

Week 14
June 1, 3

①

Review

Given a subspace V of $\mathbb{R}^m$. Let $\vec{b}$ be a vector of $\mathbb{R}^m$, $\vec{p}$ be the projection of $\vec{b}$ onto V . This implies that $\vec{b} - \vec{p}$ is a projection of $\vec{b}$ onto $V^\perp$.

$$\vec{p} = A \vec{x}' \quad \text{where} \quad C(A) = V.$$

(因為 $\vec{p}$ 在 V 中, $\vec{p}$ 自然可以由 V 中的向量來表示; 所以只要取一組向量 $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ 可以 $\text{Span } V$, $\vec{p} = \sum_{i=1}^n c_i \vec{v}_i$

$$A^T A \vec{x}' = A^T \vec{b} \quad \dots \dots \dots \text{Normal Equation}$$

(*) If $A^T A$ is invertible, then $\vec{x}' = (A^T A)^{-1} A^T \vec{b}$.

$$\vec{p} = A \vec{x}' \quad \dots \dots \text{Easy to obtain.}$$

So, when $A^T A$ is nonsingular (invertible)?

Theorem Let A be an $m \times n$ matrix and $\text{rank}(C(A)) = n$. Then $A^T A$ is nonsingular.

Proof. Note that $A^T A$ is a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^n$. It suffices to claim that the nullity of $A^T A$ is zero, i.e., $\text{Null}(A^T A) = \{\vec{0}\}$.

Let $A^T A \vec{x} = \vec{0}$. Then $A \vec{x} \in \text{Null}(A^T) = C(A)^\perp$.

Since $A \vec{x} \in C(A)$, $A \vec{x} \in C(A)^\perp \cap C(A)$ and thus $A \vec{x} = \vec{0}$.

Now, since $\text{rank}(C(A)) = n$, the only solution for $A \vec{x} = \vec{0}$ is $\vec{x} = \vec{0}$. This concludes the proof. $\blacksquare$

Example If we are dealing with $V = \text{Span}(\{\vec{v}\})$ where $\vec{v} \neq \vec{0}$, then $A^T A = \langle \vec{v}, \vec{v} \rangle \neq 0$. Hence $\vec{x}' = \frac{A^T \vec{b}}{\langle \vec{v}, \vec{v} \rangle} = \frac{\langle \vec{v}, \vec{b} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}$. The projection of $\vec{b}$ onto V is $\vec{p} = \frac{\langle \vec{v}, \vec{b} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}$. $\blacksquare$

(**) Let $\underline{P} = A(A^T A)^{-1} A^T$. (Projection Matrix!)

Then $P \vec{b} = \vec{p}$.

(Note) $(A^T A)^{-1} \neq A^{-1} (A^T)^{-1}$. (?)

3

Application

Definition (Least-Squares Solution)

A least-squares solution to an inconsistent system $A\vec{x} = \vec{b}$ is a solution to the normal equations $A^T A \vec{x}' = A^T \vec{b}$ (无解)

($\vec{x}'$)

Example

$$\begin{cases} x_1 - x_2 = -3 \dots ① \\ 2x_1 + x_2 = 0 \dots ② \\ 3x_1 + 2x_2 = 6 \dots ③ \end{cases} \quad \left(\begin{array}{l} \text{No solution! From ① and ②} \\ \text{we have } x_1 = -1 \text{ and } x_2 = 2. \\ \text{But, by ③ } 3x_1 + 2x_2 = 6 \text{ (not 1).} \end{array} \right.$$

⇓

$$\begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}$$

$$\text{Now, } A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}$$

$$A^T A \vec{x}' = A^T \vec{b}$$

//

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 14 & 7 \\ 7 & 6 \end{bmatrix}$$

(invertible)

$$\Rightarrow \vec{x}' = (A^T A)^{-1} A^T \vec{b} = \begin{bmatrix} -\frac{3}{7} \\ 3 \end{bmatrix}$$

④

↙ $\vec{x}''$ Consider solution $(-1, 2)$.

$$x_1 - x_2 = -3$$

$$2x_1 + x_2 = 0$$

$$3x_1 + 2x_2 = 1$$

So, the corresponding $\vec{b}' = (-3, 0, 1)$ Consider solution $(-\frac{3}{7}, 3)$.

$$x_1 - x_2 = -\frac{24}{7}$$

$$2x_1 + x_2 = \frac{15}{7}$$

$$3x_1 + 2x_2 = \frac{33}{7}$$

$$A\vec{x}' = \begin{bmatrix} -\frac{24}{7} \\ \frac{15}{7} \\ \frac{33}{7} \end{bmatrix}$$

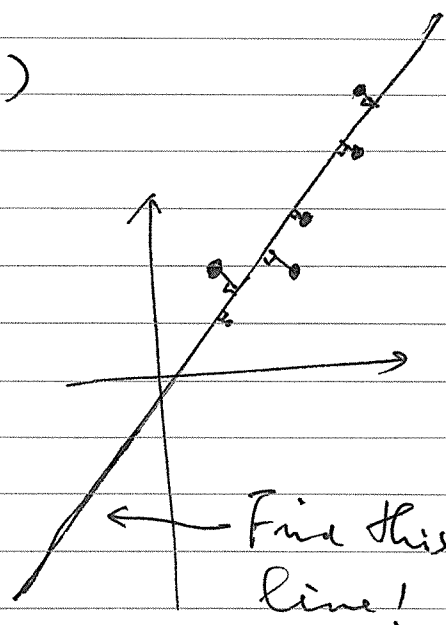
$$\|A\vec{x}' - \vec{b}\|$$

$$= \left(\left(\frac{3}{7}\right)^2 + \left(\frac{15}{7}\right)^2 + \left(\frac{9}{7}\right)^2 \right)^{1/2}$$

$$= \frac{(9 + 225 + 81)^{1/2}}{7} = \frac{\sqrt{315}}{7} \approx \frac{17.7}{7}$$

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

$$= \left(0^2 + 0^2 + 5^2 \right)^{1/2} = 5$$



≤ 3

($A\vec{x}'$ 所得到的答案最接近 $\vec{b}'$)

Why?

 $(-3, 0, 1)$

⑤

Orthogonal Bases and Projections

Note that if $A^T A$ is nonsingular, then we have a unique solution $\vec{x}' = (A^T A)^{-1} A^T \vec{b}$ for each $\vec{b} \in \mathbb{R}^m$, i.e. the projection of $\vec{b}$, $\vec{p} = A \vec{x}'$ is uniquely determined.

But, sometimes find $(A^T A)^{-1}$ is very time-consuming. (浪费时间)

(*) If $A = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n]$ is composed with independent columns (basis) and satisfies the following conditions, then $A^T A = I$;

Condition (1): $\langle \vec{v}_i, \vec{v}_j \rangle = 0$, and Condition (2) $\|\vec{v}_i\| = 1, \forall i$.

Definition (Orthonormal Basis)

Let V be a subspace of $\mathbb{R}^m$ and $B = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, $n \leq m$, be a basis of V . B is an orthonormal basis of

V if (1) $\langle \vec{v}_i, \vec{v}_j \rangle = 0$, $i \neq j$ and (2) $\|\vec{v}_i\| = 1 \forall i = 1, 2, \dots, n$

Example

$\{(1, 0, 0), (0, 1, 0)\}$ is an orthonormal basis of $x-y$ plane in $\mathbb{R}^3$

6

$$A = [\vec{w}_1 \ \vec{w}_2 \ \dots \ \vec{w}_n]$$

$\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$ is an orthonormal basis of $C(A)$.

$\Rightarrow$

$$A^T A = \begin{bmatrix} \vec{w}_1 \\ \vec{w}_2 \\ \vdots \\ \vec{w}_n \end{bmatrix} [\vec{w}_1 \ \vec{w}_2 \ \dots \ \vec{w}_n]$$

$$= \begin{bmatrix} \langle \vec{w}_i, \vec{w}_j \rangle \end{bmatrix}_{n \times n} = I_n.$$

Normal Equation becomes

$$\vec{x}' = (A^T A)^{-1} A^T \vec{b}$$

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

(Great!)

More properties:

$$\vec{p} = A \vec{x}' =$$

$$\begin{bmatrix} \vec{w}_1 \\ \vec{w}_2 \\ \vdots \\ \vec{w}_n \end{bmatrix} \begin{bmatrix} \langle \vec{w}_1, \vec{b} \rangle \\ \langle \vec{w}_2, \vec{b} \rangle \\ \vdots \\ \langle \vec{w}_n, \vec{b} \rangle \end{bmatrix}$$

$$= \sum_{i=1}^n \vec{w}_i \langle \vec{w}_i, \vec{b} \rangle$$

(7)

If $\vec{b} \in \text{Span}(\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\})$,

then $\vec{b} = \underbrace{\langle \vec{w}_1, \vec{b} \rangle}_{c_1} \vec{w}_1 + \dots + \underbrace{\langle \vec{w}_m, \vec{b} \rangle}_{c_m} \vec{w}_m$.

Proof. $\vec{b} = \sum_{i=1}^m c_i \vec{w}_i$

$\underbrace{\langle \vec{w}_i, \vec{b} \rangle}_{c_i} = c_i \quad \checkmark$

Orthonormal!
Basis

Note

If we use orthogonal basis instead of orthonormal, then some adjustment is needed!

$\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ orthogonal basis



$\left\{ \frac{\vec{v}_1}{\|\vec{v}_1\|}, \frac{\vec{v}_2}{\|\vec{v}_2\|}, \dots, \frac{\vec{v}_m}{\|\vec{v}_m\|} \right\}$ orthonormal basis!

我们不用 (5.3.5) 的 $\vec{w}_i$ 而是 $\frac{\vec{v}_i}{\|\vec{v}_i\|}$ 来构造

(8)

How to use orthonormal basis?

If $\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$ is an orthonormal basis of $\mathbb{R}^m$, then for any two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^m$,

$$\langle \vec{u}, \vec{v} \rangle = \sum_{i=1}^m \langle \vec{u}, \vec{w}_i \rangle \langle \vec{v}, \vec{w}_i \rangle.$$

Proof. Let $\vec{u} = \alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2 + \dots + \alpha_m \vec{w}_m$ and

$$\vec{v} = \beta_1 \vec{w}_1 + \beta_2 \vec{w}_2 + \dots + \beta_m \vec{w}_m.$$

$$\text{Then, } \langle \vec{u}, \vec{v} \rangle = \left\langle \sum_{i=1}^m \alpha_i \vec{w}_i, \sum_{j=1}^m \beta_j \vec{w}_j \right\rangle$$

$$= \sum_{i=1}^m \alpha_i \beta_i. \quad \text{Since } \alpha_i = \langle \vec{u}, \vec{w}_i \rangle \text{ and } \beta_j = \langle \vec{v}, \vec{w}_j \rangle,$$

we have the proof. 

How to obtain an orthonormal basis?

Theorem (Gram-Schmidt)

Every subspace V of $\mathbb{R}^m$ has an orthogonal basis and hence an orthonormal basis.

Proof.

Can you do it yourself?

(9)

Gram-Schmidt Process

Given : An n -dim subspace of $\mathbb{R}^m$ ($n \leq m$).

Goal : To produce an orthogonal basis for V .

Procedure :

Step 1. Choose any nonzero vector $\vec{v}_1 \in V$.

Step 2. Choose any $\vec{u}_2 \in V \setminus \text{Span}(\{\vec{v}_1\})$ and let $\vec{v}_2$ be its projection onto the orthogonal complement of this subspace, i.e. the subspace $(\text{Span}(\{\vec{v}_1\}))^\perp$:

$$\vec{v}_2 = \vec{u}_2 - \frac{\langle \vec{v}_1, \vec{u}_2 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1.$$

Step 3. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$ be selected vectors after ≥ 2 steps and $V_k = \text{Span}(\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\})$. Let $\vec{u}_{k+1}$ be a vector in $V \setminus V_k$ and $\vec{v}_{k+1}$ be its projection onto $V_k^\perp$.

$$\begin{aligned} \vec{v}_{k+1} &= \vec{u}_{k+1} - \frac{\langle \vec{v}_1, \vec{u}_{k+1} \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 - \dots - \frac{\langle \vec{v}_k, \vec{u}_{k+1} \rangle}{\langle \vec{v}_k, \vec{v}_k \rangle} \vec{v}_k \\ &= \vec{u}_{k+1} - \sum_{i=1}^k \frac{\langle \vec{v}_i, \vec{u}_{k+1} \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \vec{v}_i. \end{aligned} \quad (1)$$

Step 4. $k=n$, stops.

Proof. It suffices to prove that, $\forall k \geq 2$, (1) holds.

Note that " $\vec{v}_{k+1}$ is a projection onto $V_k^\perp$ " $\Rightarrow \vec{v}_{k+1} \in V_k^\perp$ and

thus $\vec{v}_{k+1} \perp \vec{v}_i \quad \forall i = 1, 2, \dots, k$.

By induction

We show an easier case where $k = 2$.

$$\vec{v}_2 = \vec{u}_2 - \frac{\langle \vec{v}_1, \vec{u}_2 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1.$$

$\vec{v}_3$ is the projection of $\vec{u}_3$ onto $V_2 = \text{Span}(\{\vec{v}_1, \vec{v}_2\})$.

Let $\vec{p}_3$ be the projection of $\vec{u}_3$ onto $\text{Span}(\{\vec{v}_1, \vec{v}_2\})$.

$$\text{Then } \vec{p}_3 = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \vec{x}'$$

$$= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \left(\begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}^T \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}^T \vec{u}_3$$

$$= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \left(\begin{bmatrix} \langle \vec{v}_1, \vec{v}_1 \rangle & 0 \\ 0 & \langle \vec{v}_2, \vec{v}_2 \rangle \end{bmatrix} \right)^{-1} \begin{bmatrix} \vec{v}_1^T \vec{u}_3 \\ \vec{v}_2^T \vec{u}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \begin{bmatrix} \frac{\langle \vec{v}_1, \vec{u}_3 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \\ \frac{\langle \vec{v}_2, \vec{u}_3 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \end{bmatrix} = \frac{\langle \vec{v}_1, \vec{u}_3 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 + \frac{\langle \vec{v}_2, \vec{u}_3 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2.$$

This implies that $\vec{v}_3 = \vec{u}_3 - \vec{p}_3 = \vec{u}_3 - \sum_{i=1}^2 \frac{\langle \vec{u}_3, \vec{v}_i \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \vec{v}_i$.

(*) The general case can be obtained by the same way of calculation.