Oct 12, 2020

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

model choice: linear

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

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

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

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

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

$A\hat{x}_1 - b$

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

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

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

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

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

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

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

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

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

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

can lower variance

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