# Least Squares

In terms of linear algebra we want to solve an overdetermined system of equations (i.e sets of linear system of equations in which there are more equations than unknowns)

i.e $\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$ where $\underset{\sim}{A}$ having more rows than columns.

The idea of least squares solution is to find $\underset{\sim}{x}$ that minimizes 2-norm of the residual $\underset{\sim}{r} = \underset{\sim}{b} - \underset{\sim}{A}\,\underset{\sim}{x}$

<u>Example</u>:- Let us consider the data

| $x$<br>log(GDP) | $y$<br>% urbanization |
| --- | --- |
| $x_1$ | $y_1$ |
| $x_2$ | $y_2$ |
| $\vdots$ | |
| $x_m$ | $y_m$ |

Say we want to fit a model

$$y = c_1 + c_2 x + c_3 x^2 + \dots c_n x^n$$

$n < m$

It would be very nice to have

$$\left.\begin{aligned} c_1 + c_2 x_1 + c_3 x_1^2 + \dots c_n x_1^n &= y_1 \\ c_1 + c_2 x_2 + c_3 x_2^2 + \dots c_n x_2^n &= y_2 \\ &\vdots \\ c_1 + c_2 x_m + c_3 x_m^2 + \dots c_n x_m^n &= y_m \end{aligned}\right\}$$

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ 1 & x_3 & x_3^2 & \dots & x_3^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \dots & x_m^n \end{bmatrix}}_{\underset{\sim}{A} \quad m \times n} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{bmatrix}}_{\underset{\sim}{x} \; n \times 1} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}}_{\underset{\sim}{b} \; m \times 1}$$

i.e we want to solve

$$\underset{\sim}{A} \underset{\sim}{x} = \underset{\sim}{b}$$

$$\underset{\sim}{A} \in \mathbb{R}^{m \times n} ; \; \underset{\sim}{x} \in \mathbb{R}^{n \times 1} ; \; \underset{\sim}{b} \in \mathbb{R}^{m \times 1}$$

$m > n$

$\underset{\sim}{A}$ is a full rank matrix

In general there is no solution to this problem unless $\underset{\sim}{b} \in \text{range}(\underset{\sim}{A})$

and this will be true for special choices of $\underset{\sim}{b}$

$$\underset{\sim}{A}\underset{\sim}{x} \approx \underset{\sim}{b}$$

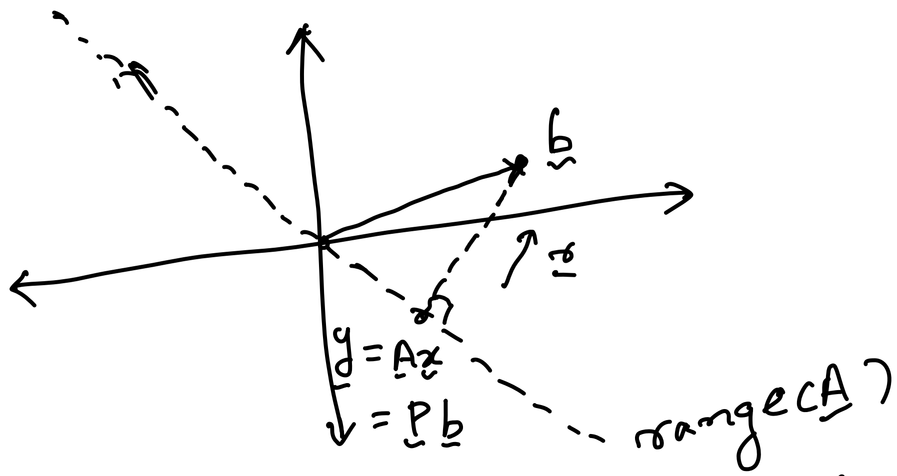
but can $\underset{\sim}{r} = \underset{\sim}{b} - \underset{\sim}{A}\underset{\sim}{x}$ be made smaller? Smallness of $\underset{\sim}{r}$ hints us to use a norm, and if we choose 2-norm, the problem becomes

Given $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$; $m \geq n$, $b \in \mathbb{R}^m$
$\underset{\sim}{A}$ is a full rank matrix, then find
$\underset{\sim}{x} \in \mathbb{R}^n$ such that
$\|\underset{\sim}{b} - \underset{\sim}{A}\underset{\sim}{x}\|_2^2$ is minimized

$$\min_{c_j} \sum_i (p(x_i) - y_i)^2$$

The 2-norm corresponds to Euclidian distance and the geometric interpretation is that we want to find vector $\underset{\sim}{x}$ such the vector $\underset{\sim}{A}\underset{\sim}{x} \in \mathbb{R}^m$ in range $(\underset{\sim}{A})$ is closest to $\underset{\sim}{b}$

# Orthogonal projection and normal equations!




→ Orthogonal projection will minimize norm $\underset{\sim}{r} = \underset{\sim}{b} - \underset{\sim}{A}\underset{\sim}{x}$ in the 2-norm.

→ That magical $\underset{\sim}{x}$ that minimizes $\|r\|_2$ satisfies $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{P}\underset{\sim}{b}$ where $\underset{\sim}{P} \in \mathbb{R}^{m\times m}$ is an orthogonal projector onto range$(A)$

(i.e. residual $\underset{\sim}{r}$ must be orthogonal to range$(\underset{\sim}{A})$)

Thm 1:- Let $\underset{\sim}{A} \in \mathbb{R}^{m\times n}$ $(m \geq n)$ and full rank and $\underset{\sim}{b} \in \mathbb{R}^m$ be given. Then a vector

$x \in \mathbb{R}^n$ that minimizes $\|r\|_2 = \|b - Ax\|$

(i.e $x$ is a least squares solution) if and only if

$r \perp \text{range}(A)$

Remarks:- Let $y$ be any vector in the range$(A)$, then there exists a $d \in \mathbb{R}^n$ such that $y = Ad$

Since $r \perp \text{range}(A)$

$$y^T r = 0$$

$$\Rightarrow d^T A^T r = 0 \quad \forall d \in \mathbb{R}^n$$

$$\Rightarrow A^T r = 0$$

$$\Rightarrow A^T(b - Ax) = 0$$

$$\Rightarrow \boxed{A^T A x = A^T b} \quad — (*)$$

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

$$Ax = \underbrace{A(A^T A)^{-1} A^T}_{} b$$

$$= Pb$$

where $P \in \mathbb{R}^{m \times m}$

orthogonal projector onto range$(A)$

$$\Rightarrow \boxed{A^T A x = A^T b} \longrightarrow \text{Normal Equations}$$

$(A^T A) x = A^T b$

is $n \times n$ system of equations

that has a unique solution

if and only if $A$ has

full rank!

(*) When $A$ has a full rank, the solution $x$ to the least squares problem is unique and formally can be written as

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

This allows us to define pseudo-inverse of $A$ denoted by $A^+ = (A^T A)^{-1} A^T \in \mathbb{R}^{n \times m}$

$$\boxed{A^+ A = I} \qquad \boxed{x = A^+ b}$$

## Algorithms to solve least squares:-

$$\underset{\sim}{A}^T \underset{\sim}{A} \underset{\sim}{x} = \underset{\sim}{A}^T \underset{\sim}{b} \rightarrow \text{(Normal equation)}$$

$$\underset{\sim}{A} \underset{\sim}{x} = \underset{\sim}{P} \underset{\sim}{b} \rightarrow \text{(least squares)}$$

(i) Cholesky Factorization:- $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ $m \geq n$

If $\underset{\sim}{A}$ has full rank, then $\underset{\sim}{A}^T \underset{\sim}{A}$ is square, symmetric and positive definite

Use cholesky factorization, which factors a symmetric positive definite matrix into the form $\underset{\sim}{R}^T \underset{\sim}{R}$ where $\underset{\sim}{R}$ is upper triangular

$$\underset{\sim}{A}^T \underset{\sim}{A} = \underset{\sim}{R}^T \underset{\sim}{R}$$

$$\underset{\sim}{A}^T \underset{\sim}{A} \underset{\sim}{x} = \underset{\sim}{A}^T \underset{\sim}{b}$$

$$\underset{\sim}{R}^T \underset{\sim}{R} \underset{\sim}{x} = \underset{\sim}{A}^T \underset{\sim}{b}$$

Algo:- (1) Form $\underset{\sim}{A}^T \underset{\sim}{A}$ and $\underset{\sim}{A}^T \underset{\sim}{b}$

(2) Cholesky factorization of $\underset{\sim}{A}^T \underset{\sim}{A}$

$$\underset{\sim}{A}^T \underset{\sim}{A} = \underset{\sim}{R}^T \underset{\sim}{R}$$

to obtain $\underset{\sim}{R}^T \underset{\sim}{R} \underset{\sim}{x} = \underset{\sim}{A}^T \underset{\sim}{b}$

③ Solve Lower triangular system

$$R^T w = A^T b \text{ for } w$$

$$(w = Rx)$$

④ Solve upper triangular system

$$Rx = w \text{ for } x$$

Work $\sim mn^2 + \frac{1}{3}n^3$ flops

(ii) -via- QR factorization

$A = \hat{Q}\hat{R}$ obtain Householder triangularization

$$Ax = Pb$$

$P = \hat{Q}\hat{Q}^T$ is orthogonal projector onto range(A)

$$Ax = Pb$$

$$\Rightarrow \hat{Q}\hat{R}x = \hat{Q}\hat{Q}^T b$$

$$\Rightarrow \hat{R}x = \hat{Q}^T b \quad — (\#)$$

Algo :- ① Compute reduced QR factorization

$$A = \hat{Q}\hat{R}$$

② Compute vector $\hat{Q}^T b$

③ Solve upper triangular system

$\hat{R}x = \hat{Q}^T b$ for $x$

Work $\sim 2mn^2 - \frac{2}{3}n^3$ flops

Least squares using SVD:-

We use reduced SVD

$\underset{m\times n}{A} = \underset{m\times n}{\hat{U}} \hat{\Sigma} V^T$ to least squares problem

$Ax = Pb$ projector $P = \hat{U}\hat{U}^T$

$Ax = \hat{U}\hat{U}^T b$

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

$\Rightarrow \underbrace{\hat{\Sigma}V^T x = \hat{U}^T b}$

Algo:- ① Compute reduced SVD $A = \hat{U}\hat{\Sigma}V^T$

② Compute $\hat{U}^T b$

③ Solve diagonal system

$\hat{\Sigma} w = \hat{U}^T b$

for $w = V^T x$

(4) Solve $\underset{\sim}{V}^T \underset{\sim}{z} = \omega$ for $\underset{\sim}{x} \Rightarrow \boxed{\underset{\sim}{x} = \underset{\sim}{V}\underset{\sim}{\omega}}$

Work $\sim 2mn^2 + 11n^3$

Cholesky $\sim mn^2 + \frac{1}{3}n^2$ flops

QR factorization $\sim \underline{2mn^2 - \frac{2}{3}n^3}$ flops

SVD $\sim \underline{2mn^2 + 11n^3}$