Oct 12, 2020

data points (a_i^T, b_i) $i=1,\dots,m$

model choice: linear

best model: $\hat{x} = \arg\min_x \|Ax-b\|_2^2$

$\hat{x} = A^+ b$, $A\hat{x} = b + r$ $\quad A^T r = 0$

$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ $\quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ have $(A_1, b_1 + \delta b_1)$

$\hat{x}_1 = A_1^+ (b_1 + \delta b_1)$

How good is $\hat{x}_1$?

$A\hat{x}_1 - b$

$= A\hat{x}_1 - A\hat{x} + r$ $\quad (A\hat{x} = b + r)$

$= A(\hat{x}_1 - \hat{x}) + r$

$\|A(\hat{x}_1 - \hat{x}) + r\|_2^2$ $\quad A^T r = 0$

$= \boxed{\|A(\hat{x}_1 - \hat{x})\|_2^2} + \boxed{\|r\|_2^2} + \cancel{2(\hat{x}_1 - \hat{x})^T A^T r}$

bias from linear model

variance from sensitivity to data perturbations

$\hat{x} = A^+ b$ $\quad A\hat{x} = b + r$

Claim: $\hat{x} = A_1^+ (b_1 + r_1)$

Proof: $A_1 \hat{x} = b_1 + r_1$.

$A_1^+ (b_1 + r_1) = \arg\min_y \|A_1 y - (b_1 + r_1)\|_2^2$

$\|A(\hat{x}_1 - \hat{x})\|_2$

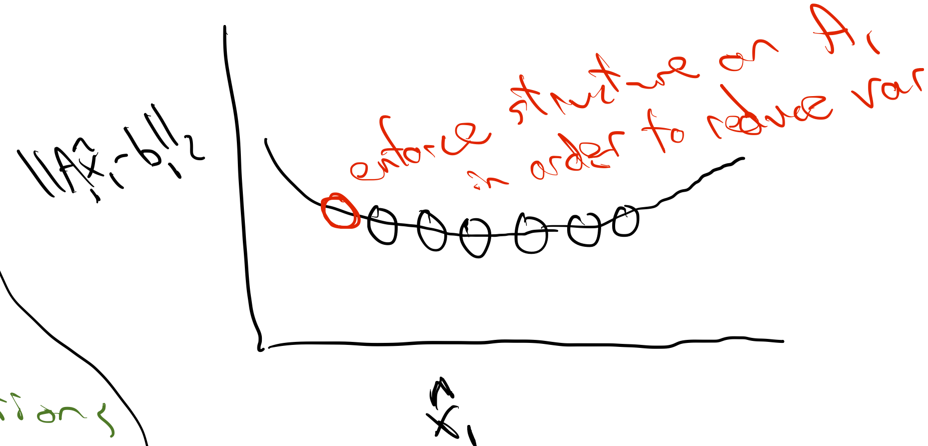
$\le \|A\|_2 \|A_1^+ (\cancel{b_1} + \delta b_1) - A_1^+ (\cancel{b_1} + r_1)\|_2$

$\le \|A\|_2 \|A_1^+\|_2 [\|\delta b_1\|_2 + \|r_1\|_2]$

$= \sigma_1(A) / \sigma_n(A_1) \, [\|\delta b_1\|_2 + \|r_1\|_2]$

can lower variance

cost: increase bias $\|r_1\|_2$ $\|r\|_2$



Last time: truncated SVD

$$\hat{x} = A_\tau^+ b = V_k \Sigma_k^{-1} U_k^T b \qquad k = \arg\max_i \; \sigma_i > \tau$$

Structure: solution $\hat{x}$ living in leading k-dim subspace $\Rightarrow \sigma_1/\sigma_k$

Explicitly encouraging small solutions

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

Normal equations:

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

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

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

$$\Sigma_\lambda^{-1} = \begin{pmatrix} \sigma_1/\sigma_1^2+\lambda^2 & & \\ & \ddots & \\ & & \sigma_n/\sigma_n^2+\lambda^2 \end{pmatrix} \Rightarrow \sqrt{\frac{\sigma_1^2+\lambda^2}{\sigma_n^2+\lambda^2}}$$

Another idea: sparse solutions

Goal: pick k columns that are "most independent"

Pivoted QR: $A\Pi = QR$ (MGS, unstable)

1. choose col t with largest 2-norm
2. pivot t to front
3. $r_{11} = \|A(:,1)\|_2$ $\quad q_1 = A(:,1)/r_{11}$
4. $A(:,2:n) = (I - q_1 q_1^T)\,A(:,2:n)$ and repeat

$$\|A\Pi e_2\|_2^2 = \|q_1 r_{12} + q_2 r_{22}\|_2^2 = |r_{12}|^2 + |r_{22}|^2 \le |r_{11}|^2 = \|A\Pi e_1\|_2^2$$

$$|r_{22}|,\ |r_{12}| \le |r_{11}|$$

$$|r_{j+1,\,j+1}| \le |r_{jj}|$$

Exactly rank r

$$A\Pi = (Q_1\ \ Q_2)\ \begin{array}{c|cc} & r & n-r \\ \hline r & R_1 & R_2 \\ n-r & 0 & 0 \end{array}$$

first r cols: $Q_1 R_1$

$Q_1 R_2$

Nearly low rank

$$A\Pi = (Q_1\ \ Q_2)\begin{pmatrix} R_1 & R_2 \\ 0 & E \end{pmatrix}$$

E is small

$\|E\|$ is small

$$\begin{aligned} A\Pi &\approx Q_1[R_1\ \ R_2] \\ &= Q_1 R_1 [I\ \ R_1^{-1} R_2] \\ &= \underbrace{(A\Pi)(:,1:k)[I\ \ T]} \end{aligned}$$

Result of PQR: r columns of A that approx. range(A)

$$\min_{x \in \mathbb{R}^r} \|A_r x - b\|_2^2$$

L_1-regularization / Lasso

$$\min \|Ax - b\|_2^2 \quad \text{s.t. } \operatorname{nnz}(x) \le r \quad \Big\}\ \text{NP-hard}$$

$$\min \|Ax-b\|_2 \quad \text{s.t. } \|x\|_1 \le B$$

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

$$\operatorname{nnz}(x) \Rightarrow \|x\|_1$$

$$r = 1 \quad \operatorname{nnz}(x) \le 1 \iff \|x\|_\infty \le B$$

$\|x\|_1 \le B$ = convex hull $\{x \mid \|x\|_\infty \le B\}$