Jan 29, 2019

Last time:

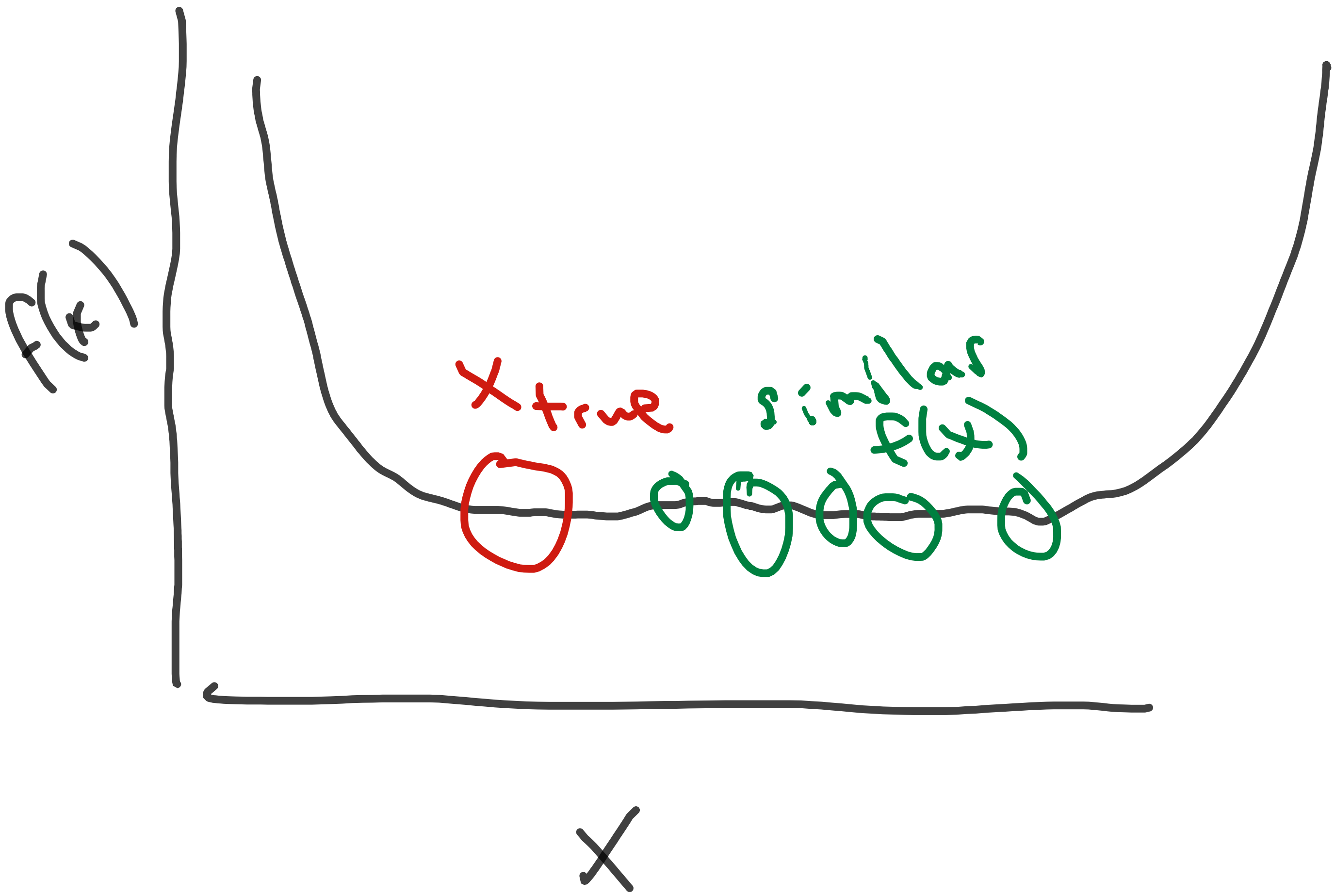
LLS with

- QR
- Cholesky
- SVD

Near-linear correlation in features $\Rightarrow$ ill-conditioning

$\Rightarrow$ many near-optimal solns

Today: how do we pick a solution?



Impose structure

$\Rightarrow$ some bias

better conditioning

Idea: Discourage large solutions

ML: ℓ_2-reg. LLS

Stat: Ridge regression

NA: Tikhonov reg.

Objective

$$\min_x \|Ax - b\|^2 + \lambda^2 \|x\|^2$$

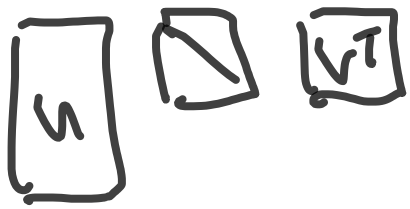
$$\Leftrightarrow \min_x \left\| \begin{bmatrix} A \\ \lambda I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2$$

Normal eqns:

$$(A^T A + \lambda^2 I)\hat{x} = A^T b$$

SVD view: $A = U \Sigma V^T$

$U^T U = V^T V = I$



$V V^T = V^T V = I$

Normal eqns:

$$A^T A + \lambda^2 I = V \Sigma^2 V^T + \lambda^2 I$$
$$= V(\Sigma^2 + \lambda^2 I) V^T$$

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

$$\Leftrightarrow \quad \hat{x} = V \underbrace{(\Sigma^2 + \lambda^2 I)^{-1} \Sigma}_{\begin{pmatrix} f^{-1}(\sigma_1) & & \\ & \ddots & \\ & & f^{-1}(\sigma_n) \end{pmatrix}} U^T b$$

$$f(\sigma)^{-1} = \frac{\sigma}{\sigma^2 + \lambda^2}$$

Truncated SVD

$$A = U \Sigma V^T \qquad \text{SVD}$$

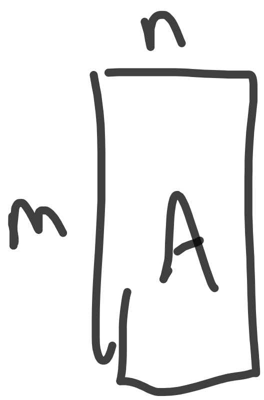
$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}$$

ill-conditioned if $\sigma_1 >> \sigma_j$ for $j > i$

(informal)

just throw out small singular vals

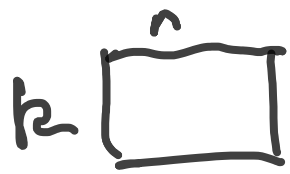
$$A \approx \underset{n \times k}{U_k} \, \underset{k \times k}{\Sigma_k} \, \underset{k \times n}{V_k^T} = A_k$$



Normal equations:

$$\min_{x} \|A_k x - b\|^2$$

$$V_k^T \hat{x} = \Sigma_k^{-1} U_k^T b$$



$$V_k V_k^T \hat{x} = \underbrace{V_k \Sigma_k^{-1} U_k^T b}_{\bar{x}}$$

$$V_k V_k^T \bar{x} = \underbrace{V_k \overbrace{\cancel{V_k^T V_k}}^{I} \Sigma_k^{-1} U_k^T b}_{\bar{x}}$$

$$\boxed{\hat{x} \triangleq V_k \Sigma_k^{-1} U_k^T b}$$

Equivalently:

$$\hat{x} = V f(\Sigma) U^T b$$

$$f(\Sigma) = \begin{pmatrix} f(\sigma_1) & & \\ & \ddots & \\ & & f(\sigma_n) \end{pmatrix}$$

$$f(\sigma) = \begin{cases} \sigma & \sigma > \varepsilon \\ 0 & \text{o/w} \end{cases}$$

with A_k $\varepsilon = \sigma_{k+1}$

Idea # 2: sparse solutions

$$\min_x \|Ax-b\|^2 + \lambda \|x\|_1$$

$$\|x\|_1 = \sum_i |x_i|$$

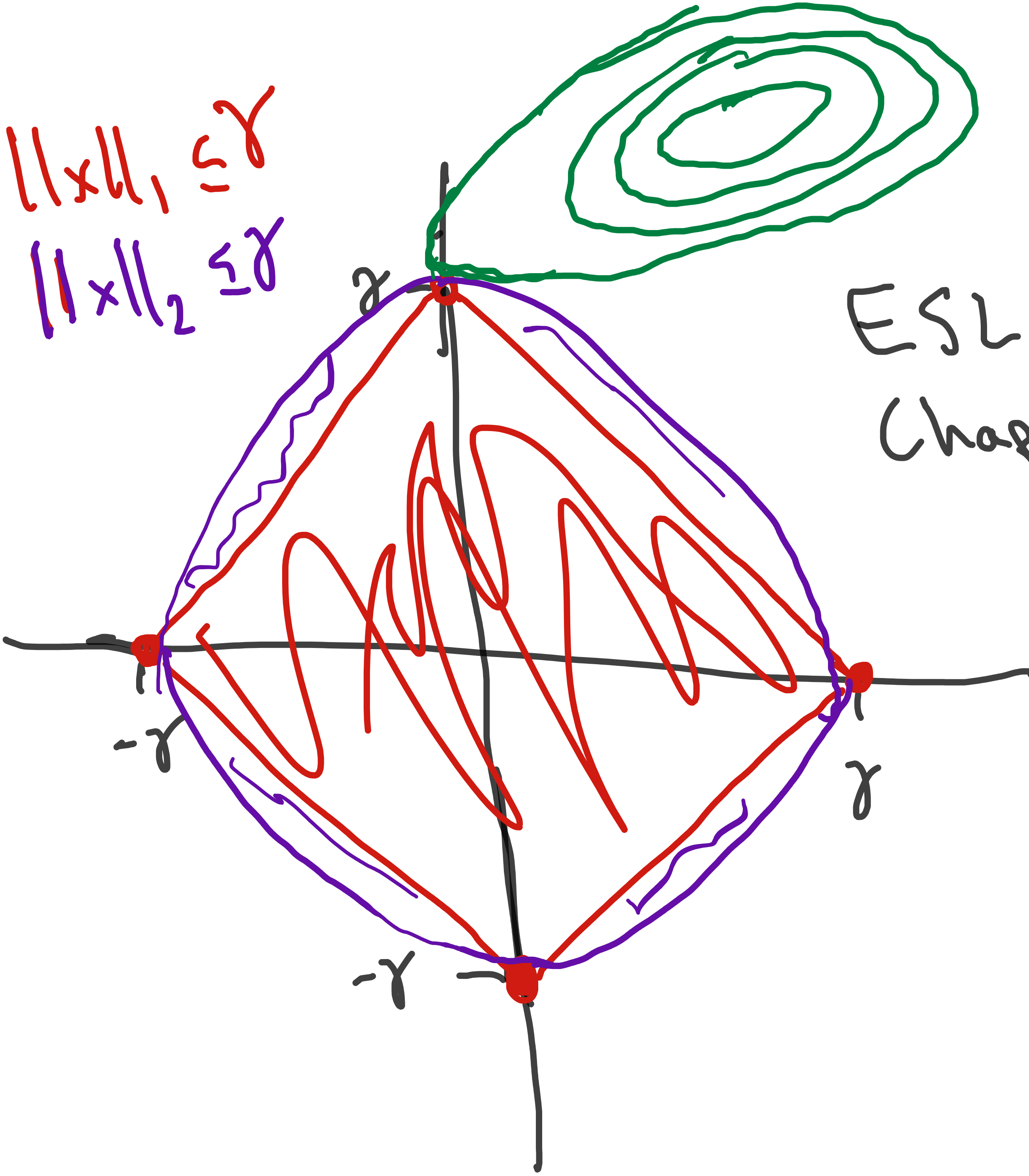
ML: ℓ_1-reg. LLS

Stat: Lasso

NA: basis pursuit denoising

$$\min_x \|Ax-b\|^2$$

$$\text{s.t.} \quad \|x\|_1 \le \gamma$$



ESL Chap. 3

Can we solve?

$$\min_x \underbrace{\|Ax - b\|^2}_{\text{convex}} + \underbrace{\lambda \|x\|_1}_{\text{convex}}$$

smooth

discont. first deriv.

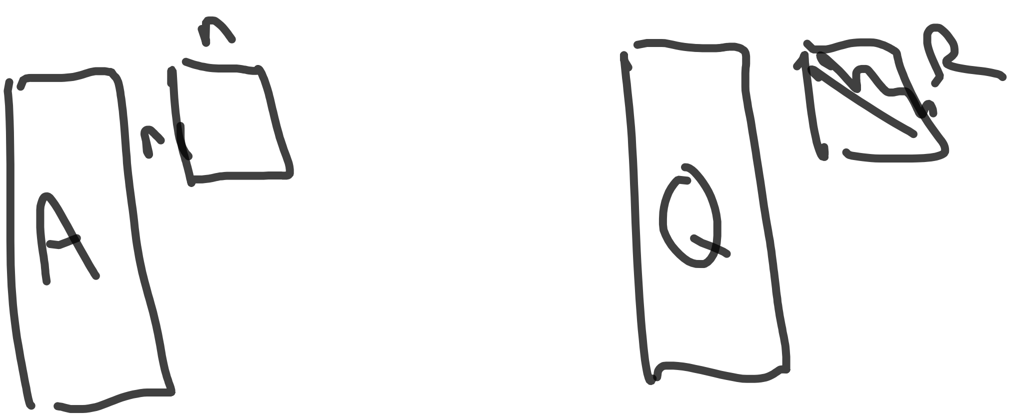
Matrix computations for sparse solutions

pivoted QR

permute columns

pivots

$$A\Pi = QR$$



Alg: RRQR Chan

1. pick col of A w/ largest 2-norm
2. pivot the col to front
3. Orthogonalize out

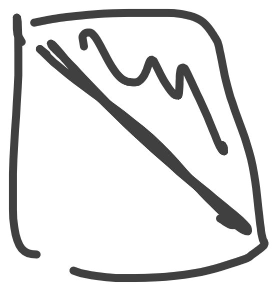
$$Q[:,1] = A[:,1] / \|A[:,1]\|$$

$$A[:,2] = (I - q_1 q_1^T)\, A[:,2:n]$$

$Q[:,1]$ update R

4. Repeat on $A[:,2:n]$

Similar to trunc. SVD



R

At some point $R_{k+1,k+1} < \varepsilon$

$$A \approx \underset{m \times k}{Q_k} \; \underset{k \times n}{R_k}$$

LLS: soln:

1. Solve LLS with the k columns we picked out

$$\min_{x} \| A_k \tilde{x} - b \|^2$$

$\rightarrow$ k-dimensional solution

but $\hat{x} \in \mathbb{R}^n$

$\rightarrow$

2. set other indices in $\hat{x}$ to 0