Jan 28, 2020

$$\hat{x} = \arg\min_x \|Ax-b\|_2^2 \qquad \hat{x} = \underbrace{(A^TA)^{-1}A^T}_{A^+}b \qquad r = A\hat{x}-b, \quad A^Tr = 0$$

---

true data: (A, b) A full rank, model choice: linear

best model: $x = \arg\min_y \|Ay-b\|_2^2$ $\quad Ax = b+r \quad A^Tr=0$

$$A = \begin{bmatrix} A_{tr} \\ A_{te} \end{bmatrix} \quad b = \begin{bmatrix} b_{tr} \\ b_{te} \end{bmatrix} \quad (A_{tr}, b_{tr}+e)$$

Model fit: $\hat{x} = \arg\min_y \|A_{tr}y - (b_{tr}+e)\|_2^2$

① $\hat{x} = A_{tr}^+(b_{tr}+e)$

② $x = A_{tr}^+(b_{tr}+r_{tr})$

$$\boxed{A_{tr}}\,\boxed{x} = \boxed{}\; b_{tr}+r_{tr}$$

$$\|A(\hat{x}-x)\|_2^2 = ?$$

$$A(\hat{x} - x) + r = A\hat{x} - Ax + r = A\hat{x} - b \quad \left(\begin{matrix} Ax = \\ b + r \end{matrix}\right)$$

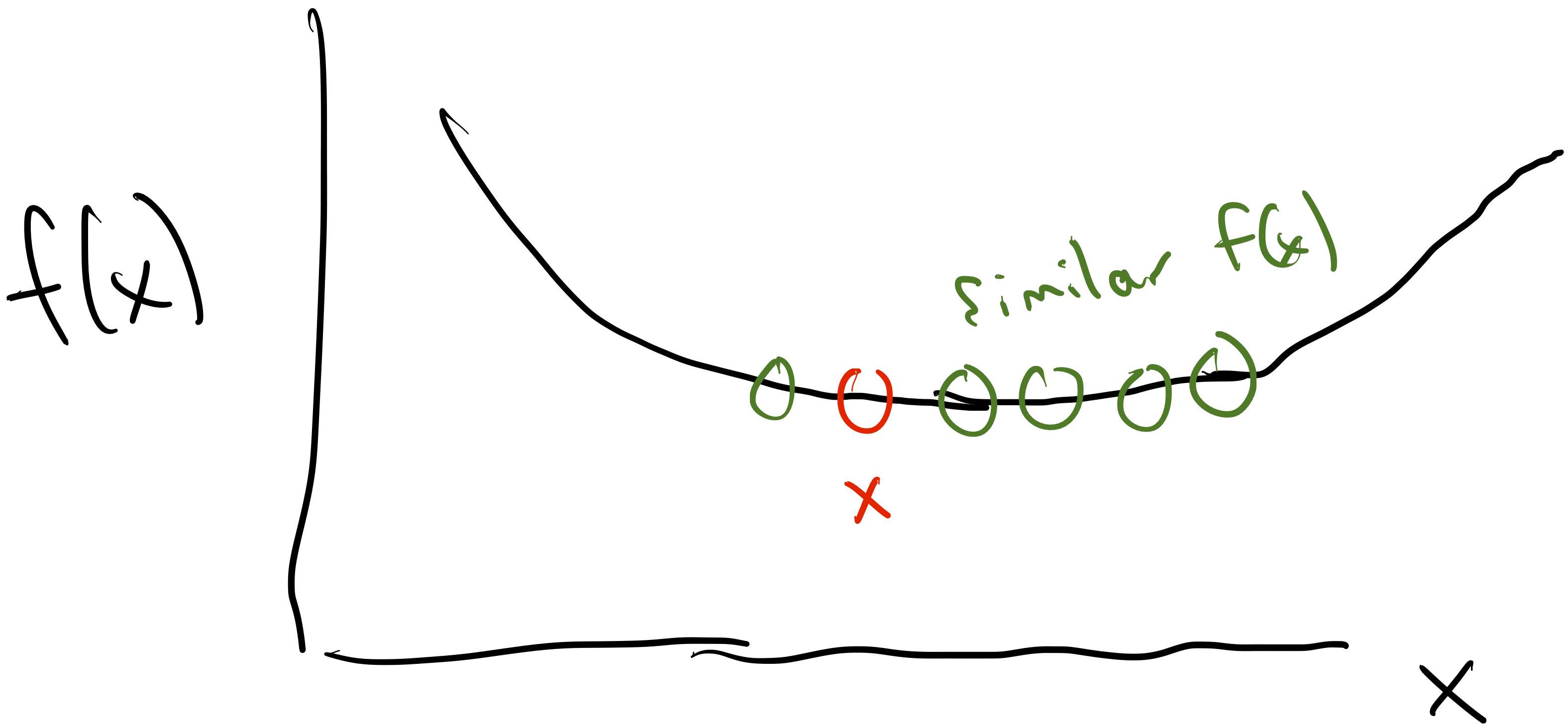
$$\|A\hat{x} - b\|_2^2 = \|A(\hat{x} - x) + r\|_2^2 = \underbrace{\|A(\hat{x} - x)\|_2^2}_{\text{variance}} + \|r\|_2^2 \quad \text{bias}$$

$$r^T A = 0$$

$$\|A(\hat{x} - x)\| = \|AA_{tr}^{+}(\cancel{b}_{tr} + e - (\cancel{b}_{tr} + r_{tr}))\|$$

$$\leq \underbrace{\|A\| \, \|A_{tr}^{+}\|}_{\text{conditioning}} (\|r_{tr}\| + \|e\|)$$

$$\|\hat{x} - x\| \leq \|A_{tr}^{+}\| (\|r_{tr}\| + \|e\|)$$



impose structure on which solution we choose

- some bias
- lower var by better conditioning

Idea 1: encourage "small" solutions

$$\min_x \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2$$

$$\min_x \left\| \begin{bmatrix} A \\ \lambda I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

$$(A^T A + \lambda^2 I)\hat{x} = \begin{bmatrix} A^T & \lambda I \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix} = A^T b$$

SVD view

$$\boxed{A} = \boxed{U}\ \overset{\Sigma}{\boxed{\diagdown}}\ \boxed{V^T} \qquad U^T U = V^T V = V V^T = I$$

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

$$= V \Sigma^2 V^T + \lambda^2 I$$

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

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

$$\hat{x} = V \underbrace{(\Sigma^2 + \lambda^2 I)^{-1} \Sigma}_{\Sigma^{-1} \text{ last time}} U^T b$$

$$\begin{pmatrix} \frac{\sigma_1}{\sigma_1^2 + \lambda^2} & & \\ & \ddots & \\ & & \frac{\sigma_n}{\sigma_n^2 + \lambda^2} \end{pmatrix}$$

$$\hat{x} = A^+ b \qquad A^+ = (A^TA)^{-1}A^T = V\Sigma^{-1}U^T$$

$$\|B\|_2 = \sigma_{max}(B) \qquad \|A^+\|_2 = \frac{1}{\sigma_{min}(A)}$$

"throw out" small σ

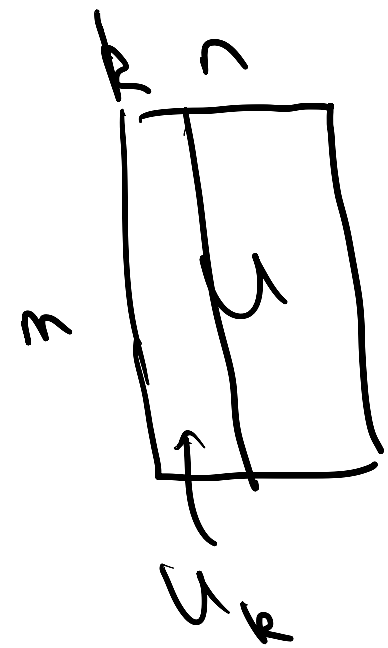
$$f(\Sigma)^{-1} = \begin{pmatrix} 1/\sigma_1 & & & & & \\ & \ddots & & & & \\ & & 1/\sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \qquad f(\sigma)^{-1} = \begin{cases} \frac{1}{\sigma} & \sigma > \varepsilon \\ 0 & o/w \end{cases}$$

$$\bar{x} = V f(\Sigma)^{-1} U^T b$$

$$= \left(\sum_{i=1}^{n} f(\sigma_i)^{-1} v_i u_i^T\right) b$$

$$= \left(\sum_{i=1}^{k} \frac{1}{\sigma_i} v_i u_i^T\right) b$$

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



$$\underset{m}{\overset{n}{A}} \approx \underbrace{\underset{m}{\overset{k}{U_k}}\; \overset{k}{\Sigma_k}\; \overset{n}{V_k^T}}_{\text{truncated SVD}} = A_k$$

Fact: $A_k = \underset{\substack{B \\ \operatorname{rank}(B)=k}}{\arg\min} \|A-B\|_F$

$V_k^T V_k = I$

$\Sigma_k^{-2} V_k^T \qquad\qquad \Sigma_k^{-1} V_k^T$

$$\min_x \|A_k x - b\|_2^2 \qquad V_k \Sigma_k^2 V_k^T \hat{x} = V_k \Sigma_k U_k^T b$$

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

$$\hat{x} = \bar{x} + y, \quad y \in N(V_k^T)$$

$b^T U_k \Sigma_k^{-1} V_k^T y$

$$\|\hat{x}\|^2 = \|\bar{x} + y\|^2 = \|\bar{x}\|^2 + \|y\|^2 + 2\bar{x}^T y$$

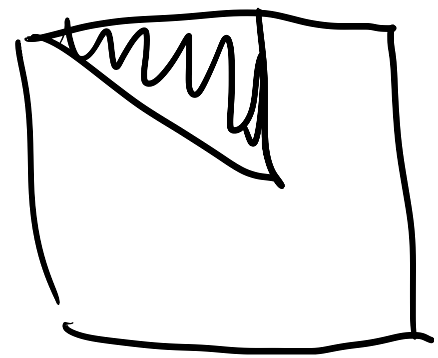
Idea 2: encourage sparse solutions

Problem: some cols are nearly linearly dep.

Idea: choose subset of cols that are most "indep"

Alg: (greedy) pivoted QR

$$A \, \Pi = Q \, R$$



$$\bar{A}_k = A[:, \Pi(1:k)]$$

$$\hat{x}_k = \underset{x}{\operatorname{argmin}} \|\bar{A}_k x - b\|$$

1. pick col of A w/ largest 2-norm Chan 87
2. "pivot" col to front
3. $Q[:,1] = A[:,1] / \|A[:,1]\|_2$

   $A[:,2:n] = (I - Q[:,1]Q[:,1]^T)\, A[:,2:n]$

   (update R)
4. Repeat on $A[:,2:n]$ until

   $R_{kk} < \varepsilon$

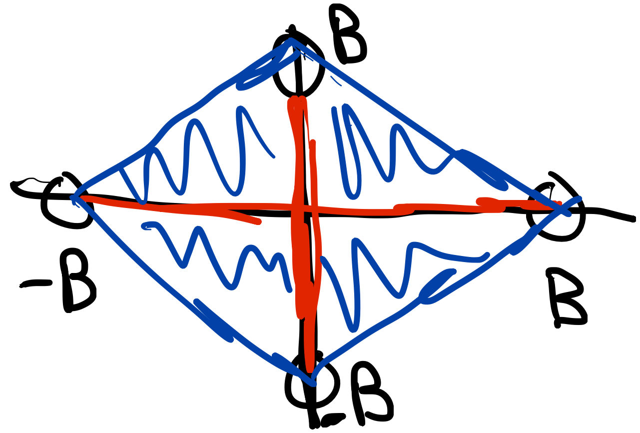
$$\min \|Ax - b\|_2^2$$
$$\text{s.t.} \quad nnz(x) \le k$$

Proposed solution: $nnz(x) \Rightarrow \|x\|_1 = \sum |x_i|$

$\|x\|_1$ is convex

Why? $k = 1$

$$\min \|Ax-b\|$$
$$\text{s.t.} \quad nnz(x) \le 1 \qquad \|x\|_\infty \le B$$



$\|x\|_1 \le B$

"l_1 heuristic"

$$\min \|Ax-b\|_2^2 + \lambda \|x\|_1$$

LASSO

LARS