Motivation from LS

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

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

$$x^* = (A^TA)^{-1}A^Tb$$

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

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

$$\Downarrow$$

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

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

This is the Projection matrix onto col(A).

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

Def: The i-th leverage score of A is

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

where $U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{pmatrix}_{m \times n}$

Observations

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

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

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

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

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

$= \ell_i$

d. Therefore, if $\ell_i = 1$

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

$= 0$

$\underline{e}$. Therefore, if $\ell_i = 1$

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

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

<u>Example:</u> $x_1 = 6$, $x_2 = 8$ 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 $\overset{b_1}{\overset{\shortparallel}{y_1}}, \overset{b_2}{\overset{\shortparallel}{y_2}}$, 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.

<u>Example</u> Let $x_1 = 0$, $x_2 = 1$.

Then $\ell_2 = 1$ so we exactly fit x_2.

<u>TODO:</u>

<u>a.</u> Analyze sampling ✓

<u>b.</u> How to approx ℓ;

<u>c.</u> ~~JL makes leverage scores uniform~~

<u>d.</u> App to low-rank

Leverage scores 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.:

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

$$\Updownarrow$$

$$(\forall x) \quad |x^T A^T S^T S A x - x^T A^T A x| \leq \varepsilon \underbrace{x^T A^T A x}_{\text{As opposed to } \|x\|^2} \tag{*}$$

Often more ambitious since if x corresponds to tiny eigenvalue of $A^T A$, the RHS is small.

<u>Preconditioning</u>: How can we express the desired inequality as spectral norm bound?

Let $x \in \mathbb{R}^d$. We can restrict our attention to $x \in row(A)$. 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 $AV\Sigma^{-1} = U\Sigma V^T V \Sigma^{-1} = U$, we obtain that $(*)$ is equivalent to

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

$$\Updownarrow$$

$$\| U^T S^T S U - U^T U \|_2 \le \varepsilon$$

$$\Uparrow$$

Sample $s \ge c d \ln(n) / \varepsilon^2$ rows according to squared-length sampling according to U!

$\iff$ Leverage score sampling!

<u>Obstacle</u>: computing leverage scores (exactly) requires computing the SVD. We'll

show how to approximate them using sketching

Ridge leverage scores

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

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

Notation: $A = U \Sigma V^T$,

$$A^T A + \lambda I = V \bar{\Sigma}^2 V^T$$

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

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

$$\left| y^T \bar{\Sigma}^{-1} \Sigma U^T S^T S U \Sigma \bar{\Sigma}^{-1} y - y^T \bar{\Sigma}^{-1} \Sigma U^T U \Sigma \bar{\Sigma}^{-1} y \right| \le \varepsilon \|y\|^2$$

$$\Uparrow$$

Sample according to rows of $Z = U \Sigma \bar{\Sigma}^{-1}$. That is,

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

$$= \underbrace{a_i^T (A^T A + \lambda I)^{-1} a_i}_{\tilde{\ell}_i :=}$$

Ridge leverage score

$$\|Z\|_2 \le 1, \quad \|Z\|_F^2 = \sum \tilde{\lambda}_i$$

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

effective dimension

$\Downarrow$

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 $a_i \notin \mathrm{span}\{a_j : j \neq i\}$.

Show that $\ell_i = 1$.

(2) Show that $\ell_i = \min \{ \|x\|^2 : x \in \mathbb{R}^n, A^T x = a_i \}$.

Use this to reprove Ex. 1.

(3) Prove that $\ell_i = \min \{ t \geq 0 : a_i a_i^T \preceq t A^T A \}$

Solutions

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

$$\text{rank}(A_{-i}) = \sum_{j \neq i} \ell_j(A_{-i}) \geq \sum_{j \neq i} \ell_j(A) \geq \sum_{j \neq i} \ell_j(A) + (\ell_i(A) - 1)$$

$$= \text{rank}(A) - 1.$$

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

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

$$\min \; f(x)$$
$$\text{s.t.} \; A^T x = a_i$$

$\Downarrow$ (opt conditions)

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

$\Downarrow$

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

$\Downarrow$

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

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

$$\underset{i}{a} = \underset{\underset{\text{sp}}{\in}}{\tilde{a}_i} + \underset{\underset{\text{sp}}{\perp}}{\bar{a}_i} = \sum_{j \neq i} x_j a_j + x_i a_i$$

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

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

$\Downarrow$

$$x_i = 1$$

③ "$\Rightarrow$" $A^T x = a_i$, $\|x\|^2 = \ell_i$

$$\Downarrow$$

$$a_i a_i^T = A^T x x^T A$$

$$(\forall v)\ v^T a_i a_i^T v = v^T A^T x x^T A v \le \|x x^T\|_2\, v^T A^T A v$$

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

Taking $\quad Av = x = (A^T)^+ a_i$

$$v = (A A^T)^+ a_i$$

$$\Downarrow$$

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

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

Equality holds