Least Squares

Sometimes no solution to

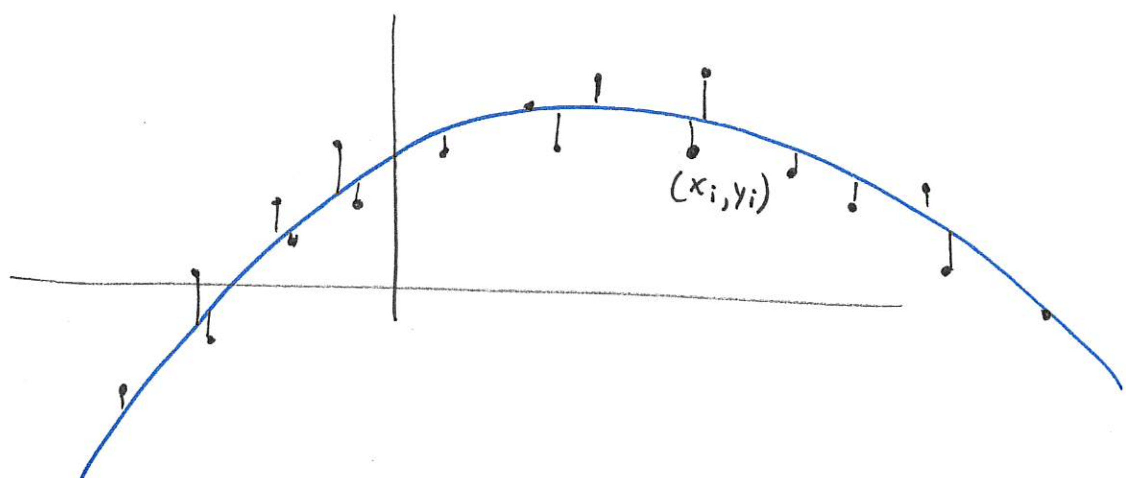
$$Ax = b$$

We want to find the "best" solution $\hat{x}$.

Least squares choose $\hat{x}$ so as to minimize the norm of the residual, $r = b - A\hat{x}$, i.e.,

$$\min_{\hat{x}} \| b - A\hat{x} \|_2^2$$

Example: Overdetermined system from noisy data



We collect data (x_i, y_i), $i = 1, \dots, m$. Fit the data with a parabola

$$y = ax^2 + bx + c = f(x)$$

Such that

$$\sum_{i=1}^{m} \left(f(x_i) - y_i\right)^2 = \min$$

The unknowns are a, b, and c. Each data point gives an equation

$$a x_i^2 + b x_i + c = y_i \quad , \quad i = 1, \dots, m$$

Combine into a matrix equation

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ & \vdots & \\ x_m^2 & x_m & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

$$\underset{m \times 3}{A} \underset{3 \times 1}{\vec{x}} = \underset{m \times 1}{\vec{b}}$$

$$m > 3$$

As illustrated in plot above, this system has no solution. Seek to minimize $\| \vec{b} - A\vec{x} \|_2^2$ "least squares"

How to find the $\hat{x}$ that minimizes $\| \vec{b} - A\hat{x} \|_2^2$?

$$\begin{aligned} \| b - A\hat{x} \|_2^2 &= (b - A\hat{x})^T (b - A\hat{x}) \\ &= b^T b - b^T A \hat{x} - \hat{x}^T A^T b + \hat{x}^T A^T A \hat{x} \quad = \min \\ &= \phi(\hat{x}) \end{aligned}$$

to find min, find $\hat{x}$ s.t. $\frac{\partial \phi}{\partial x}(\hat{x}) = 0$

Normal Equations (derivation by calculus)

$$\phi(x) = x^T A^T A x - 2b^T A x + b^T b$$

Differentiate to find $\hat{x}$ s.t. $\nabla\phi(\hat{x}) = 0$

Two ways

1. vector

$$\delta\phi = \delta x^T A^T A x + x^T A^T A \delta x - 2b^T A \delta x$$

$$= 2x^T A^T A \delta x - 2b^T A \delta x$$

$$= (2x^T A^T A - 2b^T A)\delta x$$

$$\Rightarrow \nabla\phi = (2x^T A^T A - 2b^T A)^T$$

$$= 2A^T A x - 2A^T b$$

$$\nabla\phi(\hat{x}) = 0 \Rightarrow 2A^T A \hat{x} = 2A^T b$$

$$\Rightarrow \boxed{A^T A \hat{x} = A^T b}$$

2. scalar

$$\phi(x) = \sum_{i,j} x_i [A^T A]_{ij} x_j - 2\sum_{i,j} b_i a_{ij} x_j + b^T b$$

$$\frac{\partial\phi}{\partial x_k}(x) = \sum_{i,j} \frac{\partial x_i}{\partial x_k}[A^T A]_{ij} x_j + x_i [A^T A]_{ij}\frac{\partial x_j}{\partial x_k} - 2\sum_{i,j} b_i a_{ij}\frac{\partial x_j}{\partial x_k}$$

$$= \sum_{i,j} \delta_{ik}[A^T A]_{ij} x_j + x_i [A^T A]_{ij}\delta_{jk} - 2\sum_{i,j} b_i a_{ij}\delta_{jk}$$

$$= \sum_j [A^T A]_{kj} x_j + \sum_i x_i [A^T A]_{ik} - 2\sum_i b_i a_{ik}$$

2. (continued)

$$\frac{\partial \phi}{\partial x_k}(x) = \sum_j [A^T A]_{kj} x_j + \sum_i x_i [A^T A]_{ik} - 2 \sum_i b_i a_{ik}$$

$$\nabla \phi(x) = A^T A x + A^T A x - 2 A^T b$$

$$= 2 A^T A x - 2 A^T b$$

$$\nabla \phi(\hat{x}) = 0 \iff \boxed{A^T A \hat{x} = A^T b}$$

Normal Equations

The $\hat{x}$ that minimizes $\| b - A\hat{x} \|_2^2$ satisfies the normal equations

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

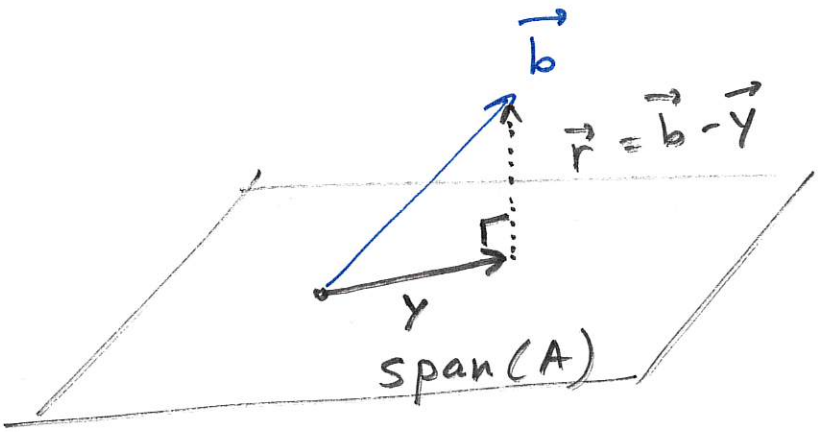
Normal Equations (geometric intuition)

We want to find $\hat{x}$ so that

$\| A\hat{x} - b \|_2$ is as small as possible.

Here is a picture depicting the problem

(span(A) = range(A) = column space(A))



Notice that $\vec{b} \notin$ span(A). If it was, there would be a solution to $Ax = b$. We want to find $\vec{y} \in$ span(A) as close to $\vec{b}$ as possible

The "residual" is

$$\vec{r} = \vec{b} - \vec{y} = \vec{b} - A\hat{x} \qquad (\vec{y} = A\hat{x})$$

and we want $\|\vec{r}\|_2 = \min$. This is achieved when

$$\vec{r} \perp \vec{A}$$

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

$$\boxed{A^T A x = A^T b}$$

Notice that $\vec{y}$ is the orthogonal projection of $\vec{b}$ onto span(A)

$$y = AA^{+} b$$

If $A^T A$ is invertible,

$$y = A(A^T A)^{-1} A^T b$$

$\vec{y} \in \text{span}(A)$

closest vector to $\vec{b}$ in span(A)

$\vec{y}$ is unique, but

$$A\hat{x} = \vec{y}$$

$\hat{x}$ may not be unique

$\hat{x}$ is unique iff A has indep. columns

Least Squares Sol'n by pseudo-inverse.

$$\min_{x} \| \underset{m\times 1}{b} - Ax \|_2$$

Let $\underset{m\times n}{A} = \underset{m\times m}{U}\underset{m\times n}{\Sigma}\underset{n\times n}{V^T}$, with rank r

$$\|b - Ax\|_2^2 = \|b - U\Sigma V^T x\|_2^2$$

$$= \|U^T b - \Sigma V^T x\|_2^2$$

$$= \left\| \begin{pmatrix} u_1^T b \\ \vdots \\ u_r^T b \\ u_{r+1}^T b \\ \vdots \\ u_m^T b \end{pmatrix} - \begin{pmatrix} \sigma_1 v_1^T x \\ \vdots \\ \sigma_r v_r^T x \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} u_1^T b - \sigma_1 v_1^T x \\ \vdots \\ u_r^T b - \sigma_r v_r^T x \end{pmatrix} \right\|_2^2 + \left\| \begin{pmatrix} u_{r+1}^T b \\ \vdots \\ u_m^T b \end{pmatrix} \right\|_2^2$$

$$= \sum_{i=1}^{r} \| u_i^T b - \sigma_i v_i^T x \|_2^2 + \sum_{i=r+1}^{m} \| u_i^T b \|_2^2$$

<u>Solution</u>: $u_i^T b = \sigma_i v_i^T x \qquad i = 1, \dots, r$

$$\Rightarrow \quad v_i^T x = \frac{1}{\sigma_i} u_i^T b \qquad i = 1, \dots, r$$

$$x = \sum_{i=1}^{r} \left(\frac{u_i^T b}{\sigma_i} \right) v_i + \sum_{i=r+1}^{n} \alpha_i v_i$$

$x = V\Sigma^+ U^T b$ is the

minimum norm solution to the L.S.

Or, from normal equations:

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

$$V \Sigma^T \cancel{U^T U} \Sigma V^T \hat{x} = V \Sigma^T U^T \vec{b}$$

$$\cancel{V} \Sigma^T \Sigma V^T \hat{x} = \cancel{V} \Sigma^T U^T \vec{b}$$

$$\Rightarrow \quad \sigma_i^2 \vec{v}_i^T \hat{x} = \sigma_i \vec{u}_i^T \vec{b} \quad , \quad i = 1, \dots, r$$

$$\Rightarrow \quad \sigma_i \vec{v}_i^T \hat{x} = \vec{u}_i^T \vec{b} \quad , \quad i = 1, \dots, r$$

$$\vec{v}_i^T \hat{x} = \frac{1}{\sigma_i} \vec{u}_i^T \vec{b} \quad , \quad i = 1, \dots, r$$

$$\hat{x} = \sum_{i=1}^{r} \frac{\vec{u}_i^T \vec{b}}{\sigma_i} \vec{v}_i + \sum_{i=r+1}^{n} \alpha_i \vec{v}_i$$

minimum norm solution is

$$\hat{x} = \sum_{i=1}^{r} \frac{\vec{u}_i^T \vec{b}}{\sigma_i} \vec{v}_i$$