

Recap: Column space of a matrix, rank of a matrix

Q We are interested in NFEs

$$\underline{x}^T \underline{x} \underline{\beta} = \underline{x}^T \underline{y}$$

$\text{rk}(\underline{x}^T \underline{x}) = p$ [Note that \underline{x} is an $n \times p$ matrix]

$$\text{then } (\underline{x}^T \underline{x})^{-1} \underline{x}^T \underline{y} = \hat{\underline{\beta}}$$

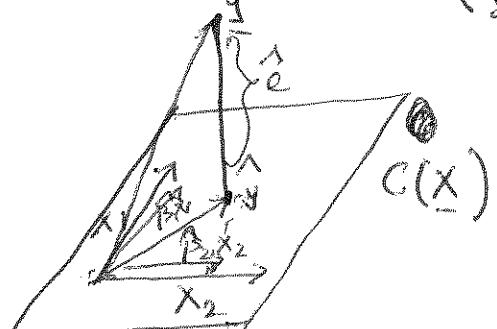
Q Our goal is to study another case when $\text{rk}(\underline{x}^T \underline{x}) < p$

Solution of $\underline{\beta}$ is not unique.

Instead of finding $\underline{\beta}$, we will find $\hat{\underline{y}} \in C(\underline{x})$ that is closest to \underline{y} in the Euclidean distance.

Once we find $\hat{\underline{y}}$ we will characterize the class of all $\hat{\underline{\beta}}$ s.t. $\hat{\underline{y}} = \underline{x} \hat{\underline{\beta}}$.

Let $n=3$, $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, there be two predictions x_1 and x_2 .



We want to find $\hat{\underline{y}}$ which is in the column space of \underline{x} and closest to \underline{y} .

We are minimizing $\|\underline{y} - \underline{x} \underline{\beta}\|^2$

Least square actually obtains $\underline{x} \underline{\beta}$ which is going to give us the $\hat{\underline{y}}$.

Let us find one such \hat{y} . Once we are done with it, we will be left with the error $\hat{e} = y - \hat{y}$. From this figure it is easy to see that \hat{e} is orthogonal to the column space of X .

note that, $\underline{X^T X \beta} - \underline{X^T y} = 0$

$$\Rightarrow \underline{X^T(y - X\beta)} = 0 \quad \text{If } \hat{\beta} \text{ is a solution}$$

then $\hat{e} = y - X\hat{\beta}$ satisfies the condition that

$$\underline{X^T \hat{e}} = 0$$

Null space of a matrix

The null space of an $m \times n$ matrix A , denoted by $N(A)$ is defined as

$$N(A) = \left\{ \underline{y} \mid A\underline{y} = \underline{0} \right\} \subset \mathbb{R}^n.$$

Ex: $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & 1 & -3 \\ 1 & 2 & 0 \end{bmatrix} \quad 4 \times 3$ Find $N(A)$.

$$A\underline{y} = \underline{0}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & 1 & -3 \\ 1 & 2 & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow y_1 + y_2 + y_3 = 0 \quad \dots \quad (1)$$

$$2y_1 + 2y_2 + 2y_3 = 0 \quad \dots \quad (2)$$

$$-y_1 + y_2 - 3y_3 = 0 \quad \dots \quad (3)$$

$$y_1 + 2y_2 = 0 \quad \dots \quad (4)$$

From (4) $y_1 = -2y_2$

Use this in ③ to obtain

$$2y_2 + 0y_2 = 3y_3 \Rightarrow y_2 = y_3$$

$$y_2 = y_3 \text{ and } y_1 = -2y_2$$

① and ② won't give any more independent equations. Any solution

should be of the form $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} -2y_2 \\ y_2 \\ y_2 \end{pmatrix}$

$$N(\underline{A}) = \{ \underline{y} \mid y_1 = -2y_2, y_3 = y_2 \} = \left\{ c \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} : c \in \mathbb{R} \right\}$$

Null space is spanned by one vector $\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$.

$$N(\underline{A}) = \{ \underline{y} \mid \underline{A}\underline{y} = \underline{0} \}$$

$\hat{\underline{e}}$ must satisfy $\underline{x}^T \hat{\underline{e}} = 0 \Rightarrow \hat{\underline{e}} \in N(\underline{x}^T)$

$$\underline{x}^T \hat{\underline{e}} = \hat{y} \in C(\underline{x})$$

Def: Two vectors $\underline{x}, \underline{y}$ are said to be orthogonal if $\underline{x}^T \underline{y} = 0$

Def: (Orthogonal space)

Two subspaces M_1 & M_2 are orthogonal spaces if $\underline{x} \in M_1$ and $\underline{y} \in M_2$ means $\underline{x}^T \underline{y} = 0$

④ Notation: Let S be a vector space and let M be a subspace of S . $M^\perp_S = \{ \underline{y} \in S \mid \underline{y} \perp M \}$ is used to denote the orthogonal ~~on~~ space of M . Sometimes orthogonal space of M is called the orthogonal complement of M . \perp is used to denote orthogonality. ③

Thus: For any matrix $\underline{A}_{m \times n}$, $C(\underline{A})$ and $N(\underline{A}^T)$ are orthogonal spaces.

Pf: $\underline{v} \in C(\underline{A})$, $\underline{w} \in N(\underline{A}^T)$

since $\underline{v} \in C(\underline{A})$, there exists some \underline{h}

$$\text{s.t. } \underline{v} = \underline{A} \underline{h}$$

$$\underline{v}^T \underline{w} = \underline{h}^T \underline{A}^T \underline{w} . \text{ Since } \underline{w} \in N(\underline{A}^T), \underline{A}^T \underline{w} = \underline{0}$$

$$\text{by definition. Thus } \underline{v}^T \underline{w} = 0$$

This implies that $C(\underline{A}) \& N(\underline{A}^T)$ are orthogonal spaces.

Result: $C(\underline{A}) \cap N(\underline{A}^T) = \{\underline{0}\}$.

Pf: Let $\underline{v} \in C(\underline{A}) \cap N(\underline{A}^T)$

there exists \underline{h} s.t. $\underline{v} = \underline{A} \underline{h} \quad \dots \quad (1)$

Also, $\underline{v} \in N(\underline{A}^T) \Rightarrow \underline{A}^T \underline{v} = \underline{0} \quad \dots \quad (2)$

\Rightarrow by (1) & (2) that $\underline{A}^T \underline{A} \underline{h} = \underline{0}$

$\Rightarrow \underline{h}^T \underline{A}^T \underline{A} \underline{h} = 0 \Rightarrow \|\underline{A} \underline{h}\|^2 = 0$

$\Rightarrow \underline{A} \underline{h} = \underline{0} \Rightarrow \underline{v} = \underline{0}$

$\Rightarrow C(\underline{A}) \cap N(\underline{A}^T) = \{\underline{0}\}$.

This result implies that ~~$C(\underline{x})$ and $N(\underline{x}^T)$~~ $C(\underline{x})$ and $N(\underline{x}^T)$ are orthogonal complements and they have no vector in common other than $\underline{0}$.

$$\hat{y} = \underline{x}\hat{\beta} \in C(\underline{x}) \text{ and } \hat{e} \in N(\underline{x}^T)$$

Since $C(\underline{x})$ is the orthogonal complement of $N(\underline{x}^T)$ $\Rightarrow \hat{y}^T \hat{e} = 0$

Thm: Let S be a vector space. Let M be a subspace of S . The orthogonal complement M^\perp

M_S^\perp is a subspace of S .

Most importantly, if $\underline{x} \in S$ then it can be written uniquely as $\underline{x} = \underline{x}_0 + \underline{x}_1$ with $\underline{x}_0 \in M$ and $\underline{x}_1 \in M_S^\perp$.

In our case $M = C(\underline{x})$ and $M_S^\perp = N(\underline{x}^T)$

$$\hat{y} \in C(\underline{x}), \hat{e} \in N(\underline{x}^T)$$

~~$$y = \hat{y} + \hat{e}$$~~

Ex: $S = \mathbb{R}^3, A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 3 \\ 1 & 0 & 3 \end{bmatrix}$

Let's find $C(A)$ and $N(A^T)$.

You can check $C(A) = \{ \text{all vectors of the form } \begin{pmatrix} a \\ b \\ b \end{pmatrix}, a, b \in \mathbb{R} \}$

~~$C(A)$~~ To find $N(A^T)$ we have to look at the system of equations

$$A^T \underline{y} = \underline{0}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 2 & 3 & 3 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow y_1 + y_2 + y_3 = 0 \quad \text{--- --- ①}$$

$$y_1 = 0 \quad \text{--- --- ②}$$

$$2y_1 + 3y_2 + 3y_3 = 0 \quad \text{--- --- ③}$$

Any solution should be of the form

$$N(\underline{A}^T) = \left\{ \begin{pmatrix} 0 \\ -b \\ b \end{pmatrix} \mid b \in \mathbb{R} \right\}$$

Let $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ be any vector.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} x \\ \frac{y+z}{2} \\ \frac{y+z}{2} \end{bmatrix} + \begin{pmatrix} 0 \\ \frac{y-z}{2} \\ -\frac{y-z}{2} \end{pmatrix}$$

clearly $\begin{pmatrix} x \\ \frac{y+z}{2} \\ \frac{y+z}{2} \end{pmatrix} \in C(\underline{A})$ and $\begin{pmatrix} 0 \\ \frac{y-z}{2} \\ -\frac{y-z}{2} \end{pmatrix} \in N(\underline{A}^T)$

the above is the unique decomposition for this case.

Message so far:

Suppose $\hat{\beta}$ is a solution to the NEs $\underline{x}^T \underline{x} \beta = \underline{x}^T \underline{y}$

$$\textcircled{1} \quad \underline{y} \in \mathbb{R}^n \quad \textcircled{2} \quad \hat{\underline{y}} = \underline{x} \hat{\beta} \in C(\underline{x})$$

$$\textcircled{3} \quad \underline{x}^T \hat{\underline{e}} = 0 \Rightarrow \hat{\underline{e}} \in N(\underline{x}^T)$$

$$\textcircled{4} \quad \hat{\underline{y}} \perp \hat{\underline{e}} \quad \textcircled{5} \quad \underline{y} = \hat{\underline{y}} + \hat{\underline{e}} \quad \text{is the unique decomposition}$$

$$\underline{y} = \hat{\underline{y}} + \hat{\underline{e}} \text{ and } \hat{\underline{y}}^T \hat{\underline{e}} = 0$$

$$\|\underline{y}\|^2 = \|\hat{\underline{y}} + \hat{\underline{e}}\|^2 = \|\hat{\underline{y}}\|^2 + \|\hat{\underline{e}}\|^2 + 2 \underbrace{\hat{\underline{y}}^T \hat{\underline{e}}}_0$$

$$= \|\hat{\underline{y}}\|^2 + \|\hat{\underline{e}}\|^2$$

$\|\underline{y}\|^2$: Total sum of squares (SST)

$\|\hat{\underline{y}}\|^2 = \|\underline{x}\hat{\beta}\|^2$: The regression sum of squares (SSR)

$\|\hat{\underline{e}}\|^2 = \|\underline{y} - \hat{\underline{y}}\|^2$: The error sum of squares (SSE).

Qn: Is there any mechanical way to find $\hat{\underline{y}}$.
Can we find a matrix \underline{M} s.t.

$$\hat{\underline{y}} = \underline{M} \underline{y}$$

\underline{M} matrix projects any vector $\underline{y} \in \mathbb{R}^n$ onto the column space of \underline{x} (remember $\hat{\underline{y}} \in C(\underline{x})$). These type of matrices are called projection matrices.

Definition (Projection matrix)

A square matrix \underline{M} is a perpendicular projection matrix onto $C(\underline{x})$ if and only if

(i) $\underline{v} \in C(\underline{x})$ implies $\underline{M} \underline{v} = \underline{v}$

(ii) $\underline{w} \perp C(\underline{x})$ implies $\underline{M} \underline{w} = \underline{0}$

Thus: $C(\underline{M}) = C(\underline{x})$

Thus: \underline{M} is projection matrix \Leftrightarrow if and only if (i) $\underline{M} \cdot \underline{M} = \underline{M}$ (idempotent) (ii) $\underline{M}^T = \underline{M}$ (symmetric)

Thus: $\textcircled{2}$ Projection matrix operator is unique.

Ex: $\underline{x} = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \quad C(\underline{x}) = \mathbb{R}^2$

Consider $M = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$

$$C(\underline{x}) = \left\{ \begin{pmatrix} 2a \\ a \end{pmatrix} : a \in \mathbb{R} \right\}$$

$$N(\underline{x}^T) = \left\{ \begin{pmatrix} -a \\ 2a \end{pmatrix} : a \in \mathbb{R} \right\}$$

$$\underline{M}^2 = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = M$$

$M^T = M$, so M is a projection matrix.

~~Let~~ $\underline{x} \in C(\underline{x})$ then $\underline{x} = \begin{pmatrix} 2a \\ a \end{pmatrix}$

$$\underline{M} \underline{x} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{pmatrix} 2a \\ a \end{pmatrix} = \begin{pmatrix} 2a \\ a \end{pmatrix} = \underline{x}$$

$\underline{w} \in N(\underline{x}^T)$ then $\underline{w} = \begin{pmatrix} -a \\ 2a \end{pmatrix}$

$$\underline{M} \begin{pmatrix} -a \\ 2a \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

\underline{M} is clearly the projection matrix onto the column space of \underline{x} .

Thm: If \underline{M} and \underline{M}_0 are projection matrices with $C(\underline{M}_0) \subset C(\underline{M})$ then $\underline{M} - \underline{M}_0$ is also a projection matrix.

Thm: \underline{M} and \underline{M}_0 are projection matrices with $C(\underline{M}_0) \subset C(\underline{M})$ then $C(\underline{M} - \underline{M}_0)$ is the orthogonal complement of $C(\underline{M}_0)$.

If $\underline{x} \in C(\underline{M})$ and $\underline{x} \perp \underline{\circlearrowleft} C(\underline{M}_0)$, then
 $\underline{x} = \underline{M}\underline{x} = \underbrace{\underline{M}_0\underline{x}}_0 + (\underline{M} - \underline{M}_0)\underline{x} = (\underline{M} - \underline{M}_0)\underline{x}$
 $\Rightarrow \underline{x} \in C(\underline{M} - \underline{M}_0)$