2^2 \leq (1+\varepsilon)\cdot \mathrm{Opt}$$

- We obtain faster algorithms by requiring weaker guarantee from S:

Weaker Req's

- S is a ½ subspace embedding
- For arbitrary orthonormal matrix V and vec. r:

(AMM) $$\|V^T S S^T V r - r\|_2 \leq \frac{\sqrt{\varepsilon\lambda/\sigma^2}}{\sqrt{n}} \|V\|_F \|r\|_2$$

Near-Optimality of Sketch Sizes:

- Guarantees satisfied by OSNAP with $m = \tilde{O}\left(\frac{n\sigma^2}{\lambda\varepsilon}\right)$
- Can also quickly compute AS
- We show near optimality of sketch size by reducing from Ridge Regression $\Rightarrow$ AMM
- We then prove tight lowerbounds for AMM:

$$\varepsilon\text{-AMM} \Rightarrow \Omega\left(1/\varepsilon^2\right) \text{ rows in sketch.}$$

Thank You