Welcome back to Ma221! Lecture 20, Oct 9

Dealing with (nearly) low rank matrices

Motivation: Real data often low rank (or nearly redundant)

1. take precautions to avoid inaccurate LS solution
2. use it to compress data, go faster, both deterministic and randomized

use LS to illustrate compression useful elsewhere.

Solving a LS problem when A rank deficient

Thm: A $m \times n$ $m \geq n$ rank $r < n$

$$A = U\Sigma V^T = \begin{array}{c} \\ m \end{array}\begin{array}{c} \begin{array}{ccc} r & n-r & m-n \end{array} \\ \left[U_1, U_2, U_3 \right] \end{array} \begin{array}{c} \\ r \\ n-r \\ m-n \end{array}\begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ n \end{array}\begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \left[V_1, V_2 \right]^T \end{array}$$

Σ_1 full rank r

$\Sigma_2 = 0$

The set of vectors minimizing $\|Ax-b\|_2$ is $\{ x = V_1 \Sigma_1^{-1} U_1^T b + V_2 y_2, \text{ any } y_2 \in \mathbb{R}^{n-r} \}$

Unique x minimizing $\|Ax-b\|_2$ and $\|x\|_2$ is gotten from $y_2 = 0$

$x = \underline{V_1 \Sigma_1^{-1} U_1^T} b$

Def: $A^+ = V_1 \Sigma_1^{-1} U_1^T$ is Moore-Penrose pseudoinverse of A (includes full rank $r=n$ case)

(in practice Σ_2 will be all singular values less than user-defined threshold)

So square or not, full rank or not "best solution" is $x = A^+ b$

Proof: $\|Ax - b\|_2 = \|U \Sigma V^T x - b\|_2$

$= \|\Sigma V^T x - U^T b\|_2$ since U orthog

$= \|\Sigma y - U^T b\|_2$ where $y = V^T x$

$\|x\|_2 = \|y\|_2$ so okay to minimize either one

$$= \left\| \begin{bmatrix} \Sigma_1 y_1 - U_1^T b \\ -U_2^T b \\ -U_3^T b \end{bmatrix} \right\|_2 \qquad y = \begin{array}{c} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \end{array} \begin{array}{l} r \\ n-r \end{array}$$

minimized by $y_1 = \Sigma_1^{-1} U_1^T b$

and $\|x\|_2^2 = \|y\|_2^2 = \|y_1\|_2^2 + \|y_2\|_2^2$ minimized by $y_2 = 0$

$x = Vy = [V_1, V_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = V_1 y_1 + V_2 y_2$

$= V_1 \Sigma_1^{-1} U_1^T b$

$= A^+ b$

Solving LS when A (nearly) rank deficient using truncated SVD

$\kappa(A) = \frac{\sigma_{max}}{\sigma_{min}} = \infty$ if A low rank

$$\underset{x}{\operatorname{argmin}} \left\| \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\|_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\underset{x}{\operatorname{argmin}} \left\| \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\|_2 = \begin{bmatrix} 1 \\ 1/\epsilon \end{bmatrix} \quad \epsilon \text{ tiny}$$

What does a "solution" mean if it can change discontinuously?

Often A not known exactly, just up to some tolerance $\|A - A'\|_2 \le tol$

What to do?

Def: Truncated SVD:

if $A = U \Sigma V^T$

$$A(tol) = U\, \Sigma(tol)\, V^T$$

$$\Sigma(tol) = \operatorname{diag}(\sigma_1(tol), \sigma_2(tol), \ldots, \sigma_n(tol))$$

$$\sigma_i(tol) = \begin{cases} \sigma_i & \text{if } \sigma_i \ge tol \\ 0 & \text{if } \sigma_i < tol \end{cases}$$

$A(tol)$ = lowest rank matrix within distance tol to A

Use $A(tol)$ instead of A for LS

reduces $\kappa = \frac{\sigma_{max}}{\sigma_{min}}$ to $\frac{\sigma_{max}}{tol}$

tol is a "knob" to trade off sensitivity and how well LS problem can be solved, i.e. how small you can make $\|Ax-b\|_2$

Replacing A by an "easier" matrix, called regularization, $A(tol)$ is one way, others too

Lemma: $x_1 = \arg\min_x \|A(tol)x - b_1\|_2$

$x_2 = \arg\min_x \|A(tol)x - b_2\|_2$

choose x_i of smallest norm

then $\|x_1 - x_2\|_2 \le \frac{\|b_1 - b_2\|_2}{tol}$

proof: $\|x_1 - x_2\|_2 = \|(A(tol))^+ (b_1 - b_2)\|_2$

$= \| V (\Sigma(tol))^+ U^T (b_1 - b_2)\|_2$

$= \| \text{diag}(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots \frac{1}{\sigma_k}, 0 \dots 0) U^T (b_1 - b_2)\|_2$

last nonzero $\sigma_k \ge tol$

$\le \frac{1}{\sigma_k} \|b_1 - b_2\|_2$

$\le \frac{1}{tol} \|b_1 - b_2\|_2$

How does $A(tol)$ depend on tol?

piecewise continuous, changes when $tol = \sigma_i$

An advantage of $A(tol)$:

we can use it for compression of A

$A(tol)$ has rank $k \Rightarrow$ need

$m \cdot k$ words for U_1

$+ k$ words for Σ_1

$+ n \cdot k$ words for V_1

to store SVD, vs $m \cdot n$ for full svd

$k \cdot (m + 1 + n)$ can be $<< m \cdot n$ if k small

Solving (nearly) low rank LS

using Tikhonov regularization

or ridge regression

Replace $\underset{x}{\text{argmin}} \|Ax - b\|_2^2$

by $\underset{x}{\text{argmin}} \|Ax - b\|_2^2 + \lambda \|x\|_2^2$

you choose $\lambda > 0$

λ "penalizes" very large x

tuning parameter like tol

$$\underset{x}{\text{argmin}} \|Ax - b\|_2^2 + \lambda \|x\|_2^2$$

$$= \underset{x}{\text{argmin}} \left\| \underbrace{\begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix}}_{\text{full rank } \forall \lambda > 0} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

$$NE \Rightarrow \quad x = \left(\begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix} \right)^{-1} \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix}^T b$$

$$(*) \qquad = (A^T A + \lambda I)^{-1} A^T b$$

$\Rightarrow$ add λ to diagonal of NE

How does λ change SVD solution?

plug $A = U \Sigma V^T$ into $(*)$

$$x = V(\Sigma(\Sigma^2 + \lambda I)^{-1}) U^T b$$

$$= V \operatorname{diag}\left(\frac{\sigma_i}{\sigma_i^2 + \lambda} \right) \cdot U^T b$$

usual solution if $\lambda = 0$

$$\sigma_i >> \lambda^{1/2} \Rightarrow \frac{\sigma_i}{\sigma_i^2 + \lambda} \sim \frac{1}{\sigma_i}$$

$$\sigma_i < \lambda^{1/2} \Rightarrow \frac{\sigma_i}{\sigma_i^2 + \lambda} \leq \frac{1}{\lambda^{1/2}}$$

i.e. $\lambda^{1/2}$ and tol in A(tol)

play analogous roles

but solution a smooth function of λ

Solving low rank LS with QR

"QR with column pivoting"

Suppose we did $A = QR$ exactly

for A: $\operatorname{rank}(A) = r < n$.

What would R look like?

If leading r columns of A were independent (true for "almost all" A)

$$R = \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{c} r \\ n-r \end{array} & \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & 0 \end{array}\right] \end{array}$$

R_{11} full rank

$R_{22} = 0$

If A nearly low rank, hope that

$\|R_{22}\| < tol$, set $R_{22} = 0$

Assuming true, solve LS as follows

$$A = Q \cdot R = \begin{array}{c c} & \begin{array}{cc} n & m-n \end{array} \\ m & [Q, Q'] \end{array} \begin{bmatrix} R \\ 0 \end{bmatrix}$$

$$= \begin{array}{c c} & \begin{array}{ccc} r & n-r & m-n \end{array} \\ m & [Q_1, Q_2, Q'] \end{array} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}$$

$$\operatorname*{argmin}_x \|Ax - b\|_2$$

$$= \operatorname*{argmin}_x \left\| [Q_1, Q_2, Q'] \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} x - b \right\|_2$$

$$= \operatorname*{argmin}_x \left\| \begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} \end{array} x - \begin{bmatrix} Q_1^T b \\ Q_2^T b \\ Q'^T b \end{bmatrix} \right\|_2 \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{array}{l} r \\ n-r \end{array}$$

$$= \operatorname*{argmin}_x \left\| \begin{bmatrix} R_{11} x_1 + R_{12} x_2 - Q_1^T b \\ Q_2^T b \\ Q'^T b \end{bmatrix} \right\|_2$$

solution $x_1 = R_{11}^{-1} Q_1^T b - R_{11}^{-1} R_{12} x_2$
for all x_2

How to pick x_2 to minimize $\|x\|_2$?
What can go wrong?

Ex $A = \begin{bmatrix} e & 1 \\ 0 & 0 \end{bmatrix}$ e tiny $R_{11} = e$, $R_{12} = 1$

$x = \begin{bmatrix} (b_1 - x_2)/e \\ x_2 \end{bmatrix}$ e tiny $\Rightarrow$ very sensitive to small changes in x_2, b_1

$A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = AP = \begin{bmatrix} 1 & e \\ 0 & 0 \end{bmatrix} \Rightarrow x = \begin{bmatrix} b_1 - e x_2 \\ x_2 \end{bmatrix}$

insensitive to changes in b, x_2

What would a "perfect" R factor look like?

Compare $\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$ to $\begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}$

Def: Rank Revealing QR factorization (RRQR) for short is $A \cdot P = QR$
where:
R_{11} approximates Σ_1
R_{22} " Σ_2
$\|R_{11}^{-1} R_{12}\|$ small