

# Motivation from LS

"Leverage scores capture the leverages of each example on the best fit."

$$\min_x \|Ax - b\|^2, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

$$x^* = (A^T A)^{-1} A^T b$$

$$\hat{b} := Ax^* = A(A^T A)^{-1} A^T b$$

$$A = \sum_{i=1}^n \sigma_i u_i v_i^T \quad (S \quad D)$$



$$A(A^T A)^{-1} A^T = \left( \sum \sigma_i u_i v_i^T \right) \left( \sum \sigma_i^{-1} v_i v_i^T \right)$$

$$\left( \sum \sigma_i v_i u_i^T \right) = \sum u_i u_i^T$$

This is the Projection matrix onto  $\text{col}(A)$ .

Denote by  $H = \sum_{i=1}^n u_i u_i^T$ .

Def: The  $i$ -th leverage

Score of  $A$  is

$$h_i = H(i, i) = \|U(i, :)\|^2,$$

$$\text{where } U = \begin{pmatrix} u'_1 & \dots & u'_n \\ | & & | \\ | & & | \end{pmatrix}_{m \times n}$$

Observations

a.  $h_i \in [0, 1],$

$$\underline{b.} \quad \sum_{i=1}^n \lambda_i = \sum_{i=1}^n U(i,:) ^2 = \|U\|_F^2$$

$$= \sum_{i=1}^n \|u_i\|^2 = n$$

$$\underline{c.} \quad H(i,:) = \sum_{j=1}^n U(i,j) u_j^T$$

$$\Rightarrow \|H(i,:)\|^2 = \|U(i,:)\|^2$$

$$= \lambda_i$$

d. Therefore, if  $\lambda_i = 1$

then  $(\forall j \neq i) \quad H_{i,j} = \langle u(i,:), u(j,:) \rangle$

$$= 0$$



c. Therefore, if  $x_i = 1$

$$\hat{b}_i = H(i, :) \cdot b = H_{i,:} \cdot b = b_i$$

Similarly if  $x_i$  is small,  
then  $b_i$  has little effect  
on  $\hat{b}_i$ .

Example:  $x_1 = 6, x_2 = 8$  so

$$A = \begin{pmatrix} 6 \\ 8 \end{pmatrix} = \underbrace{\begin{pmatrix} 0.6 \\ 0.8 \end{pmatrix}}_U (10) \cdot (1)$$

$$UU^T = \begin{pmatrix} 0.36 & 0.48 \\ 0.48 & 0.64 \end{pmatrix}$$

Given  $y_1, y_2$   <sup>$b_1$</sup>   <sup>$b_2$</sup>  - we have

$$\hat{y}_1 = 0.36 \cdot y_1 + 0.48 y_2$$

$$\hat{y}_2 = 0.48 y_1 + 0.64 y_2$$

The more distant example is more influential.

Example Let  $x_1=0, x_2=1$ .

Then  $r_2=1$  so we exactly fit  $x_2$ .

TODO: d. App to low-rank

a. Analyze sampling ✓

b. How to approx  $r$ ;

c. ~~FL makes leverage~~

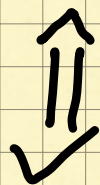
~~scores uniform~~

# Leverage<sup>scores</sup> sampling

Goal: relative error bounds.

Given  $A \in \mathbb{R}^{n \times d}$ , we would like to sample rows of  $A$  s.t. w.h.p. the rescaled sampling matrix  $S \in \mathbb{R}^{s \times n}$  satisfies w.h.p.:

$$(\forall x) \quad \|SAx\|^2 = (1 \pm \epsilon) \|Ax\|^2$$



$$(\forall x) \quad |x^T A^T S^T S A x - x^T A^T A x| \leq \epsilon \underbrace{x^T A^T A x}_{(*)}$$

As opposed

$$\leq \epsilon \|x\|^2$$

Often more ambitious since

if  $x$  corresponds to tiny eigenvalue of  $A^T A$ , the RHS



is small.

Preconditioning: How can we express the desired inequality as spectral norm bound?

Let  $x \in \mathbb{R}^d$ . We can restrict our attention to  $x \in \text{row}(A)$ . Then

$\exists y \in \mathbb{R}^d$  s.t.  $x = V \Sigma^{-1} y$ , where

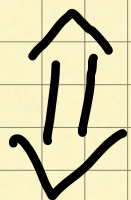
$A = U \Sigma V^T$  is the SVD of  $A$ .

Since  $A V \Sigma^{-1} = U \Sigma V^T V \Sigma^{-1}$

$= U$ , we obtain that (\*)

is equivalent to

$$(\forall y) \quad |y^T U^T S^T S U y - y^T U^T U y| \leq \varepsilon \|y\|^2$$



$$\|U^T S^T S U - U^T U\|_2 \leq \epsilon$$

↑↑

Sample  $s \geq c d \ln(n) / \epsilon^2$  rows

according to squared-length sampling according to  $U$ !

$\Leftrightarrow$  Leverage score sampling!

Obstacle: computing leverage scores (exactly) requires computing the SVD. We'll

show how to approximate them using sketching



# Ridge leverage scores

Idea: replace  $\text{rank}(A)$  by "effective rank" of  $A$ . For parameter  $\lambda > 0$ , we aim at sampling rows of  $A$  s.t. w.h.p.

$$(\forall x) \quad |x^T A^T S^T S A x - x^T A^T A x| \leq \underbrace{\varepsilon x^T (A^T A + \lambda I) x}_{\text{relative error} + \text{small slack}} \quad (**)$$

Notation:  $A = U \Sigma V^T$ ,  
 $A^T A + \lambda I = V \bar{\Sigma}^2 V^T$

Given  $x \in \mathbb{R}^d \exists y \text{ s.t. } x = V \bar{\Sigma}^{-1} y.$

Then  $(**) \Leftrightarrow (\forall y)$

$$|y^T \bar{\Sigma}^{-1} \Sigma U^T S^T S U \Sigma \bar{\Sigma}^{-1} y - y^T \bar{\Sigma}^{-1} \Sigma U^T U \Sigma \bar{\Sigma}^{-1} y| \leq \epsilon \|y\|^2$$

↑↑

Sample according to rows of

$Z = U \Sigma \bar{\Sigma}^{-1}$ . That is,

$$p_i = \sum_j u_{ij}^2 \sigma_j^2 / (\sigma_j^2 + \lambda)$$

$$= a_i^T (A^T A + \lambda I)^{-1} a_i$$

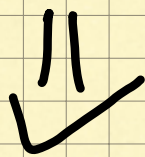
$$\tilde{\ell}_i :=$$

Ridge leverage score

$$\|Z\|_2 \leq 1, \|Z\|_F^2 = \Sigma \tilde{\lambda}_i$$

$$= \text{tr} \left( A^T A (A^T A + \lambda I)^{-1} \right) =: d_\lambda$$

effective  
dimension



$$\text{Sample size} \approx d_\lambda / \varepsilon^2$$



# Exercises

① Let  $A = \begin{pmatrix} -a_1 \\ \vdots \\ -a_n \end{pmatrix} \in \mathbb{R}^{n \times d}$  and suppose that  $a_i \notin \text{span}\{a_j : j \neq i\}$ .

Show that  $\lambda_i = 1$ .

② Show <sup>that</sup>  $\forall \lambda_i = \min \{ \|x\|^2 : x \in \mathbb{R}^n, A^T x = a_i \}$ .

Use this to reprove Ex. 1.

③ Prove that  $\lambda_i = \min \{ t \geq 0 : a_i a_i^T \leq t A^T A \}$

# Solutions

① Let  $A_{-i}$  correspond to removing  $a_i$ .  
Then  $\forall j \neq i, r_j(A_{-i}) \geq r_j(A)$ . Since adding  $a_i$  increases the rank by 1, we have

$$\begin{aligned} \text{rank}(A_{-i}) &= \sum_{j \neq i} r_j(A_{-i}) \geq \sum_{j \neq i} r_j(A) \geq \sum_{j \neq i} r_j(A) + (r_i(A) - 1) \\ &= \text{rank}(A) - 1. \end{aligned}$$

Since the LHS is equal to the RHS, all the inequalities are equalities, hence  $r_i(A) = 1$ .

② Define  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  by  $f(x) = \|x\|^2$ .

$$\min f(x)$$

$$\text{s.t. } A^T x = a_i$$

|| (opt conditions)

$$2x = \nabla f(x) \in \text{row}(A^T)$$

||

$$x^* = (A^T)^+ a_i$$

||

$$\|x^*\|^2 = a_i (A^T)^+ A^+ a_i = a_i (A^T A)^+ a_i$$

Reproving 1: Suppose  $a_i \notin \text{SP}\{a_j : j \neq i\}$ .

Then given  $x$  satisfying  $a = A^T x = \sum x_j a_j$ , we have

$$a = \underbrace{\tilde{a}_i}_{\text{SP}} + \underbrace{\bar{a}_i}_{\text{SP}} = \sum_{j \neq i} x_j a_j + x_i a_i$$

Taking inner product with  $\bar{a}_i$ , we get

$$\|\bar{a}_i\|^2 = x_i a_i^T \bar{a}_i = x_i \|\bar{a}_i\|^2$$

||

$$x_i = 1$$



$$\textcircled{3} \Rightarrow A^T x = a, \quad \|x\|^2 = 8;$$

$$a; a^T = A^T x x^T A$$

$$\begin{aligned} (\forall v) \quad v^T a; a^T v &= v^T A^T x x^T A v \leq \|x x^T\|_2 v^T A^T A v \\ &= \|x\|^2 v^T A^T A v = 8; v^T A^T A v \end{aligned}$$

Taking

$$A v = x = (A^T)^+ a;$$

$$v = (A A^T)^+ a;$$

$\Downarrow$

$$v^T A^T x x^T A v = \|x\|^4$$

$$v^T A^T A v = \|x\|^2$$

Equality holds