$Ax = b$, $\quad b \notin \text{colspan}(A)$. $\quad A \in \mathbb{R}^{m \times n}$

$r = b - Ax \in \mathbb{R}^m$

$= \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix}$

$\|r\|_2 = \sqrt{r^T r}$

$\|r\|_2^2 = r^T r = r_1^2 + r_2^2 + \dots + r_m^2$

$$\min_{x \in \mathbb{R}^n} r^T r$$

$$A = \begin{bmatrix} \\ - \; a_i \; - \\ \\ \end{bmatrix}_{m \times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_i \\ \\ \\ \end{bmatrix}$$

$r_i = b_i - \overbrace{a_i x}^{\hat{y}_i} \doteq$

where $x_i = \underbrace{(\_\_\_\_\_\_)}_{n\text{-length}}$

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

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \\ 0 \end{bmatrix} \qquad Ax - b = \begin{bmatrix} 2x_1 + x_2 - 1 \\ x_1 + 2x_2 \\ -1 \end{bmatrix}$$

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \min_{x_1, x_2} (2x_1 + x_2 - 1)^2 + (x_1 + 2x_2)^2 + 1$$

$$\frac{\partial}{\partial x_1} \|Ax - b\|_2^2 = 0 \quad \& \quad \frac{\partial}{\partial x_2} \|Ax - b\|_2^2 = 0$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$$

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \min_{x \in \mathbb{R}^n} f(x)$$

$\hat{x} \in \mathbb{R}^n$ which minimizes $f(x)$ satisfies

$$\nabla_x f(\hat{x}) = 0$$

$$f(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)^2$$

Compute $\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \left(\sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)^2 \right) = 2 \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right) A_{ik}$$

$$= 2 \sum_{i=1}^{m} (A^T)_{ki} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)$$

$$= 2 \left[A^T (Ax - b) \right]_k = 2 \sum_{i=1}^{m} A^T_{ki} (Ax - b)_i$$

$$\Rightarrow \nabla_x f = 2A^T(Ax-b)$$

For a minimizer $\hat{x}$ of $f(x)$, we will have,

$$\nabla_x f(\hat{x}) = 2A^T(A\hat{x}-b) = 0$$

$$\Rightarrow A^TA\hat{x} = A^Tb \leftarrow \text{normal eq}^{ns}$$

Under the assumption that A has lin. indep. columns, (A^TA) is invertible.

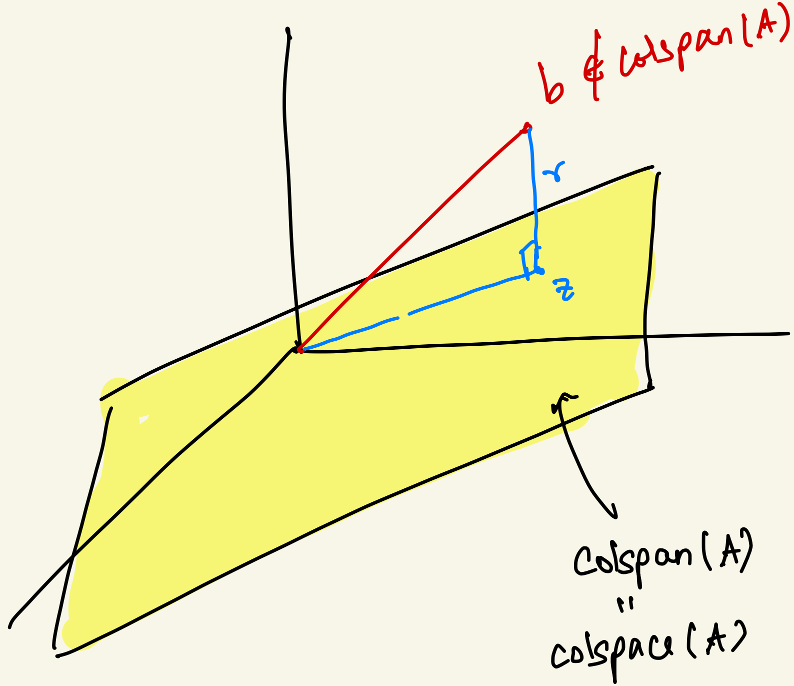
$$\Rightarrow \hat{x} = (A^TA)^{-1}A^Tb$$ (LS solution is unique).

Assume columns of A are linearly indep.

WTP A^TA is invertible

$x \in \mathbb{R}^n$ s.t. $A^TAx = 0$

$$0 = x^T(A^TAx)$$
$$= (x^TA^T)(Ax)$$
$$= (Ax)^T(Ax)$$
$$= \|Ax\|_2^2$$
$$\Rightarrow \|Ax\|_2 = 0$$
$$\Leftrightarrow Ax = 0$$
$$\Leftrightarrow x = 0 \quad \because A \text{ has lin. indep. columns.}$$



Find a vector

$$z \in \text{colspan}(A)$$

s.t. z is the closet vector to b.

normality condition

$$\iff r = A\hat{x} - b \perp z \quad \forall z \in \text{colspan}(A)$$

$$\iff z^T r = 0 \quad \forall z \in \text{colspan}(A)$$

$$\iff (Ay)^T (A\hat{x} - b) = 0 \quad \forall y \in \mathbb{R}^n$$

$$\iff y^T A^T (A\hat{x} - b) = 0 \quad \forall y \in \mathbb{R}^n$$

$$\iff A^T (A\hat{x} - b) = 0$$

$$\iff A^T A \hat{x} = A^T b$$

fixed

$$\overbrace{(y_1 \;\; y_2 \; \cdots \; y_n)}^{\text{arbitrary}} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = 0$$

LS data fitting.

$\mathcal{D} = \{ (x_i, y_i)_{i=1}^{N} \}$ given data set.

input output

$y = f(x)$ ← is the functional relationship. (unknown)

Objective: To guess/estimate/find this relationship f.

Soln: fix p basis functions $f_1, f_2, \ldots, f_p$ such that

$$\hat{f} = \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_p f_p$$

$$r_i = y_i - \hat{f}(x_i)$$

$$\min_{\alpha_1, \alpha_2, \ldots \alpha_p} \sum_{i=1}^{N} r_i^2 = \min_{\alpha_1, \ldots, \alpha_p} \sum_{i=1}^{N} (y_i - \hat{f}(x_i))^2$$

$$\hat{y}_i = \hat{f}(x_i) = \alpha_1 f_1(x_i) + \alpha_2 f_2(x_i) + \cdots + \alpha_p f_p(x_i) \quad \forall i = 1, 2, \ldots, N$$

$$\hat{y}_i = \begin{bmatrix} f_1(x_i) & f_2(x_i) \cdots & f_p(x_i) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \quad \forall i = 1, 2, \ldots, N$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N \quad , \quad \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{bmatrix} \in \mathbb{R}^N$$

$$\underset{N \times 1}{\hat{y}} = \overbrace{\begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \vdots & & & \\ f_1(x_i) & f_2(x_i) & \cdots & f_p(x_i) \\ \vdots & & & \\ f_1(x_N) & f_2(x_N) & \cdots & f_p(x_N) \end{bmatrix}_{N \times p}}^{\Phi} \overbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}_{p \times 1}}^{\alpha}$$

$$r = y - \hat{y} \in \mathbb{R}^N$$

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ \vdots \\ y_N - \hat{y}_N \end{pmatrix}$$

$$\min_{\alpha_1, \alpha_2, \ldots, \alpha_p} \sum_{i=1}^{N} r_i^2 = \min_{\alpha_1, \ldots, \alpha_p} r^T r = \min_{\alpha_1, \ldots, \alpha_p} \|r\|_2^2 = \min_{\alpha_1, \ldots, \alpha_p} \|y - \hat{y}\|_2^2$$

$$= \min_{\alpha \in \mathbb{R}^p} \|y - \Phi \alpha\|_2^2$$

Since columns of Φ are linearly independent, the least squares solution $\hat{\alpha} \in \mathbb{R}^N$ is given by

$$\hat{\alpha} = (\Phi^T \Phi)^{-1} \Phi^T y$$