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 $x = A_{tr}^+(b_{tr}+r_{tr})$

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

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

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

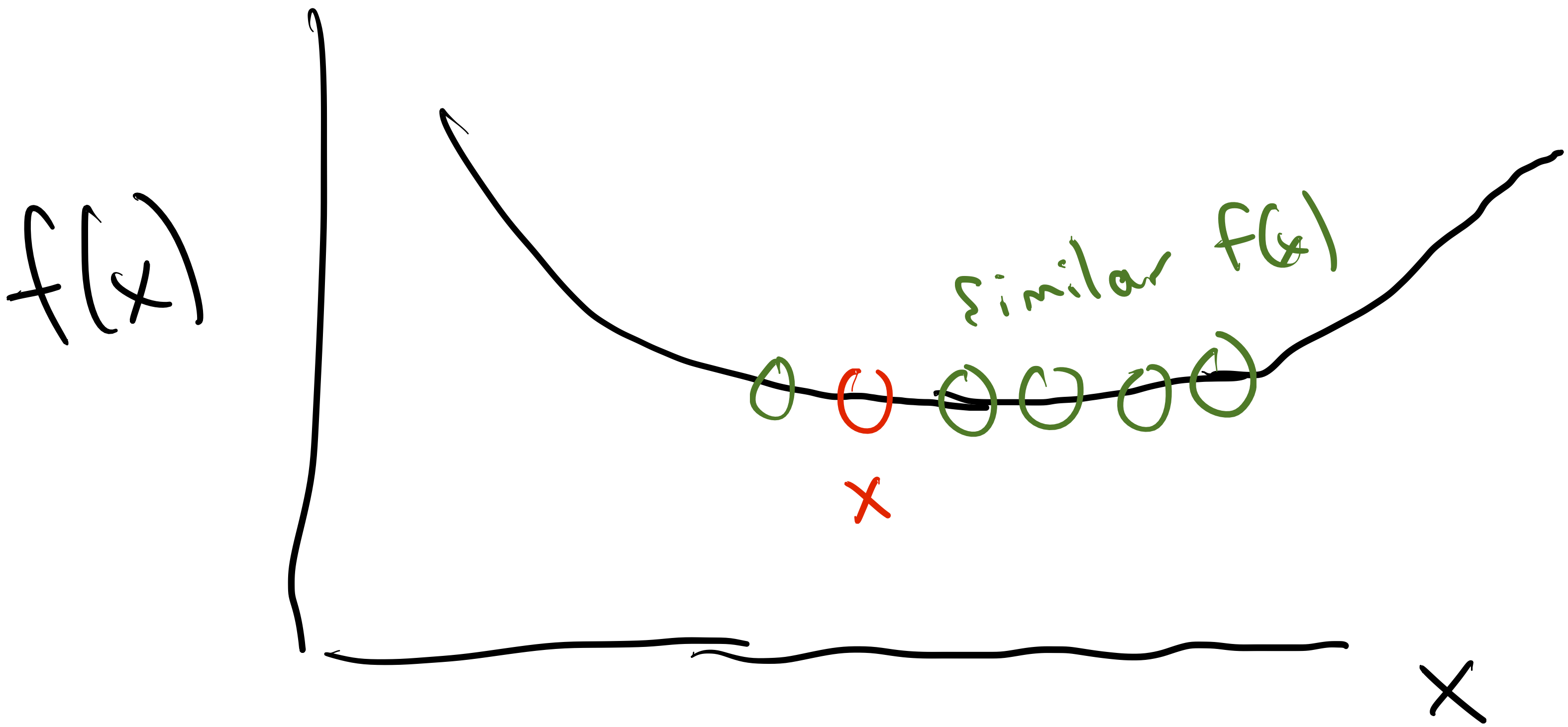
$$\|A\hat{x}-b\|_2^2 = \|A(\hat{x}-x)+r\|_2^2 = \underbrace{\|A(\hat{x}-x)\|_2^2}_{\text{variance}} + \underbrace{\|r\|_2^2}_{\text{bias}}$$

$$r^T A = 0$$

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

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

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



impose structure on which solution we choose

- some bias
- lower var by better conditioning

encourage "small" solutions

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

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