STA0067: Regression Analysis II	**Spring 2024**

Lecture 01: Matrix Theory

Lecturer: Kipoong Kim

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1. Matrix Theory

1.1 Basic Theories

1.1.1 Types of Matrix

(1) The $m \times n$ matrix A with the ij component a_{ij} is given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and denoted by $\mathbf{A} = (a_{ij})$

(2) The i th row vector of $\mathbf{A}$: $\mathbf{A}_i = (a_{i1}, a_{i2}, \cdots, a_{in})$

The j th column vector of $\mathbf{A}$:

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

(3) square matrix : $m = n$ case

(4) diagonal matrix : a square matrix with $a_{ij} = 0, i \neq j$, and denoted by $A = \operatorname{diag}(a_1, \cdots, a_n)$

(5) unit matrix or identity matrix : a diagonal matrix with $a_{ii} = 1, i = 1, \cdots, n$, and denoted by $\boldsymbol{I}$ or $\boldsymbol{I}_n$

(6) The transpose of an $m \times n$ matrix $\mathbf{A} = (a_{ij})$ is an $n \times m$ matrix $\mathbf{A}' = (a_{ji})$, and also denoted by $\mathbf{A}^T$ or $\mathbf{A}'$. Further, it is clear that $(\mathbf{A}\boldsymbol{B})' = \boldsymbol{B}'\mathbf{A}'$.

(7) A square matrix A is called idempotent if $A^2 = A$.

1.2 Trace

(1) The trace of a square matrix A is sum of diagonal elements of A, and denoted by $\mathrm{tr}(\mathbf{A})$. Hence, $\mathrm{tr}(\mathbf{A}) = \sum a_{ii}$.

(2) properties of trace

$$\begin{aligned}
&\mathrm{tr}(A \pm B) = \mathrm{tr}(A) \pm \mathrm{tr}(B) \\
&\mathrm{tr}(kA) = k\,\mathrm{tr}(A), k:\ \text{constant} \\
&\mathrm{tr}(AB) = \mathrm{tr}(BA),
\end{aligned}$$

where AB and BA should be defined.

1.2.1 Determinant

(1) The determinant of an $n \times n$ square matrix A is denoted by $\det(A)$ or $|\mathbf{A}|$, and defined as

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} c_{ij}, \forall i = 1, \cdots, n,$$

where

$$c_{ij} = (-1)^{i+j} d_{ij}$$

is called the cofactor of a_{ij} and d_{ij} is the determinant of $(n-1) \times (n-1)$ matrix $\mathbf{A}_{(i,j)}$ which is matrix $\mathbf{A}$ with the i th row and j th column deleted.

(2) Examples

$n = 1$ case : $|A| = a_{11}$, i.e, the value itself

$n = 2$ case :

$$\begin{aligned}
|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} &= a_{11}a_{22} - a_{12}a_{21}, \quad (i = 1 \text{ case }) \\
&= -a_{21}a_{12} + a_{22}a_{11}, \quad (i = 2 \text{ case })
\end{aligned}$$

$n = 3$ case :

$$\begin{aligned}
|A| = &\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
= &a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} \\
= &a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
= &a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} \\
&+ a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22}
\end{aligned}$$

(2) properties of determinant

$$
\begin{aligned}
&|\mathbf{A}\boldsymbol{B}| = |\boldsymbol{B}\mathbf{A}| = |\mathbf{A}||\boldsymbol{B}|, \text{ if } \mathbf{A} \text{ and } \boldsymbol{B} \text{ are } n \times n \text{ square matrix} \\
&|\mathbf{A}| = |\mathbf{A}'| \\
&|k\mathbf{A}| = k^n|\mathbf{A}|, k \text{ is constant}
\end{aligned}
$$

(3) A square matrix A is called non - singular if $|A| \neq 0$, and called singular if $|A| = 0$.

1.3 Inverse Matrix

1.3.1 Linearly Independence and Dependence

(1) Def: The linear combination of n-dimensional vector $v_1, \cdots, v_k$ and constants $c_1, \cdots, c_k$ is $\mathbf{0}$, i.e.,

$$c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$$

has at least one c_i which is nonzero, then $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$ are called linearly dependent.

Otherwise, if all the c_i' s are zero, then $v_1, \cdots, v_k$ are called linearly independent.

(2) Example : Two vectors $(1, 1)$ and $(-3, 2)$ are linearly independent because

$$c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2\begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

gives $c_1 - 3c_2 = 0, c_1 + 2c_2 = 0$, and we must have $c_1 = 0, c_2 = 0$.

(3) Remark: If $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$ are linearly dependent, $c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$ has nonzero c_i. Assume that if $c_j \neq 0$, then

$$\boldsymbol{v}_j = -\frac{1}{c_j}\left(c_1\boldsymbol{v}_1 + \cdots + c_{j-1}\boldsymbol{v}_{j-1} + c_{j+1}\boldsymbol{v}_{j+1} + \cdots + c_k\boldsymbol{v}_k\right)$$

i.e., if $v_1, \cdots, v_k$ are linearly dependent, then one vector can be expressed as a linear combination of other vectors.

1.3.2 Rank of Matrix

(1) Def : Let A be an $m \times n$ matrix with m row vectors and n column vectors. Let m^* be the maximum number of linearly independent vectors among m row vectors, and n^* be the maximum number of linearly independent vectors among n column vectors. Then, we must have $m^* = n^*$ which is called the rank of $\mathbf{A}$, and denoted by $r(\mathbf{A})$. Hence, $r(\mathbf{A}) \leq \min(m, n)$.

(1) Example : Compute the rank of 5×4 matrix $\mathbf{A}$.

$$A=\begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & -1 & 3 & 2 \\ 3 & 0 & 6 & 5 \\ 2 & 1 & 3 & 3 \\ 6 & 0 & 12 & 10 \end{bmatrix}$$

First, we must have $r(A) \leq 4$. Note that the 1 st and 2 nd row vectors are linearly independent, and the 3rd, 4th, and 5th row vectors can be expressed as a linear combination of the 1st and 2nd row vectors. For, example, the 3rd row vector can be expressed as $(1,2,0,1)+2(1,-1,3,2)$. Also, can show that the 1st and 2nd column vectors are linear independent. Therefore, $r(\mathbf{A})=2$.

(2) properties of rank

(I) $r(\boldsymbol{AB}) \leq \min(r(\mathbf{A}), r(\boldsymbol{B}))$

(II) If $n \times n$ matrix $\mathbf{A}$ is non-singular, then $r(\mathbf{A})=n$, and if it is singular, then $r(A)<n$.

(III) The rank of A does not change if it is multiplied by an non-singular matrix.

(IV) If $AGA=A$, then $r(A)=r(GA)$.

(V) $r(\mathbf{A}:\boldsymbol{B}) \leq r(\mathbf{A})+r(\boldsymbol{B})$, where $\mathbf{A}$ is $m \times n_1, \boldsymbol{B}$ is $m \times n_2$, and $\mathbf{A}:\boldsymbol{B}$ is $m \times (n_1+n_2)$ matrix, called augmented matrix. For example, when

$$A=\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \quad B=\begin{pmatrix} 7 & 9 \\ 8 & 10 \end{pmatrix}$$
$$A \vdots B=\begin{pmatrix} 1 & 3 & 5 & 7 & 9 \\ 2 & 4 & 6 & 8 & 10 \end{pmatrix}$$

(VI) $r(\mathbf{A}+\boldsymbol{B}) \leq r(\mathbf{A}:\boldsymbol{B}) \leq r(\mathbf{A})+r(\boldsymbol{B})$

(VII) If $\mathbf{A}$ is $n \times n$ matrix, then $r(\boldsymbol{AB}) \geq r(\mathbf{A})+r(\boldsymbol{B})-n$.

(VIII) If $\mathbf{A}$ is $n \times n$ idempotent matrix, then $r(\boldsymbol{I}-\mathbf{A})=n-r(\mathbf{A})$.

(IX) $r(\mathbf{A})=r\left(\mathbf{A}'\right)=r\left(\mathbf{A}'\mathbf{A}\right)=r\left(\mathbf{A}\mathbf{A}'\right)$

1.3.3 Inverse Matrix

(1) motivation : To compute a solution x of the equation $ax=b$, we multiply the inverse of a, i.e., $1/a$ on both sides, i.e.,

$$ax=b \Rightarrow \frac{1}{a}ax=\frac{1}{a}b \Rightarrow x=\frac{b}{a}$$

Here, we must have $a \neq 0$ since the existence of the inverse of a is possible only when $a \neq 0$. Now, consider n equations with n unknowns.

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \qquad\qquad & \quad \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}$$

Now, let $A = (a_{ij}), i = 1, \cdots, n; j = 1, \cdots, n$ be $n \times n$ square matrix, $\boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}', \boldsymbol{b} = \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix}'$, then the above equations can be expressed as

$$Ax = b$$

If A^{-1} satisfies

$$AA^{-1} = A^{-1}A = I$$

then it is called the inverse of A. For a scalar a to be invertible, we must have $a \neq 0$. Likewise, for a square matrix A to be invertible, we must have $|A| \neq 0$, i.e., non-singular. Hence, if A is non-singular, then $A^{-1}Ax = A^{-1}b => x = A^{-1}b$. In fact, A^{-1} is given by

$$A^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ \vdots & \vdots & & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix}$$

where c_{ij} is the cofactor (see A.1.3) of a_{ij}.

(2) properties

(I) A^{-1} is unique.

(II) A^{-1} is non-singular.

(III) $(A')^{-1} = (A^{-1})'$

(IV) $(AB)^{-1} = B^{-1}A^{-1}$

(3) orthogonal matrix

If a square matrix $\mathbf{A}$ satisfies $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \boldsymbol{I}$, then it is called orthogonal matrix, and it has the following properties.

(I) $\operatorname{tr}(\mathbf{A}'\boldsymbol{B}\mathbf{A}) = \operatorname{tr}(\boldsymbol{B})$

(II) $|A'BA| = |B|$

(III) $|A| = \pm 1$

1.3.4 Inverse of Special Matrices

(I) Let J be an $n \times n$ square matrix with all the components are 1 , then if $a \neq 0, a + nb \neq 0$, we have

$$(a\mathbf{I}+b\mathbf{J})^{-1}=\frac{1}{a}\left(\boldsymbol{I}-\frac{b}{a+nb}\boldsymbol{J}\right)$$

(II) $\{\operatorname{diag}(a_1,\cdots,a_n)\}^{-1}=\operatorname{diag}(1/a_1,\cdots,1/a_n)\,,a_i\neq 0,i=1,\cdots,n$

(III) $\boldsymbol{I}+\mathbf{A}+A^2+\cdots+A^{n-1}=(A^n-\boldsymbol{I})\,(\mathbf{A}-\boldsymbol{I})^{-1}$

(IV) $\left(I+A^{-1}\right)^{-1}=A(A+I)^{-1}$

(V) $(I+AB)^{-1}=I-A(I+BA)^{-1}B$

(VI) $(A+UBV)^{-1}=A^{-1}-A^{-1}UBV\left(I+A^{-1}UBV\right)^{-1}A^{-1}$

1.3.5 Generalized Inverse

Here we define the inverse of A when A is non-singular, and we also define the inverse of $\mathbf{A}$ when $\mathbf{A}$ is not a square matrix.

(1) Moore-Penrose inverse

Let A be $p\times q$ matrix, then the Moore - Penrose inverse of A is a $q\times p$ matrix M satisfying the following 3 conditions;

(I) $AMA=A$

(II) $MAM=M$

(III) AM and MA are symmetric

Note that the Moore-Penrose inverse M is unique. For example, the MoorePenrose inverse of for 3×4 matrix

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

is

$$M=\frac{1}{18}\begin{bmatrix} 5 & 2 & -1\\ 1 & 1 & 1\\ -4 & -1 & 2\\ 6 & 3 & 0 \end{bmatrix}$$

Also, note that if $\mathbf{A}$ is a square matrix, then $\boldsymbol{M}$ is also a square matrix and $M=A^{-1}$.

(2) generalized inverse

Among the above 3 conditions, if G satisfies the 1st condition only, i.e., $AGA=A$, then G is called the generalized inverse of A, and denoted by A^{-}. In fact, G is not unique. Here, we introduce one popular method of computing G. First, find a square and non-singular submatrix A_{11} of $A_{p\times q}$, and let all other components be $\mathbf{0}$. Finally, find $\mathbf{A}_{11}^{-1}$, i.e.,

$$A_{p\times q} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \Rightarrow G_{q\times p} = \begin{bmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$
$$-10-$$

For example, consider

$$A = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix}$$

and let

$$A_{11} = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}, A_{12} = \begin{bmatrix} 4 & 3 \\ 2 & -2 \end{bmatrix}, A_{21} = \begin{bmatrix} 5 & -4 \end{bmatrix}, A_{22} = \begin{bmatrix} 0 & -7 \end{bmatrix}$$

then

$$G = \frac{1}{7}\begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

As an another example, consider a oneway classification

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1,2, j = 1,2,3$$

i.e.,

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{bmatrix}$$

In matrix notation,

$$y = X\beta + \varepsilon$$

then the normal equation for the least squares method becomes

$$(X'X)\,\beta = X'y$$

however,

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix}$$

is singular, so that $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ does not exist. Hence, as an estimator of $\boldsymbol{\beta}$, we use the generalized inverse $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-}$, and the estimator becomes $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-}\boldsymbol{X}'\boldsymbol{y}$. One possible method is given by

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

(3) computational issues in inverse matrix

Note that

$$A = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix}$$

is singular since $|\mathbf{A}| = 0$, and therefore the inverse does not exist. Now,

$$B = \begin{bmatrix} 1.9998 & 0.9999 \\ 5.9994 & 3.0009 \end{bmatrix}$$

is non-singular since its determinant is 0.0024 . Note that even though $\boldsymbol{B}$ is mathematically non-singular, it is almost singular since its determinant is close to 0 . Now, consider

$$\begin{bmatrix} 2.0 & 2.5 \\ 2.5 & 3.0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

then the solution is $(0.1, 0.1)'$, however, the solution of

$$\begin{bmatrix} 2.04 & 2.49 \\ 2.49 & 3.04 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

is $(-1, 1)'$, and the determinant is very close to 0 (in fact, 0.0015). In this case, a small change in components may result in a big change in solution. A matrix with a small value of determinant is called ill - conditioned matrix, and it can be measured by the condition number which is defined as a ratio of the largest singular value (see A.6.2 for definition) to the smallest singular value of the $n \times p$ matrix A. If a condition number is large, then the matrix can be ill-conditioned.

1.4 Partitioned Matrix

Consider partitioning the matrix $\boldsymbol{P}$.

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

Then, by the following identity,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

we have

$$\begin{aligned} \begin{vmatrix} \mathbf{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{vmatrix} &= |\mathbf{A}| \left|\boldsymbol{D} - \boldsymbol{C}\mathbf{A}^{-1}\boldsymbol{B}\right|, \text{ if } \mathbf{A}^{-1} \text{ exists} \\ &= |\boldsymbol{D}| \left|\mathbf{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right|, \text{ if } \boldsymbol{D}^{-1} \text{ exists} \end{aligned}$$

Also, the inverse of $\boldsymbol{P}$ is

$$\begin{aligned} &\begin{bmatrix} \mathbf{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} \\ =&\begin{bmatrix} \boldsymbol{I} & -\mathbf{A}^{-1}\boldsymbol{B} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\left(\boldsymbol{D} - \boldsymbol{C}\mathbf{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\mathbf{A}^{-1} & \left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{bmatrix}, \text{ if } \mathbf{A}^{-1} \text{ exists} \\ =&\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}^{-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I} \\ -\boldsymbol{D}^{-1}\boldsymbol{C} \end{bmatrix} \left(\mathbf{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} \left[\boldsymbol{I} - \boldsymbol{B}\boldsymbol{D}^{-1}\right], \text{ if } \boldsymbol{D}^{-1} \text{ exists} \end{aligned}$$

For example, if

$$\boldsymbol{P} = \begin{bmatrix} 2 & 0 & 0 & 1 & -1 \\ 0 & 2 & 0 & 1 & -1 \\ 0 & 0 & 2 & 1 & -1 \\ 2 & 2 & 2 & 1 & 0 \\ -1 & -1 & -1 & 0 & 1 \end{bmatrix}$$

then let

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix},$$
$$\boldsymbol{C} = \begin{bmatrix} 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and we have

$$\begin{aligned} &A^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &\boldsymbol{D} - \boldsymbol{C}\mathbf{A}^{-1}\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} \cdot \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 3 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \end{aligned}$$

so that

$$|\boldsymbol{P}| = 8\left(1 - \frac{9}{2}\right) = -28$$

$$\boldsymbol{P}^{-1} = \frac{1}{14}\begin{bmatrix} 4 & -3 & -3 & 2 & -2 \\ -3 & 4 & -3 & 2 & -2 \\ -3 & -3 & 4 & 2 & -2 \\ 4 & 4 & 4 & 2 & 12 \\ -2 & -2 & -2 & 6 & 8 \end{bmatrix}$$

1.5 Eigenvalues and Eigenvectors

1.5.1 Eigenvalues and Eigenvectors

For a square matrix $\mathbf{A}$, a vector $\boldsymbol{u}$, and a constant λ, consider

$$Au = \lambda u$$

which is equivalent to

$$(\mathbf{A} - \lambda \boldsymbol{I})\boldsymbol{u} = \mathbf{0}$$

If $\mathbf{A} - \lambda \boldsymbol{I}$ is non-singular, then the solution for $\boldsymbol{u}$ is $\boldsymbol{u} = \mathbf{0}$. But, if $\mathbf{A} - \lambda \boldsymbol{I}$ is singular, then there exist nonzero solution for u and a constant λ, i.e.,

$$|A - \lambda I| = \mathbf{0}$$

gives nonzero solution for u and a constant λ. The above equation is called the characteristic equation for A, and it is the nth degree polynomial in λ since A is an $n \times n$. Let the n solutions be $\lambda_1, \cdots, \lambda_n$, then these are called the eigenvalue, characteristic roots, or latent roots. Further, for each λ_i, a vector $\boldsymbol{u}_i$ satisfying

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \quad i = 1, \cdots, n$$

is called the eigenvector corresponding to λ_i.

For example, if

$$A = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

then the characteristic equation is

$$\begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0$$

and eiegnvalues are $\lambda = -5$ or $\lambda = 7$. Also, we have

$$A\begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5\begin{bmatrix} 2 \\ -3 \end{bmatrix}$$
$$A\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

and hence, the corresponding eigenvectors for $\lambda = -5$ and $\lambda = 7$ are

$$\begin{bmatrix} 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

respectively.

1.5.2 Properties of Eigenvalues

(I) If λ is an eigenvalue of $\mathbf{A}$, then

(a) Eigenvalue of A^k is λ^k.

(b) Eigenvalue of A^{-1} is $1/\lambda$.

(c) Eigenvalue of A is $c\lambda$.

(d) Eigenvalue of $\mathbf{A} + c\boldsymbol{I}$ is $\lambda + c$.

(e) Eigenvalue of $(\mathbf{A} + c\boldsymbol{I})^{-1}$ is $1/(\lambda + c)$.

(II) $\operatorname{tr}(A) = \sum \lambda_i, |A| = \prod \lambda_i$

(III) If $\mathbf{A}$ is symmetric and its components are real values, then

(a) All the eigenvalues are real.

(b) Eigenvectors are orthogonal.

(c) The rank of A is the number of nonzero eigenvalues.

(IV) Eigenvalues of idempotent matrix is either 0 or 1 , but the converse is not true.

1.6 Quadratic Forms and Positive Definite Matrix

1.6.1 Quadratic Forms

For vector $\boldsymbol{x}$ and matrix $\mathbf{A}, \boldsymbol{x}'\mathbf{A}\boldsymbol{x}$ is called the quadratic form in $\boldsymbol{x}$. For example, if $\boldsymbol{x} = (x_1 x_2 x_3)'$ and

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix}$$

then

$$x'Ax = x_1^2 + 7x_2^2 + 5x_3^2 + (4+2)x_1x_2 + (2+3)x_1x_3 + (-2+6)x_2x_3$$

which is a quadratic function in x_1, x_2, x_3. In general, $x = (x_1, \cdots, x_n)'$ and $A_{n\times n} = (a_{ij})$, then

$$\begin{aligned}\boldsymbol{x}'\mathbf{A}\boldsymbol{x} &= \sum a_{ii}x_i^2 + \sum_{i\neq j}\sum a_{ij}x_ix_j \\ &= \sum a_{ii}x_i^2 + \sum_{i<j}\sum (a_{ij}+a_{ji})\, x_ix_j\end{aligned}$$

Since the coefficient of x_ix_j is the sum of a_{ij} and a_{ji}, the corresponding matrix A is not unique, i.e.,

$$\begin{aligned}&x_1^2 + 7x_2^2 + 5x_3^2 + (4+2)x_1x_2 + (2+3)x_1x_3 + (-2+6)x_2x_3 \\ =&x_1^2 + 7x_2^2 + 5x_3^2 + (3+3)x_1x_2 + (3+2)x_1x_3 + (0+4)x_2x_3\end{aligned}$$

Hence, if we take

$$B = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 7 & 0 \\ 2 & 4 & 5 \end{bmatrix}$$

then $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$. Therefore, when we define a quadratic form, we often assume A is symmetric. Then, the matrix corresponding to a certain quadratic form becomes unique. A symmetric matrix corresponding to the above quadratic form is

$$A = \begin{bmatrix} 1 & 3 & 5/2 \\ 3 & 7 & 2 \\ 5/2 & 2 & 5 \end{bmatrix}$$

and it is unique.

1.6.2 Positive Definite Matrix

(1) definition

It is clear that a scalar is either positive or negative, however, it is not possible to define a matrix is positive or negative. Hence, the sign of a matrix is defined using the quadratic form. Consider a matrix

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

and the corresponding quadratic form is

$$\begin{aligned}x'Ax &= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3 \\ &= (x_1+2x_2)^2 + (x_1+x_3)^2 + (x_2+x_3)^2\end{aligned}$$

is always positive except that all the x_1, x_2, x_3 are zero. In this case, we call A is positive definite. In general, A is called positive definite (p.d.) if $x'Ax > 0, \forall x \neq 0$.

On the other hand, consider

$$A = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}$$

and its quadratic form is

$$\begin{aligned} x'\mathbf{A}\boldsymbol{x} &= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_1x_3 - 6x_2x_3 \\ &= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2 \end{aligned}$$

and which can be 0 even though $\boldsymbol{x} \neq \mathbf{0}$ because $\boldsymbol{x} = (213)'$ gives $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = 0$.

This matrix is called positive semidefinite (p.s.d.) matrix. In general,

A is called positive semidefinite (p.s.d.) if $x'Ax \geq 0, \forall x$ and $x'Ax = 0$, for some $\boldsymbol{x}$.

Positive definite and positive semidefinite matrix is often called nonnegative definite (n.n.d.). In this sense, negative de finite (n.d.), negative semide finite (n.s.d.), and nonpositive definite (n.p.d.) can be defined. Further, a matrix A is called indefinite if A cannot be classified into any kind of definiteness.

(2) properties

(I) If $A_{n\times n} = (a_{ij})$ is p.d., then

(a) $r(A) = n$ (b) $a_{ii} > 0, \quad i = 1, \cdots, n$

(c) $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$ is p.d. for any $n \times n$ square matrix $\boldsymbol{P}$.

(II) If $\mathbf{A}_{n\times n} = (a_{ij})$ is p.s.d., then

(a) $r(\mathbf{A}) < n$

(b) $a_{ii} \geq 0, \quad i = 1, \cdots, n$

(c) $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$ is n.n.d. for any $n \times n$ square matrix $\boldsymbol{P}$.

(III) The necessary and sufficient condition for a symmetric matrix $A_{n\times n}$ to be p.d. is

(a) There exists a full rank matrix $\boldsymbol{B}_{n\times n}$ s.t. $\boldsymbol{B}'\boldsymbol{B} = \mathbf{A}$

(b) All the eigenvalues of A are positive

(c) $a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \cdots, |\mathbf{A}| > 0$

(IV) The necessary and sufficient condition for a symmetric matrix $A_{n\times n}$ to be p.s.d. is

(a) There exists a matrix $\boldsymbol{B}_{n\times n}$ s.t. $\boldsymbol{B}'\boldsymbol{B} = \mathbf{A}$ with $r(\mathbf{A}) < n$

(b) All the eigenvalues of $\mathbf{A}$ are greater than 0 or equal to 0 , and at least one eigenvalues should be 0

(V) If the rank of $\mathbf{A}_{m\times n}$ is $m (m < n)$, then

(a) $A'A$ is p.s.d.

(b) AA' is n.n.d.

(VI) If the rank of $\mathbf{A}_{m\times n}$ is $r(r<m, r<n)$, then

(a) $A'A$ is p.s.d.

(b) AA' is n.n.d.

1.7 Projection and Decomposition of Matrix

1.7.1 Projection

(I) projection on vectors

If we project y onto a vector x, then it becomes cx, where $c = \frac{x'y}{x'x}$.

(II) projection on column space

Let $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{p-1})$ be $n \times p$ matrix, where $\boldsymbol{x}_i, i = 1, \cdots, p-1$ is n-vectors. Then, the column space of $\boldsymbol{X}$ is defined as follows;

$$\begin{aligned}
\mathcal{C}_{\boldsymbol{X}} &\equiv \operatorname{span}\{\mathbf{1}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{p-1}\} \\
&= \{\beta_0 \mathbf{1} + \beta_1 x_1 + \cdots, +\beta_{p-1} x_{p-1} \mid \beta_0, \cdots, \beta_{p-1} \in R\} \\
&= \{\boldsymbol{X}\boldsymbol{\beta} \mid \boldsymbol{\beta} \in R^p\}
\end{aligned}$$

If we project vector y onto the column space $\mathcal{C}_{\boldsymbol{X}}$, then it becomes $\boldsymbol{H}\boldsymbol{y}$, where $\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$, and it is called a projection matrix.

(III) Gram-Schmidt orthogonalization

Let $\boldsymbol{X} = (x_0, x_1, \cdots, x_{p-1})$ be an $n \times p$ matrix with $x_0 = \mathbf{1}$. Also, let $\Pi(x \mid z)$ be the projection of vector x onto a vector z, i.e.,

$$\Pi(x \mid z) = (z'x/z'z)\, z$$

Transform $\boldsymbol{X} = (x_0, x_1, \cdots, x_{p-1})$ to $\boldsymbol{Z} = (z_0, z_1, \cdots, z_{p-1})$ s.t.

$$\begin{aligned}
z_0 &= x_0 \\
z_1 &= x_1 - \Pi(x_1 \mid z_0) \\
z_2 &= x_2 - \Pi(x_2 \mid z_0) - \Pi(x_2 \mid z_1) \\
&\vdots \\
z_{p-1} &= x_{p-1} - \Pi(x_{p-1} \mid z_0) - \Pi(x_{p-1} \mid z_1) - \cdots - \Pi(x_{p-1} \mid z_{p-2})
\end{aligned}$$

Then, we have $\mathcal{C}_{\boldsymbol{X}} = \mathcal{C}_{\boldsymbol{Z}}$, and note that $z_0, z_1, \cdots, z_{p-1}$ are orthogonal to each other, and this transformation is known as Gram Schmidt orthogonalization.

1.7.2 Decomposition of Matrix

(I) QR Decomposition

Def. An $n \times p$ matrix $\boldsymbol{X}$ can be written as a form of $\boldsymbol{X} = \boldsymbol{QR}$, where Q is $n \times p$ orthogonal matrix and $\boldsymbol{R}$ is $p \times p$ upper triangular matrix. This is called QR decomposition of $\boldsymbol{X}$.

There are 3 methods of computing QR decomposition; (i) GramSchmidt process, (ii) Householder transformation, (iii) Givens rotation. Here we introduce the Gram-Schmidt process. Assume that $\boldsymbol{X} = (x_0, x_1, \cdots, x_{p-1})$ is transformed to $\boldsymbol{Z} = (z_0, z_1, \cdots, z_{p-1})$ by the Gram-Schmidt orthogonalization, and define

$$\gamma_{ij} = \begin{cases} \boldsymbol{x}_j'\boldsymbol{z}_i/\boldsymbol{z}_i'\boldsymbol{z}_i & , i < j \\ 1 & , i = j \\ 0 & , i > j \end{cases}$$

Now, let $\boldsymbol{\Gamma} = (\gamma_{ij})$, then it is $p \times p$ upper triangular matrix, and we have $\boldsymbol{X} = \mathbf{Z\Gamma}$. Therefore, $\boldsymbol{X} = QR$, where $Q = Z$ and $Q = \Gamma$. QR decomposition is very useful when we compute the inverse of $\boldsymbol{X}'\boldsymbol{X}$ because $X'X = R'Q'QR = R'R$ and it is very easy to compute the inverse of $\boldsymbol{R}'\boldsymbol{R}$.

(II) Cholesky Decomposition

For a symmetric and p.d. matrix A, there exists an upper triangular matrix $\boldsymbol{R}$ s.t. $\mathbf{A} = \boldsymbol{R}'\boldsymbol{R}$, and this result is called the Cholesky decomposition.

The computation of $\boldsymbol{R}$ for a given $\mathbf{A}$ is based on the following algorithm.

STEP 1. $r_{11} = a_{11}^{1/2}, r_{ij} = a_{1j}/r_{11}, j = 2, \cdots, p$

STEP 2. For $2 \leq i \leq p$,

$$\begin{aligned} r_{ii} &= \left(a_{ii} - \sum_{k=1}^{i-1} r_{ki}^2\right)^{1/2}, \\ r_{ij} &= \left(a_{ij} - \sum_{k=1}^{i-1} r_{ki}r_{kj}\right)/r_{ii}, \quad i+1 \leq j \leq p \end{aligned}$$

(III) Spectral Decomposition

Any $n \times n$ symmetric matrix A can be written as

$$A = \Gamma D \Gamma'$$

where $\boldsymbol{\Gamma} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ is $n \times n$ orthogonal matrix and $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$ is a diagonal matrix. Here, $\lambda_i, i = 1, \cdots, n$ are eigenvalues of A, and $\boldsymbol{u}_i$ is an eigenvector corresponding to λ_i. This decomposition is called spectral decomposition (or eigen decomposition). Further, note that we may write

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i'$$

(IV) Singular Values Decomposition

Any $n \times p(p < n)$ matrix $\boldsymbol{X}$ can be written as

$$X = USV'$$

where

(i) U is $n \times p$ orthogonal matrix consisting of eigenvectors corresponding to p largest eigenvalues among n eigenvalues of $\boldsymbol{X}\boldsymbol{X}'$

(ii) $S = \operatorname{diag}(s_1, \cdots, s_p), s_1 \geq \cdots \geq s_p \geq 0$ is $p \times p$ diagonal matrix, where $s_i, i = 1, \cdots, p$ is called singular values of X. Recall that the singular values of $\boldsymbol{X}$ is the positive square root of the eigenvalues of $X'X$ because

$$X'X = VSU'USV' = VS^2V'$$

and we see that the eigenvalues of $X'\boldsymbol{X}$ are diagonal elements of $\boldsymbol{S}^2$ by the spectral decomposition. (iii) $\boldsymbol{V}$ is $p \times p$ orthogonal matrix consisting of the eigenvectors of $\boldsymbol{X}'\boldsymbol{X}$.

This decomposition is called Singular Values Decomposition (SVD) of $n \times p(p < n)$ matrix $\boldsymbol{X}$.

1.8 Miscellanea in Matrix

1.8.1 Summing Vector and Centering Matrix

$\mathbf{1}_n = (1, 1, \cdots, 1)'$ is called a summing vector because for an n-dimensional vector $\boldsymbol{x} = (x_1, \cdots, x_n)'$, we have $\mathbf{1}'\boldsymbol{x} = \sum x_i$. Also, $\mathbf{1}_r\mathbf{1}_s'$ is $r \times s$ matrix with all the components 1, and often denoted by $\boldsymbol{J}_{r\times s}$. Further, we denote $\boldsymbol{J}_{n\times n}$ as $\boldsymbol{J}_n$, and it is easy to show $\boldsymbol{J}_n^2 = n\boldsymbol{J}_n$. Especially,

$$\boldsymbol{C} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{J}_n$$

is called a centering matrix, and we have

$$C = C' = C^2, C1 = 0, CJ = JC = 0$$

As an example for a centering matrix, consider a sample variance s^2 for $\boldsymbol{x} = (x_1, \cdots, x_n)'$.

$$\begin{aligned}(n-1)s^2 &= \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 = x'x - n\left(\frac{1}{n}\mathbf{1}'\boldsymbol{x}\right)^2 \\ &= x'\boldsymbol{x} - \frac{1}{n}\boldsymbol{x}'\mathbf{1}\mathbf{1}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x} - \frac{1}{n}\boldsymbol{x}'\boldsymbol{J}\boldsymbol{x} = \boldsymbol{x}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{C}\boldsymbol{x}\end{aligned}$$

– 27

1.8.2 Derivatives of Matrix

(i) If x, y are n-dimensional vectors, then

$$\frac{\partial}{\partial x}(x'y) = \frac{\partial}{\partial x}(y'x) = y$$

(ii) If x is n-dimensional vector and A is $n \times n$ matrix, then

$$\frac{\partial}{\partial x}(x'A) = A$$
$$\frac{\partial}{\partial x}(Ax) = A'$$

(iii) The derivative of the quadratic form of the n-dimensional vector x becomes

$$\begin{aligned}\frac{\partial}{\partial x}(x'Ax) &= Ax + A'x \\ &= 2Ax, \text{ if } A \text{ is symmetric}\end{aligned}$$

(iv) The 2nd derivative of $f(x) = f(x_1, \cdots, x_n)$ w.r.t. x_i and x_j is

$$\boldsymbol{H} = \left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\} = \frac{\partial^2 f}{\partial x \partial x'}$$

which is $n \times n$ matrix and called the Hessian matrix.

1.8.3 Kronecker Product

(1) definition

The Kronecker product or direct product of $\mathbf{A}_{p\times q}$ and $\boldsymbol{B}_{m\times n}$ is defined as

$$\mathbf{A}_{p\times q} \otimes \boldsymbol{B}_{m\times n} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1q}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{p1}\boldsymbol{B} & \cdots & a_{pq}\boldsymbol{B} \end{bmatrix}$$

and which is $pm \times qn$ matrix.

(2) properties

(I) $(\mathbf{A} \otimes \boldsymbol{B})' = \mathbf{A}' \otimes \boldsymbol{B}'$

(II) $(\mathbf{A} \otimes \boldsymbol{B})(\boldsymbol{X} \otimes \boldsymbol{Y}) = \mathbf{A}\boldsymbol{X} \otimes \boldsymbol{B}\boldsymbol{Y}$

(III) $(\mathbf{A} \otimes \boldsymbol{B})^{-1} = \mathbf{A}^{-1} \otimes \boldsymbol{B}^{-1}$,where $\mathbf{A}$ and $\boldsymbol{B}$ are square matrices

(IV) $r(\mathbf{A} \otimes \boldsymbol{B}) = r(\mathbf{A})r(\boldsymbol{B})$

(V) $\mathrm{tr}(\mathbf{A} \otimes \boldsymbol{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\boldsymbol{B})$

(VI) $|\mathbf{A}_{p\times p} \otimes \boldsymbol{B}_{m\times m}| = |\mathbf{A}|^m |\boldsymbol{B}|^p$

(VII) Eigenvalues of $\mathbf{A} \otimes \boldsymbol{B}$ is the product of the eigenvalue of $\mathbf{A}$ and the eigenvalue of $\boldsymbol{B}$

1.8.4 Vectorization

(1) definition

When we write a matrix $\mathbf{A}_{m\times n}$ as $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_n]$, where $\mathbf{A}_i$ is m-dimensional i th column vector, then vec($\mathbf{A}$) is defined as

$$\operatorname{vec} A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

i.e., vec $\mathbf{A}$ is mn-dimensional vector. For example,

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}, \quad \operatorname{vec}(A) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

(2) properties

(I) $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$

(II) $\operatorname{tr}(\boldsymbol{AB}) = (\operatorname{vec} \mathbf{A}')' \operatorname{vec} \boldsymbol{B}$

(III) $\operatorname{tr}\left(\mathbf{A}\boldsymbol{Z}'\boldsymbol{B}\boldsymbol{Z}\boldsymbol{C}\right) = \left(\operatorname{vec} \boldsymbol{Z}'\right)' \left(\boldsymbol{C}\mathbf{A} \otimes \boldsymbol{B}'\right) \operatorname{vec} \boldsymbol{Z}$