Least - Squares Problems

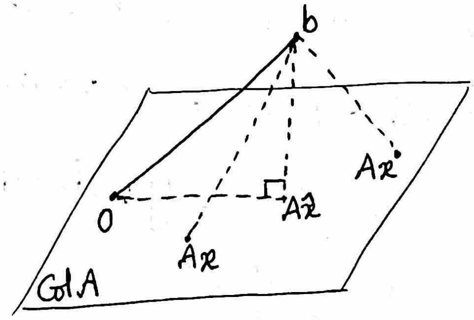
Suppose ~~that~~ the system $Ax = b$ is inconsistent, i.e. the solution doesnot exist. The best one can do is to find an x that makes Ax as close as possible to b.

Think of Ax as an approximation to b. The smaller the distance between b and Ax, given by $|b - Ax|$, the better the approximation. The general least-squares problem is to find an x that makes $|b - Ax|$ as small as possible.

Definition:-

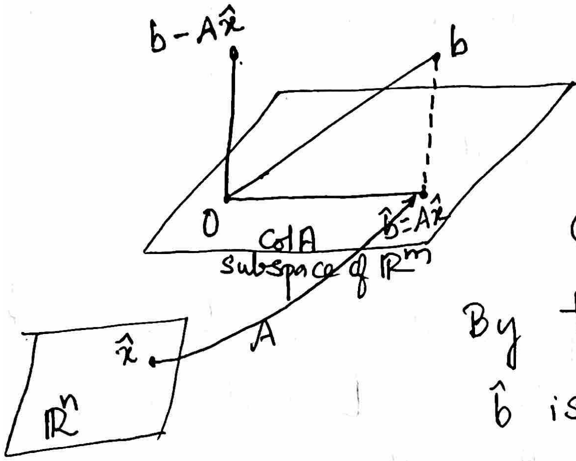
If A is $m\times n$ and b is in $\mathbb{R}^m$, a least-squares solution of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that $|b - A\hat{x}| \leq |b - Ax|$ for all x in $\mathbb{R}^n$.

No matter what x we select, the vector Ax will necessarily be in the column space of A. So we seek an x that makes Ax the closest point in $\text{Col}A$ to b.



The vector b is closer to $A\hat{x}$ than to Ax.

Solution of the General Least-Squares Problem.



Let A be an $m \times n$ matrix and b is in $\mathbb{R}^m$

Col A is a subspace of $\mathbb{R}^m$

By the Best Approximation theorem, $\hat{b}$ is the orthogonal projection of b onto Col A.

Because $\hat{b}$ is in the column space of A, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbb{R}^n$ such that $A\hat{x} = \hat{b}$. — (1)

Since $\hat{b}$ is the closest point in Col A to b, a vector $\hat{x}$ is a least-squares solution of $Ax = b$ iff $\hat{x}$ satisfies (1).

Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.

By the Orthogonal Decomposition Theorem, the projection $\hat{b}$ which lies in Col A is orthogonal to $b - \hat{b}$, i.e., $b - \hat{b}$ is orthogonal to each column of A.

i.e, $b - A\hat{x}$ is orthogonal to each column of A.

If a_j is any column of A, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$. Since each a_j^T is a row of A^T,

$$A^T(b - A\hat{x}) = 0 \quad (2) \Rightarrow A^Tb - A^TA\hat{x} = 0 \Rightarrow A^TA\hat{x} = A^Tb$$

These calculations show that each least-squares sol^n of $Ax = b$ satisfies the equation $A^TAx = A^Tb$ — (3), whose sol^n is denoted by $\hat{x}$.
The matrix equ^n (3) represents a system of equ^s called normal eq^n for $Ax = b$

Theorem

The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equations $A^TAx = A^Tb$.

example:

Find a least-squares solution of the inconsistent system $Ax = b$ for $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$

Sol:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation $A^TAx = A^Tb$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\xrightarrow{17R_2 - R_1} \begin{bmatrix} 17 & 1 \\ 0 & 84 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 168 \end{bmatrix}$$

$$\Rightarrow \quad 17x_1 + x_2 = 19 \qquad \Longrightarrow \qquad x_2 = 2$$

$$84x_2 = 168 \qquad\qquad\qquad x_1 = 1$$

$\therefore$ $\hat{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is the least-squares solution.

example:-

Find a least-squares solution of $Ax = b$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

Sol^n:-

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

Then $A^T A x = Ab \Rightarrow$

$$\begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix $[A^T A : Ab] \Rightarrow$

$$\left[\begin{array}{cccc:c} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{array}\right] \underset{3R_2-R_1;\ 3R_3-R_1;\ 3R_4-R_1}{\sim} \left[\begin{array}{cccc:c} 6 & 2 & 2 & 2 & 4 \\ 0 & 4 & -2 & -2 & -16 \\ 0 & -2 & 4 & -2 & 2 \\ 0 & -2 & -2 & 4 & 14 \end{array}\right] \underset{\substack{2R_3+R_2 \\ 2R_4+R_2}}{\sim} \left[\begin{array}{cccc:c} 6 & 2 & 2 & 2 & 4 \\ 0 & 4 & -2 & -2 & -16 \\ 0 & 0 & 6 & -6 & -12 \\ 0 & 0 & -6 & 6 & 12 \end{array}\right]$$

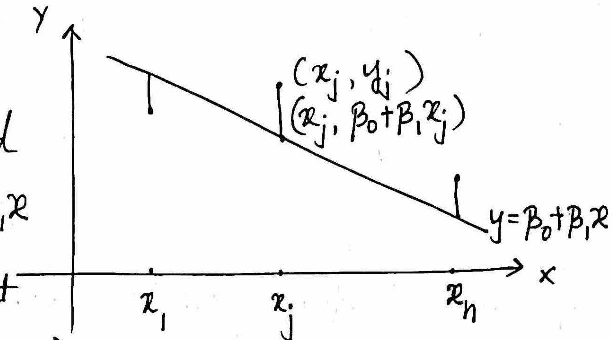
$$\underset{R_4+R_3}{\sim} \left[\begin{array}{cccc:c} 6 & 2 & 2 & 2 & 4 \\ 0 & 4 & -2 & -2 & -16 \\ 0 & 0 & 6 & -6 & -12 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \left.\begin{aligned} 6x_1 + 2x_2 + x_3 + 2x_4 &= 4 \\ 4x_2 - 2x_3 - 2x_4 &= -16 \\ 6x_3 - 6x_4 &= -12 \end{aligned}\right\} \Rightarrow \begin{aligned} x_3 &= -2 + x_4 \\ x_2 &= -5 + x_4 \\ x_1 &= 3 - x_4 \end{aligned} \quad \therefore \hat{x} = \begin{bmatrix} 3 - x_4 \\ -5 + x_4 \\ -2 + x_4 \\ x_4 \end{bmatrix}$$

~~let $x_4 = 0 \Rightarrow x_3 = -2,\ x_2 = -5,\ x_1 = 3$~~

$$\hat{x} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Applications to linear models

Suppose we want to fit a linear equation of the form $y = \beta_0 + \beta_1 x$ to the data points $(x_1, y_1) \dots, (x_n, y_n)$, that, when graphed, seem to lie close to a line. We want to determine the parameters β_0 and β_1 that make the line as close to the points as possible.



Suppose β_0 and β_1 are fixed and consider the line $y = \beta_0 + \beta_1 x$. Corresponding to each data point (x_j, y_j) there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinates.

We call y_j the observed value of y and $\beta_0 + \beta_1 x_j$ the predicted y-value. The difference between an observed y-value and a predicted y-value is called a residual. The best fit line to these given data is such that, the square of these residuals is minimum. And that is the least-square line $y = \beta_0 + \beta_1 x$. This line is also called a line of regression of y on x, because any errors in the data are assumed to be only in the y-coordinates. The coefficients β_0, β_1 of the line are called regression coefficients.

If the data points were on the line, the parameters β_0 and β_1 would satisfy the equations $\longrightarrow$

predicted y-value		observed y-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n

We can write the system as

$$X\beta = y, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

If the data points don't lie on a line, then there are no parameters β_0, β_1 for which the predicted y-values in $X\beta$ equal the observed y-values in y and $X\beta = y$ has no solution. This is a least-squares problem $Ax = b$, with different notation.

The square of the distance between the vectors $X\beta$ and y is precisely the sum of the squares of the residuals. The β that minimizes the sum also minimizes the distance between $X\beta$ and y.

Computing the least-squares solution of $X\beta = y$ is equivalent to finding the β that determines the least-squares line $(y = \beta_0 + \beta_1 x)$.

example:

Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2,1)$, $(5,2)$, $(7,3)$, $(8,3)$.

Sol^n

Using the x-coordinates and y-coordinates of the data points, we can write

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution $X\beta = y$, we obtain the normal equations by $X^T X \beta = X^T y$.

ie,
$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence
$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \begin{bmatrix} \frac{2}{7} \\ \frac{5}{14} \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14} x$$

example:

Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(-1,0), (0,1), (1,2), (2,4)$.

Sol: From the data points, we have,

$$X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

$0+1+2+4$

$0+0+2+8$

The normal eqns are $X^T X \beta = X^T y$.

ie, $$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

$24 - 4$

$42 - 20 = 22/20$

$-14 + 40 = 26/20$

Hence $$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 7 \\ 10 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 7 \\ 10 \end{bmatrix} = \begin{bmatrix} \frac{11}{20} \\ \frac{13}{10} \end{bmatrix}$$

$\therefore$ $\underline{y = \frac{11}{10} + \frac{13}{10} x}$ is the required least-squares Line.

example:

Find the line of best fit for the below data:

$b=2$ at $t=-1$, $b=0$ at $t=0$, $b=-3$ at $t=1$,

$b=-5$ at $t=2$.

sol^n: Given $(-1,2)$, $(0,0)$, $(1,-3)$, $(2,-5)$.

Using the data points, we can write,

$$X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ 0 \\ -3 \\ -5 \end{bmatrix}$$

For the least squares solution $X\beta = y$, we obtain the normal equations by $X^T X\beta = X^T y$.

ie $$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -3 \\ -5 \end{bmatrix} = \begin{bmatrix} -6 \\ -15 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} -6 \\ -15 \end{bmatrix}$$

Hence $$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} -6 \\ -15 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} -6 \\ -15 \end{bmatrix} = \begin{bmatrix} -\frac{6}{20} \\ -\frac{48}{20} \end{bmatrix} = \begin{bmatrix} -\frac{3}{10} \\ -\frac{24}{10} \end{bmatrix}$$

Thus the least-squares or the line of best fit is

$$y = -\frac{3}{10} - \frac{24}{10}x \quad \text{or} \quad \boxed{b = -0.3 - 2.4t}$$

A healthy child's systolic blood pressure p (in millimeters of mercury) and weight w (in pounds) are approximately related by the equation

$\beta_0 + \beta_1 \ln w = p.$

Use the following experimental data to estimate the systolic blood pressure of a healthy child weighing 100 pounds.

w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
p	91	98	103	110	112

$X_{5\times2}$ $X^T_{2\times5}$ $y_{5\times1}$

$$X^TX = \begin{bmatrix} 5 & 21.89 \\ 21.89 & 96.6399 \end{bmatrix}$$

$$X^Ty = \begin{bmatrix} 514 \\ 2265.8 \end{bmatrix}$$

$$X^TX\beta = X^Ty$$

$$\Rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = [X^TX]^{-1}[X^Ty]$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 18.5642 \\ 19.2407 \end{bmatrix}$$

$$P = 18.56 + 19.24 \ln w$$

at $w = 100$ $\quad P = 18.56 + 19.24 \ln 100$

$= 107.16$

$P \approx 107$