STAT 410 Linear Regression
Lab Session 2 / Sep 12, 2017 / Handout

This lab is intended to briefly review the linear algebra part related to our course. We review some basic definitions and their important properties in linear algebra.

1. Column Space.

If $\boldsymbol{A}$ is an $m \times n$ matrix with real entries, the column space of $\boldsymbol{A}$ is the subspace of $\mathbb{R}^m$ spanned by its columns. Similarly, the row space of $\boldsymbol{A}$ is the subspace of $\mathbb{R}^n$ spanned by its rows.

Example 1: Suppose

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 8 \\ 0 & 1 & 2 & 9 \end{bmatrix}$$

Then the column space of $\boldsymbol{A}$ is the subspace of $\mathbb{R}^2$ spanned by columns $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 8 \\ 9 \end{bmatrix}$. However, $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 8 \\ 9 \end{bmatrix}$ can be expressed as linear combinations of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Therefore, we can also say that the column space of $\boldsymbol{A}$ is the subspace of $\mathbb{R}^2$ spanned by columns $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

2. Rank of a Matrix.

The rank of a $n \times p$ matrix is defined to be the maximum number of linearly independent columns, or equivalently, of independent rows, in the matrix. It is a unique value, with the maximum cannot exceed $\min(n, p)$.

When a matrix is the product of two matrices, say $\boldsymbol{C} = \boldsymbol{AB}$, its rank can't exceed the smaller of the two ranks for the matrices being multiplied, i.e., $\text{rank}(\boldsymbol{C}) \leq \min\left(\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B})\right)$.

Exercise 1: Suppose

$$\boldsymbol{A} = \begin{bmatrix} 4 & 7 & 1 & 8 \\ 3 & 5 & 2 & 9 \end{bmatrix}$$

Find the row rank of $\boldsymbol{A}$ and the column rank of $\boldsymbol{A}$.

3. Identity Matrix.

The identity matrix of order k, denoted by $\boldsymbol{I}$ or $\boldsymbol{I_k}$, is a $k \times k$ square matrix whose diagonal elements are 1's and whose nondiagonal elements are 0's.

4. Inverse of a Matrix.

An inverse of a square $n \times n$ matrix exists if the rank of the matrix is n. Such a matrix is said to be nonsingular or of full rank. An $n \times n$ matrix with rank less than n is said to be singular or not of full rank, and does not have an inverse. The inverse of an $n \times n$ matrix of full rank also has rank n.

Let $\boldsymbol{A}$ be a $k \times k$ matrix. The inverse of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{-1}$, is another $k \times k$ matrix such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

If the inverse exists, it is unique.

You can show the following:

For a 2×2matrix $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^{-1} = \begin{bmatrix} d/(ad-bc) & -b/(ad-bc) \\ -c/(ad-bc) & a/(ad-bc) \end{bmatrix}$$

.

5. Transpose of a Matrix.

Let $\boldsymbol{A}$ be an $n \times k$ matrix. The transpose of $\boldsymbol{A}$, denoted by $\boldsymbol{A'}$ or $\boldsymbol{A}^T$, is a $k \times n$ matrix whose columns are the rows of $\boldsymbol{A}$. For example, if

$$\boldsymbol{A} = \begin{bmatrix} 5 & -7 \\ 3 & 8 \\ -4 & 2 \end{bmatrix},$$

then its transpose $\boldsymbol{A'}$ is $\begin{bmatrix} 5 & 3 & -4 \\ -7 & 8 & 2 \end{bmatrix}$.

Note: If $\boldsymbol{A}$ is an $n \times m$ matrix and $\boldsymbol{B}$ is an $m \times p$ matrix, then

$$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$$

Exercise 2: Try to show that $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$. (Hint: show that the kj-th element of $\boldsymbol{AB}$ equals the kj-th element of $\boldsymbol{B'A'}$).

6. Symmetric Matrix.

Let $\boldsymbol{A}$ be a $k \times k$ matrix. $\boldsymbol{A}$ is said to be symmetric if $\boldsymbol{A} = \boldsymbol{A'}$.

7. Idempotent Matrix.

Let $\boldsymbol{A}$ be a $k \times k$ matrix. $\boldsymbol{A}$ is called idempotent if

$$\boldsymbol{A} = \boldsymbol{AA}$$

If $\boldsymbol{A}$ is also symmetric, then $\boldsymbol{A}$ is called symmetric idempotent. If $\boldsymbol{A}$ is symmetric idempotent, then $\boldsymbol{I} - \boldsymbol{A}$ is also symmetric idempotent (check it!).

8. Orthonormal Matrix.

Let $\boldsymbol{A}$ be a $k \times k$ matrix. If $\boldsymbol{A}$ is an orthonormal matrix, then $\boldsymbol{A'A} = \boldsymbol{I}$. As a consequence, if $\boldsymbol{A}$ is an orthonormal matrix, then $\boldsymbol{A}^{-1} = \boldsymbol{A}'$.

9. Quadratic Form.

Let $\boldsymbol{y}$ be a $k \times 1$ vector, and let $\boldsymbol{A}$ be a $k \times k$ matrix. The function

$$\boldsymbol{y'Ay} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} y_i y_j$$

is called a quadratic form. $\boldsymbol{A}$ is called the matrix of the quadratic form.

10. Positive Definite and Positive Semi-definite Matrices.

Let $\boldsymbol{A}$ be a $k \times k$ matrix. $\boldsymbol{A}$ is said to be positive definite if $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} > 0, \forall \boldsymbol{y} \neq \boldsymbol{0}$. If $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \geq 0, \forall \boldsymbol{y} \neq \boldsymbol{0}$, then $\boldsymbol{A}$ is said to be semi-positive definite.

Example 2: The covariance matrix between two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, is defined as $\text{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = E[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{Y} - E[\boldsymbol{Y}])']$. The variance-covariance matrix of the random vector $\boldsymbol{Y}$, denoted by $\text{Cov}(\boldsymbol{Y})$, is defined as $\text{Cov}(\boldsymbol{Y}) = \text{Cov}(\boldsymbol{Y}, \boldsymbol{Y})$.

Let $\boldsymbol{Y}' = [Y_1, Y_2, \cdots, Y_n]$ be a random vector and $\boldsymbol{a}' = [a_1, a_2, \cdots, a_n]$ be a constant vector. Let $Z = \boldsymbol{a}'\boldsymbol{Y}$ be a scalar random variable. Since $\text{Var}(Z) = \boldsymbol{a}'\text{Cov}(\boldsymbol{Y})\boldsymbol{a} \geq 0$, the covariance matrix of any random vector $\boldsymbol{Y}$, $\text{Cov}(\boldsymbol{Y}, \boldsymbol{Y})$, is a positive semi-definite matrix.

11. Trace of a Matrix.

Let $\boldsymbol{A}$ be a $k \times k$ matrix. The trace of $\boldsymbol{A}$, denoted by trace($\boldsymbol{A}$) or tr($\boldsymbol{A}$), is the sum of the diagonal elements of $\boldsymbol{A}$; thus,

$$\text{trace}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}$$

. There are some important properties related to traces:

- If $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{B}$ is an $n \times m$ matrix, then
$$\text{trace}(\boldsymbol{AB}) = \text{trace}(\boldsymbol{BA})$$
.
- If the matrices are appropriately conformable, then
$$\text{trace}(\boldsymbol{ABC}) = \text{trace}(\boldsymbol{CAB})$$
.
- If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $k \times k$ matrices and a and b are scalars, then
$$\text{trace}(a\boldsymbol{A} + b\boldsymbol{B}) = a * \text{trace}(\boldsymbol{A}) + b * \text{trace}(\boldsymbol{B})$$
.

12. Rank of an Idempotent Matrix.

Let $\boldsymbol{A}$ be an idempotent matrix. The rank of $\boldsymbol{A}$ is its trace.

13. An Important Identity for a Partitioned Matrix.

Let $\boldsymbol{X}$ be an $n \times p$ matrix partitioned such that

$$\boldsymbol{X} = [\boldsymbol{X_1} \quad \boldsymbol{X_2}]$$

.

We note that

$$\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'X} = \boldsymbol{X}$$
$$\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}[\boldsymbol{X_1} \quad \boldsymbol{X_2}] = [\boldsymbol{X_1} \quad \boldsymbol{X_2}]$$

Therefore

$$\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'X_1} = \boldsymbol{X_1}$$
$$\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'X_2} = \boldsymbol{X_2}$$

14. Matrix Derivative.

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ vector of variables.

1. If $z = \boldsymbol{a}'\boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial \boldsymbol{a}'\boldsymbol{y}}{\partial \boldsymbol{y}} = \boldsymbol{a}$$
2. If $z = \boldsymbol{y}'\boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial \boldsymbol{y}'\boldsymbol{y}}{\partial \boldsymbol{y}} = 2\boldsymbol{y}$$
3. If $z = \boldsymbol{a}'\boldsymbol{A}\boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial \boldsymbol{a}'\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{y}} = \boldsymbol{A}'\boldsymbol{a}$$
4. If $z = \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$, then
$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{A}'\boldsymbol{y}$$
   If $\boldsymbol{A}$ is symmetric, then
$$\frac{\partial \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{y}} = 2\boldsymbol{A}\boldsymbol{y}$$

Exercise 3: Using the definition that $\frac{\partial z}{\partial \boldsymbol{y}} = [\frac{\partial z}{\partial y_1}, \frac{\partial z}{\partial y_2}, \cdots, \frac{\partial z}{\partial y_n}]'$, check the validity of the above equations.

15. Expectations.

Let $\boldsymbol{A}$ be a $k \times k$ matrix of constants, $\boldsymbol{a}$ be a $k \times 1$ vector of constants, and $\boldsymbol{y}$ be a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and nonsingular variance-covariance matrix $\boldsymbol{V}$.

1. $E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{\mu}$.
2. $E(\boldsymbol{A}'\boldsymbol{y}) = \boldsymbol{A}'\boldsymbol{\mu}$.
3. $\mathrm{Var}(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{V}\boldsymbol{a}$

4. $\mathrm{Var}(\boldsymbol{Ay}) = \boldsymbol{AVA}'$.

   Note: If $\boldsymbol{V} = \sigma^2\boldsymbol{I}$, then $\mathrm{Var}(\boldsymbol{Ay}) = \sigma^2\boldsymbol{AA}'$.

5. $E(\boldsymbol{y}'\boldsymbol{Ay}) = \mathrm{trace}(\boldsymbol{AV}) + \boldsymbol{\mu}'\boldsymbol{A\mu}$.

   Note: If $\boldsymbol{V} = \sigma^2\boldsymbol{I}$, then $E(\boldsymbol{y}'\boldsymbol{Ay}) = \sigma^2\mathrm{trace}(\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A\mu}$.

Most of the above equations can be easily derived from the definition of expectation and covariance. For equation 5, note that a scalar is always equal to its trace.

16. Multivariate Normal Distribution.

A random vector $\boldsymbol{Y}' = [Y_1 \quad Y_2 \quad \cdots \quad Y_n]$ is said to have a multivariate normal (or Gaussian) distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if its probability density function is given by

$$f(\boldsymbol{Y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})).$$