6.5 Least-squares Problems

Problem 1: Given

$$\vec{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 3 \\ -3 \\ 2 \\ 5 \\ 5 \end{bmatrix}, \quad \text{and} \quad \vec{x}_3 = \begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 8 \end{bmatrix}$$

i) Show that they are linearly independent

ii) Show that they are not orthogonal.

iii) Find the best approximation to $\vec{x}_3$ in the Subspace $W = \text{Span}\{\vec{x}_1, \vec{x}_2\}$

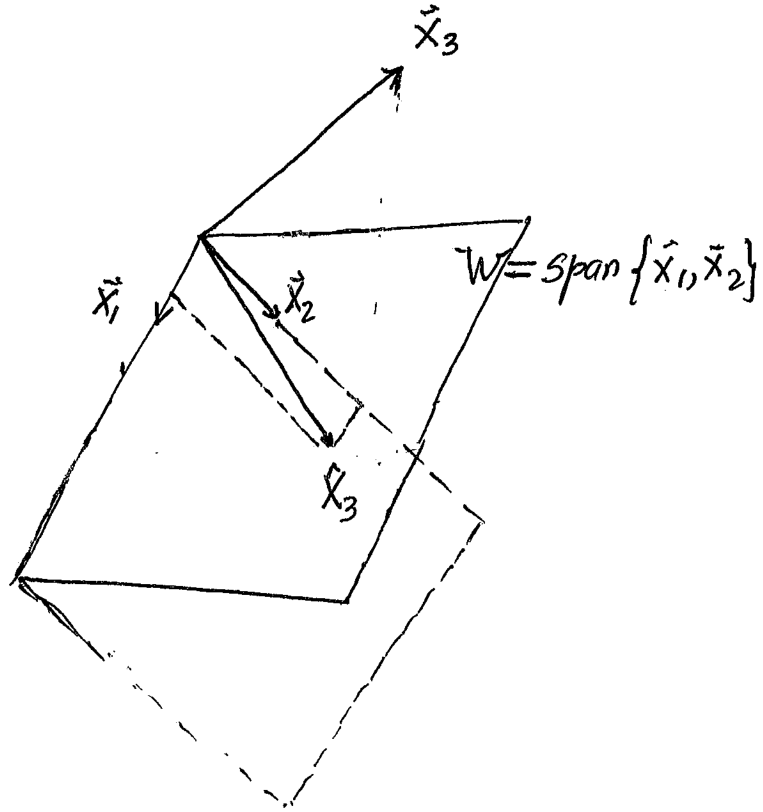
Answ:

i) Easy done in problem #12 (Sect. 6.4) by row-reduction.

ii) Clearly, $\vec{x}_1 \cdot \vec{x}_2 = 16, \neq 0$, $\vec{x}_1 \cdot \vec{x}_3 = 14$, $\vec{x}_2 \cdot \vec{x}_3 = 68$.

iii) Before we obtain the best approx. to $\vec{x}_3$ in W, we transform the basis $B = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \\ 2 \\ 5 \\ 5 \end{bmatrix} \right\}$ into an orthogonal basis for W.

The reason is because we do not know yet how to obtain an orthogonal projection into a non-orthogonal basis (See graph below).



Applying G-S process:

$$\vec{v}_1 = \vec{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_2 = \vec{x}_2 - \text{proj}_{\vec{v}_1} \vec{x}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1$$

$$\vec{v}_2 = \begin{bmatrix} 3 \\ -3 \\ 2 \\ 5 \\ 5 \end{bmatrix} - \frac{16}{4} \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

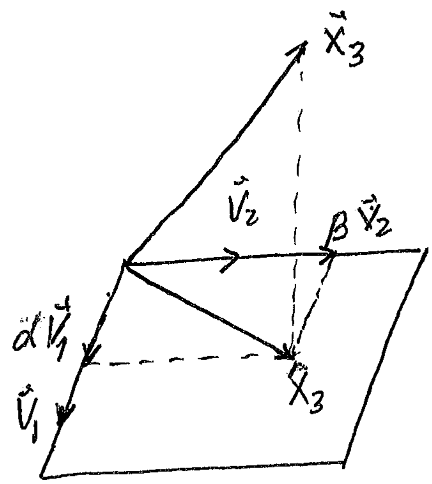
The new set $B_0 = \left\{ \overset{\vec{v}_1}{\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}}, \overset{\vec{v}_2}{\begin{bmatrix} -1 \\ 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}} \right\}$

is such that $\text{Span}\{\vec{v}_1, \vec{v}_2\} = \text{Span}\{\vec{x}_1, \vec{x}_2\} = W$

and because they are orthogonal they are linearly indep.

And as a consequence, they form an orthogonal basis for W.

Now, we can find $\hat{x}_3$ by projecting $\vec{x}_3$ into $W = \text{span}\{\vec{v}_1, \vec{v}_2\}$.



$$\hat{x}_3 = \text{proj}_W \vec{x}_3 = \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2$$

or

$$\hat{x}_3 = \frac{7}{2}\vec{v}_1 + \frac{3}{2}\vec{v}_2 = \begin{bmatrix} 2 \\ -2 \\ 3 \\ 5 \\ 5 \end{bmatrix}$$

Notice from the graph above

$$\hat{x}_3 = \alpha\vec{v}_1 + \beta\vec{v}_2$$

$$\Rightarrow \quad \alpha = \frac{7}{2}, \quad \beta = \frac{3}{2}.$$

Another way to look at this projection problem is by introducing a new matrix A.

In fact, if we define

$$A = [\hat{v}_1\ \hat{v}_2] = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 0 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

and $\hat{b} = \begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 8 \end{bmatrix}$

A question may be

i) Find a solution, if possible for

$$A\vec{x} = \vec{b}$$

ii) If not possible, find an $\hat{x}$ such that $A\hat{x} = \hat{b}$, where $\hat{b}$ is the best approximation of $\vec{b}$ in

$$\text{Span}\{\hat{v}_1, \hat{v}_2\} = \text{Col}(A).$$

This is equivalent to say

Find an $\hat{x} \in \mathbb{R}^2$ such that

$$\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{x}\|, \quad \text{for all } \vec{x} \in \mathbb{R}^2.$$

Answ:

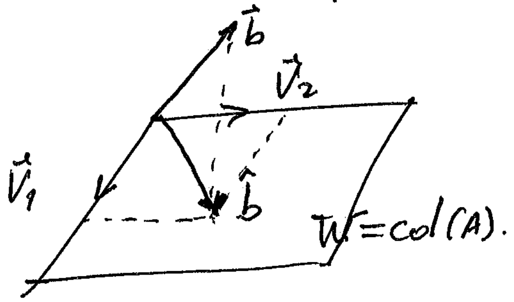
i) There is no Solution for $A\vec{x}=\vec{b}$,

Since it can be easily proved that

the set $\{\hat{v}_1, \hat{v}_2, \hat{b}\}$ is linearly independent

In other words,

$$\hat{b} \notin Col(A).$$



ii)

From our previous Computation, The best approximation to $\vec{b}=\begin{bmatrix}5\\1\\3\\2\\8\end{bmatrix}$ in $Col(A)$, is $\hat{b}=\begin{bmatrix}2\\2\\-3\\5\\5\end{bmatrix}$

that is reached when $\hat{x}=\begin{bmatrix}7/2\\3/2\end{bmatrix}$

This vector $\hat{x}$ is also called, a least-squares Solution of $A\vec{x}=\vec{b}$.

What do the components of the vector $\hat{x}$ represent?

Definition

$A_{m\times n}$ matrix

$\vec{b} \in \mathbb{R}^m$

i) Then, a least-squares solution of $A\vec{x}=\vec{b}$ is an $\hat{\vec{x}} \in \mathbb{R}^n$ such that

$$\|\vec{b} - A\hat{\vec{x}}\| \leq \|b - A\vec{x}\|, \text{ for all } \vec{x} \in \mathbb{R}^n.$$

ii) The distance

$$\|\vec{b} - A\hat{x}\|$$

is called the least-squares error of this approximation.

In our previous example the computation of a least-squares solution of $A\vec{x}=\vec{b}$ was fairly easy because the columns of A were orthogonal vectors.

In our next exercise, we will work directly with a matrix A with non-orthogonal columns and we will find out a least-squares solution for $A\vec{x}=\vec{b}$ without obtaining orthogonal columns for A first.

Exercise Reconsider Exercise 12 (Sect. 6.4).

and define

$$A = \begin{array}{c} \begin{array}{cc} \vec{v}_1 & \vec{v}_2 \end{array} \\ \begin{bmatrix} 1 & 3 \\ -1 & -3 \\ 0 & 2 \\ 1 & 5 \\ 1 & 5 \end{bmatrix} \end{array}, \qquad \vec{b} = \begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 8 \end{bmatrix}$$

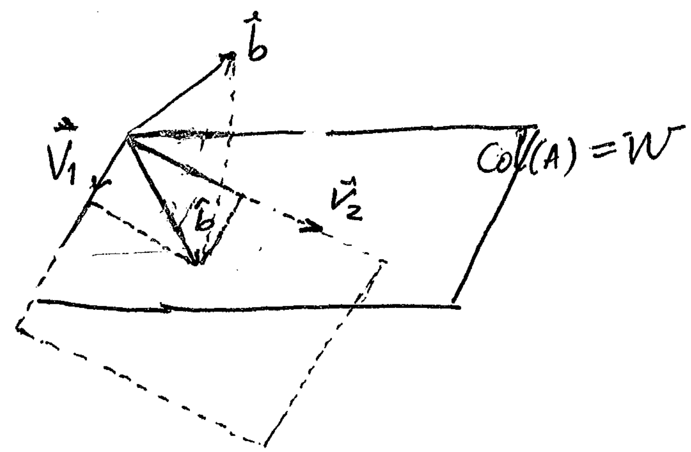
Find a least-squares solution of

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

and compute the associated least-squares error.

---

answ: Obviously $\vec{v}_1 \cdot \vec{v}_2 \neq 0$. They are non-orthogonal.



A least-squares solution will satisfy

$$A\hat{x} = \hat{b} = \text{proj}_{\text{Col}(A)} \vec{b}$$

We will obtain $\hat{x}$ without computing $\hat{b}$ first as we did in previous exercise.

In fact, $\hat{b}$ is such that (best approx. thm.)

$(\vec{b}-\hat{b})$ is orthogonal to any vector in

$Col(A) = W$.

In our case, if $\hat{x}$ is a least-squares solution of $A\vec{x}=\vec{b}$
$A\hat{x} = \hat{b}$

and $\quad \vec{v}_1 \cdot (\vec{b}-\hat{b}) = \vec{v}_1 \cdot (\vec{b} - A\hat{x}) = 0$

and $\quad \vec{v}_2 \cdot (\vec{b}-\hat{b}) = \vec{v}_2 \cdot (\vec{b} - A\hat{x}) = 0$

$\iff$

$$\vec{v}_1 \cdot A\hat{x} = \vec{v}_1 \cdot \vec{b}$$

and $\quad \vec{v}_2 \cdot A\hat{x} = \vec{v}_2 \cdot \vec{b}$

$\iff$

$$\vec{v}_1 \cdot (x_1\vec{v}_1 + x_2\vec{v}_2) = \vec{v}_1 \cdot \vec{b}$$

$$\vec{v}_2 \cdot (x_1\vec{v}_1 + x_2\vec{v}_2) = \vec{v}_2 \cdot \vec{b}$$

$$\Leftrightarrow \begin{bmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \vec{v}_1^T \vec{b} \\ \vec{v}_2^T \vec{b} \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} \vec{v}_1^T v_1 & \vec{v}_1^T \vec{v}_2 \\ \vec{v}_2^T \vec{v}_1 & \vec{v}_2^T \vec{v}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \vec{v}_1^T b \\ \vec{v}_2^T \vec{b} \end{bmatrix}$$

$$\Leftrightarrow \underbrace{\begin{bmatrix} \vec{V}_1^T \longrightarrow \\ \vec{V}_2^T \longrightarrow \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} \vec{V}_1 & \vec{V}_2 \\ \downarrow & \downarrow \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} \vec{V}_1^T \longrightarrow \\ \vec{V}_2^T \longrightarrow \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} \vec{b} \\ \downarrow \end{bmatrix}}_{\vec{b}}$$

$\Leftrightarrow$

or $$\boxed{A^T A \hat{x} = A^T \vec{b}} \qquad (8.1)$$

It means that if $\hat{x}$ is a least-squares solution of $A\vec{x} = \vec{b}$ then, it satisfies the normal equations (8.1) The reciprocal is also true. This is the content of theorem 13.

In our exercise,

$$A^T A = \begin{bmatrix} 1 & -1 & 0 & 1 & 1 \\ 3 & -3 & 2 & 5 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & -3 \\ 0 & 2 \\ 1 & 5 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 4 & 16 \\ 16 & 72 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} 1 & -1 & 0 & 1 & 1 \\ 3 & -3 & 2 & 5 & 5 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 14 \\ 68 \end{bmatrix}$$

Therefore,

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

reduces to

$$\begin{bmatrix} 4 & 16 \\ 16 & 72 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 14 \\ 68 \end{bmatrix}$$

$$\Rightarrow \quad x_1 = -\frac{5}{2}, \quad x_2 = \frac{3}{2} \Rightarrow \boxed{\hat{x} = \begin{bmatrix} -5/2 \\ 3/2 \end{bmatrix}}$$

Least-squares Solution.

This least-squares solution is different than the one obtained when the columns of A are orthogonal. It shows that the least-squares solution depends on the basis used for $Col(A)$. However, the vector $\hat{b} = proj_W \vec{b}$ is unique. It does not depend in the chosen basis.

In fact,

$$A\hat{x} = \begin{bmatrix} 1 & 3 \\ -1 & -3 \\ 0 & 2 \\ 1 & 5 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} -5/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 3 \\ 5 \\ 5 \end{bmatrix} = \hat{b} = \text{proj}_{\text{Col}(A)} \vec{b}$$

<u>least-squares error:</u>

$$\|\vec{b} - \hat{b}\|^2 = \left\| \begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 8 \end{bmatrix} - \begin{bmatrix} 2 \\ -2 \\ 3 \\ 5 \\ 5 \end{bmatrix} \right\|^2 = 9+9+9+9 = 36$$

$$\Rightarrow \boxed{\|\vec{b} - \hat{b}\| = 6}$$

<u>Thm 13</u>

The set of least-squares solutions of $A\vec{x} = \vec{b}$ coincides with the nonempty set of solutions of the normal equations

$$A^T A x = A^T \vec{b}.$$

For a given matrix A is possible to have more than one least-squares solution. The existence of only one solution is discussed in the next theorem.

Thm 14 $\quad A$ $m \times n$. The following statements are equivalent:

a) $A\hat{x} = \vec{b}$ has a unique least-squares solution for each $\vec{b} \in \mathbb{R}^m$.

b) The columns of A are linearly independent.

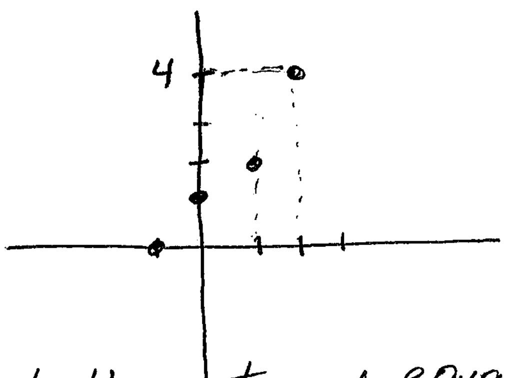
c) The matrix A^TA is invertible.

6.6. Applications to Linear Models

Least-Squares Lines.

Exercise #3 (Sect. 6.6).

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



We can write the system of equations:

$$\begin{aligned} \beta_0 + (-1)\beta_1 &= 0 \\ \beta_0 + 0\beta_1 &= 1 \\ \beta_0 + 1\beta_1 &= 2 \\ \beta_0 + 2\beta_1 &= 4 \end{aligned} \xrightarrow[\text{form}]{\text{In matrix}} \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\vec{\beta}} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}}_{\vec{y}}$$

This System has a Solution if $\begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix} = \vec{y} \in Col(X)$

Check : Augmented matrix

$$\begin{array}{ccc} \vec{v}_1 & \vec{v}_2 & \vec{b} \end{array}$$

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 3 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 3 & 4 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \sim$$

$$\sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 3 & 4 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

→ Column Vectors Lin. indep.

No solution.

$\Rightarrow \{\vec{v}_1, \vec{v}_2, \vec{b}\}$ is lin. indep.

We are interested in obtaining a line L such that the squares of the differences

$$\boxed{\text{residual}_j = r_j = y_j - \beta_0 + \beta_1 x_j, \quad j = 1,2,3,4.}$$

is minimized.

This is the same as the least-squares error of the least-squares problem

$$X\vec{\beta} = \vec{y}. \tag{13.1}$$

So let's find $\hat{\beta}$ the least-squares solution of (13.1).

Using the normal equations:

$$X^T X \hat{\beta} = X^T \vec{y}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

or

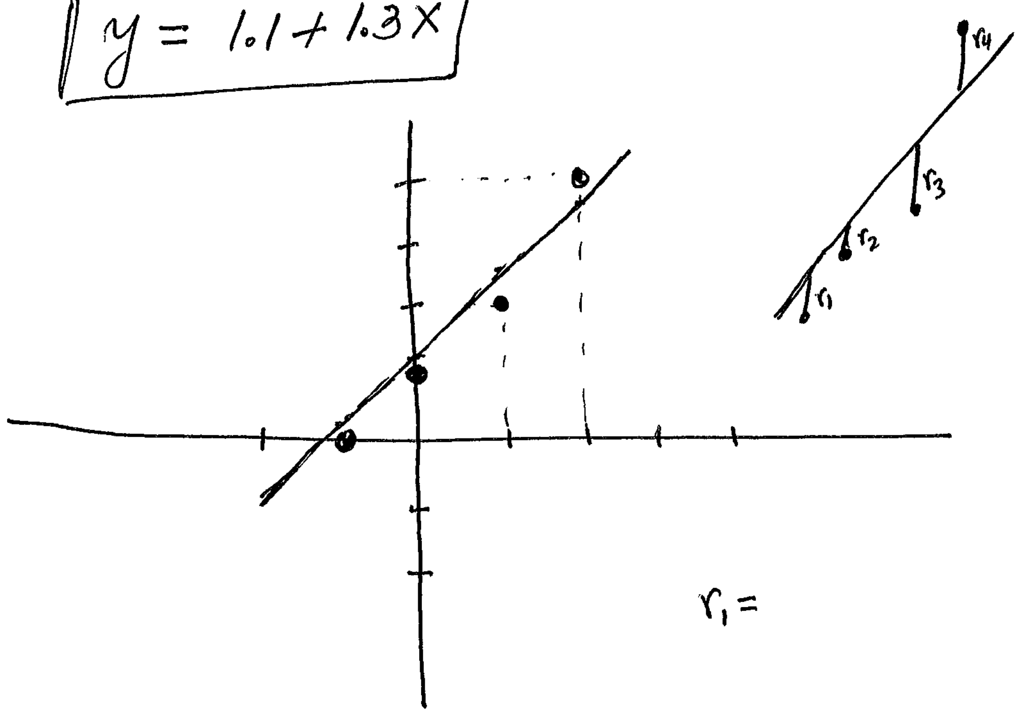
$$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 2 & 7 \\ 2 & 6 & 10 \end{bmatrix} \sim \begin{bmatrix} 2 & 1 & 7/2 \\ 2 & 6 & 10 \end{bmatrix} \sim \begin{bmatrix} 2 & 1 & 7/2 \\ 0 & 5 & 13/2 \end{bmatrix}$$

$$\Rightarrow \quad \beta_1 = 13/10, \quad \beta_0 = \frac{7/2 - 13/10}{2} = \frac{22}{20} = \frac{11}{10}$$

or

$$\boxed{y = 1.1 + 1.3X}$$



$r_1 =$

$$\hat{y} = X\hat{\beta}$$

$$\hat{y} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 11/10 \\ 13/10 \end{bmatrix} = \begin{bmatrix} -1/5 \\ 11/10 \\ 12/5 \\ 37/10 \end{bmatrix}$$

$$\vec{y} - \hat{y} = \vec{r} \quad \text{or} \quad \vec{r} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} -1/5 \\ 11/10 \\ 12/5 \\ 37/10 \end{bmatrix} = \begin{bmatrix} 1/5 \\ -1/10 \\ -2/5 \\ 3/10 \end{bmatrix}$$

Least squares error

$$\varepsilon_{\text{error}}^2 = r_1^2 + r_2^2 + r_3^2 + r_4^2 = \|\vec{r}\|^2 = \|\vec{y} - \hat{y}\|^2.$$

Common form of Least Squares Equations for Linear Regression

Sect 6.6

#15) Consider the Normal equations

A $m \times n$ matrix $A^T A \hat{x} = A^T \vec{b}$.

and

(x_1, y_1)
(x_2, y_2)
$\vdots$
(x_m, y_m)

$\hat{y} = X\vec{\beta}$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} = \begin{bmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{bmatrix}$$

Then

$X^T X \vec{\beta} = X^T \vec{y}$ is equivalent to

$$\boxed{\begin{aligned} m\beta_0 + \beta_1 \sum_{i=1}^{m} x_i &= \sum_{i=1}^{m} y_i \\ \beta_0 \sum_{i=1}^{m} x_i + \beta_1 \sum_{i=1}^{m} x_i^2 &= \sum_{i=1}^{m} x_i y_i \end{aligned}}$$