Least Squares - Motivation

Science is replete with curve fitting problems. In linear regression one has data points $\{(x_n, y_n)\}$ we wish to fit to a linear model

$$y = mx + b$$

The question is, what is a good choice of slope m and intercept b. A perfect fit would have

$$\begin{aligned} b + m x_1 &= y_1 \\ b + m x_2 &= y_2 \\ &\vdots \\ b + m x_n &= y_n \end{aligned}$$

which can be written

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

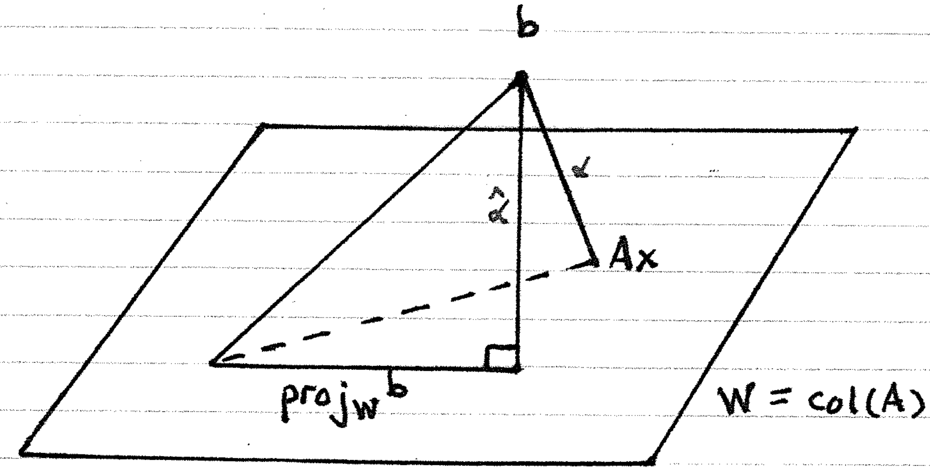
This overdetermined system (2 unknowns, n equations) then becomes a system problem

$$(1) \qquad Ax = b$$

where $b \notin col(A)$. In what is to follow we examine "least squares" solutions of (1).

Least Squares – Geometry

$$\boxed{Ax = b \qquad b \notin \operatorname{col} A}$$



$\alpha > \hat{\alpha}$

Since $b - \text{proj}_W b \in W^\perp$ we expect it to be closest. In other words, if we can find an $\hat{x}$ such that

$$A\hat{x} = \text{proj}_W b$$

then $\hat{x}$ is the "best" soln to $Ax = b$ since then $A\hat{x}$ is closest to b.

We now formalized this.

Best Approximation Theorem

Let W be a subspace of inner product space V. Then

$$\hat{w} = \text{proj}_W b \qquad \hat{w} \in W$$

is the best approximation to b in the sense that

$$\| b - \text{proj}_W b \| < \| b - w \|$$

for all $w \neq \text{proj}_W b$, $w \in W$.

Proof:

$$b - w = \underbrace{(b - \text{proj}_W b)}_{\in W^\perp} + \underbrace{(\text{proj}_W b - w)}_{\in W}$$

Pythagorus Thm applies, hence

$$\| b - w \|^2 = \| b - \text{proj}_W b \|^2 + \| \text{proj}_W b - w \|^2$$

$$\| b - w \|^2 > \| b - \text{proj}_W b \|^2 \qquad \square$$

Pythagorus for general inner product spaces

$$z = x + y \qquad \langle x, y \rangle = 0$$

then

$$\| z \|^2 = \langle x+y, x+y \rangle = \| x \|^2 + 2\cancelto{0}{\langle x, y \rangle} + \| y \|^2$$

$$\| z \|^2 = \| x \|^2 + \| y \|^2$$

Least Squares - Normal Eqns

The least squares solution x of $Ax = b$ must satisfy

$$(1) \qquad Ax = \text{proj}_W b$$

Rather than solve this for x, consider:

$$b - Ax = b - \text{proj}_W b \quad \in \text{col}(A)^{\perp} = N(A^T)^*$$

Hence

$$A^T(b - Ax) = A^T(b - \text{proj}_W b) = 0^*$$

Since the right side vanishes, x is a soln of:

$$(2) \qquad \boxed{A^T A x = A^T b} \qquad \text{normal eqns}$$

To find x, use normal eqns (2).

<u>Theorem</u> Given $Ax = b$ then the normal eqns

$$(3) \qquad A^T A x = A^T b$$

is consistent. Any soln of (3) is a least squares soln of $Ax = b$ and has

$$Ax = \text{proj}_W b$$

EXAMPLE Find the least squares solution of

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 13 \\ -13 \\ 0 \end{bmatrix}}_{b}$$

After some calculations the normal equations

$$A^T A x = A^T b$$

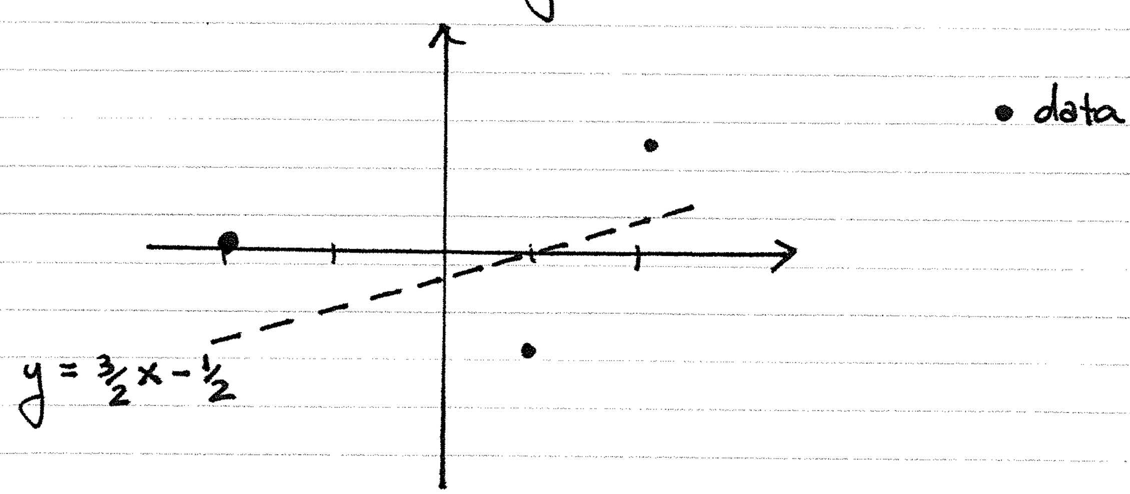
become

$$(1) \qquad \begin{bmatrix} 3 & 1 \\ 1 & 9 \end{bmatrix} x = \begin{bmatrix} 0 \\ 13 \end{bmatrix}$$

In this (typical) example $A^T A$ is invertible. Solving (1)

$$x = \begin{pmatrix} -\frac{1}{2} \\ \frac{3}{2} \end{pmatrix}$$

<u>Relation to linear regression</u> $(x_1, x_2) = (\text{intercept}, \text{slope})$



Least Squares Calculus Derivation

$$F(x) \equiv \|Ax - b\|^2$$

Seek x s.t. $F(x)$ is minimized. Define the residual $r = Ax - b$. In component form

$$r_i = \sum_{j=1}^{n} A_{ij} x_j - b_i$$

hence

$$F(x) = \sum_{i=1}^{m} r_i^2$$

To minimize $F(x)$ we must se

(1) $$\frac{\partial F}{\partial x_k} = \sum_{i=1}^{m} 2 r_i \frac{\partial r_i}{\partial x_k} = 0 \qquad \forall k.$$

From the definition of r_i

(2) $$\frac{\partial r_i}{\partial x_k} = A_{ik}$$

Use (2) in (1) to get

$$\frac{\partial F}{\partial x_k} = \sum_{i=1}^{m} 2\left(\sum_{j=1}^{n} A_{ij} x_j - b_i\right) A_{ik} = 0$$

$$\underbrace{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ik} A_{ij} x_j}_{A^T A x} = \underbrace{\sum_{i=1}^{m} A_{ik} b_i}_{A^T b}$$