Statistical Sciences Honours
Matrix Methods
Lecture 4 – Matrix Rank

Stefan S. Britz
stefan.britz@uct.ac.za

Department of Statistical Sciences
University of Cape Town

4.1 Linearly independent vectors

Let

$$_m\boldsymbol{X}_n = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \cdots \ \boldsymbol{x}_n]$$

Now consider the equation

$$\begin{aligned}
\boldsymbol{X}\boldsymbol{a} &= \boldsymbol{0} \\
[\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \cdots \ \boldsymbol{x}_n] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} &= \boldsymbol{0} \\
a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n &= \boldsymbol{0} \\
\sum_{i=1}^{n} a_i\boldsymbol{x}_i &= \boldsymbol{0}
\end{aligned}$$

4.1 Linearly independent vectors

$$\boldsymbol{X}\boldsymbol{a} = \mathbf{0}$$

- One solution is of course to let $\boldsymbol{a} = \mathbf{0}$, referred to as the trivial solution.
- If no other solution exists, then the vectors $\boldsymbol{x}_1 \dots \boldsymbol{x}_n$ are said to be **linearly independent**.
- In other words, $\boldsymbol{x}_1 \dots \boldsymbol{x}_n$ are linearly independent if $\boldsymbol{X}\boldsymbol{a} \neq \mathbf{0} \; \forall \; \boldsymbol{a} \neq \mathbf{0}$
- Note that the concept of linear independence only arises if $\boldsymbol{x}_i \neq \mathbf{0} \; \forall \; i$. If any $\boldsymbol{x}_i = \mathbf{0}$, then $a_i = c \neq 0$ and $a_j = 0, \; j \neq i$ would yield $\boldsymbol{X}\boldsymbol{a} = \mathbf{0}$.

4.1 Linearly independent vectors

- Now suppose that $\boldsymbol{x}_1 \ldots \boldsymbol{x}_n$ are linearly **dependent** and that $a_n \neq 0$.
- We can then write

$$\boldsymbol{x}_n = -\frac{a_1}{a_n}\boldsymbol{x}_1 - \frac{a_2}{a_n}\boldsymbol{x}_2 - \cdots - \frac{a_{n-1}}{a_n}\boldsymbol{x}_{n-1}$$

  with at least one of the coefficients on the RHS being non-zero.
- Since $\boldsymbol{x}_n$ is a linear combination of the other vectors, we can substitute it such that our system now contains $n-1$ different vectors.
- This process can be repeated until r linearly independent vectors remain, say $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$, and all other vectors are linear combinations of these.

4.1 Linearly independent vectors

- We have in effect partitioned the matrix $\boldsymbol{X}$ as

$$[\boldsymbol{X}_1 \colon m \times r \quad \boldsymbol{X}_2 \colon m \times (n-r)]$$

  such that each column of $\boldsymbol{X}_2$ is a linear combination of the columns forming $\boldsymbol{X}_1$

- Therefore, for some matrix $\boldsymbol{B} \colon r \times (n-r)$, we can write

$$\boldsymbol{X}_2 = \boldsymbol{X}_1 \boldsymbol{B}$$

- Let's illustrate this with an example

Example

Consider the matrix

$$\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \boldsymbol{x}_3 \ \boldsymbol{x}_4] = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 4 & 2 & 3 & 9 \\ 1 & -1 & 0 & 3 \end{bmatrix}$$

We have linear dependence, since we can see that

$$\boldsymbol{X} \begin{bmatrix} 3 \\ 0 \\ -1 \\ -1 \end{bmatrix} = \boldsymbol{0}$$

$$3\boldsymbol{x}_1 - \boldsymbol{x}_3 - \boldsymbol{x}_4 = \boldsymbol{0}$$

Example

$$3\boldsymbol{x}_1 - \boldsymbol{x}_3 - \boldsymbol{x}_4 = \mathbf{0}$$

4.2 Matrix rank

- Consider the square matrix $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}$
- Suppose the columns of $\boldsymbol{X}$ are linearly dependent

$$\boldsymbol{x}_1 = c_1\boldsymbol{x}_2 + c_2\boldsymbol{x}_3 + \cdots + c_{n-1}\boldsymbol{x}_n$$

- It is then possible to add multiples of the other columns to any one chosen column of $\boldsymbol{X}$ in such a way that this column consists entirely of 0's.

$$\boldsymbol{X}^* = \begin{bmatrix} 0 & x_{12} & \cdots & x_{1n} \\ 0 & x_{22} & \cdots & x_{2n} \\ \vdots & & & \vdots \\ 0 & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

- Note that $|\boldsymbol{X}| = \left|\boldsymbol{X}^*\right| = 0$

4.2 Matrix rank

- So if columns of $\boldsymbol{X}$ are linearly dependent, then $|\boldsymbol{X}| = 0$.
- ...which means $\boldsymbol{X}^{-1}$ does not exist.
- ...which means there is no solution to $\boldsymbol{a}$ in the equation $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{b}$ for $\boldsymbol{b} \neq \mathbf{0}$.
- However, because of the linear dependence, we know that there is some $\boldsymbol{a} \neq \mathbf{0}$ such that $\boldsymbol{X}\boldsymbol{a} = \mathbf{0}$.
- This is a crucial general property of square matrices:

Either there is a non-trivial solution to $\boldsymbol{X}\boldsymbol{a} = \mathbf{0}$, or there is a solution to $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{b} \neq \mathbf{0}$, but not both.

4.2 Matrix rank

The following are important facts regarding any arbitrary $m \times n$ matrix $\boldsymbol{X}$:

1. The maximum number of columns of $\boldsymbol{X}$ that can be linearly independent is m (the number of rows of $\boldsymbol{X}$).
2. The maximum number of rows of $\boldsymbol{X}$ that can be linearly independent is n (the number of columns of $\boldsymbol{X}$).
3. The number of linearly independent rows of $\boldsymbol{X}$ $=$ the number of linearly independent columns of $\boldsymbol{X}$.

4.2 Matrix rank

The following definitions are important regarding $\boldsymbol{X}: m \times n$:

- The **rank** of a matrix $\boldsymbol{X}$, written $rank(\boldsymbol{X})$, is the number of linearly independent columns (or rows) of $\boldsymbol{X}$.
- For an $m \times n$ matrix $\boldsymbol{X}$, $rank(\boldsymbol{X}) \leq \min(m, n)$.
- We say that $\boldsymbol{X}$ is of **full row rank** if $rank(\boldsymbol{X}) = m$.
- We say that $\boldsymbol{X}$ is of **full column rank** if $rank(\boldsymbol{X}) = n$.
- If a square matrix is of full row rank, then it is also of full column rank (and vice versa), and $\boldsymbol{X}$ is said to be of **full rank**. (What else does this imply?)

4.3 Factorisation

- It is often useful to factorize a matrix into matrices of full column and row ranks.
- Let $rank\left({}_p\boldsymbol{A}_q\right) = r$, and suppose that the rows and columns have been ordered such the first r of each are linearly independent.
- We can therefore write $\boldsymbol{A}$ in the partitioned form

$$ {}_p\boldsymbol{A}_q = \begin{bmatrix} {}_r\boldsymbol{X}_r & {}_r\boldsymbol{Y}_{q-r} \\ {}_{p-r}\boldsymbol{Z}_r & {}_{p-r}\boldsymbol{W}_{q-r} \end{bmatrix} $$

  where $\boldsymbol{X}$ is of full rank.
- For appropriate matrices $\boldsymbol{F}$ and $\boldsymbol{H}$ representing the relationships between the linearly dependent and independent rows and columns of $\boldsymbol{A}$ respectively, we can derive:

$$ \boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{X} & \boldsymbol{X}\boldsymbol{H} \end{bmatrix} $$

4.3 Factorisation

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{X} & \boldsymbol{X}\boldsymbol{H} \end{bmatrix}$$

4.4 Canonical forms

- From this we can see that the matrix $\boldsymbol{A}$ can be transformed by elementary operations into a matrix consisting only of 0's, except that the first r diagonal elements are 1's.

- To achieve this, we can pre- and post-multiply by the relevant elementary operator matrices, such that

$$\boldsymbol{PAQ} = \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \boldsymbol{C}$$

- The matrix $\boldsymbol{C}$ is termed the equivalent **canonical form** of $\boldsymbol{A}$.

- If two matrices reduce to the same canonical form, then they are said to be **equivalent**.

4.4 Canonical forms

$$\boldsymbol{PAQ} = \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \boldsymbol{C}$$

- Note that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are not unique in general, but that they are invertible.
- We can therefore also express the matrix as a function of its canonical form: $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{Q}^{-1}$.
- If $\boldsymbol{A}$ is non-singular, then $\boldsymbol{C} = \boldsymbol{I}$ and $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{Q}^{-1}$
- For a symmetric matrix $\boldsymbol{A}$ it is possible to find a matrix $\boldsymbol{P}$ such that $\boldsymbol{PAP'} = \boldsymbol{C}$.
- If $\boldsymbol{A}$ is also non-singular, this implies that $\boldsymbol{A} = \boldsymbol{KK'}$ where $\boldsymbol{K} = \boldsymbol{P}^{-1}$.

Exercises

Exercise 4.1

Verify the following useful rank theorems:

(a) $rank(\boldsymbol{AB}) \leq min\{rank(\boldsymbol{A}), rank(\boldsymbol{B})\}$

(b) If $\boldsymbol{A}$ is non-singular, then $rank(\boldsymbol{AB}) = rank(\boldsymbol{B})$

(c) If $\boldsymbol{AGA} = \boldsymbol{A}$, then $rank(\boldsymbol{GA}) = rank(\boldsymbol{A})$

(d) $rank(\boldsymbol{A} + \boldsymbol{B}) \leq rank([\boldsymbol{A}\ \boldsymbol{B}]) \leq rank(\boldsymbol{A}) + rank(\boldsymbol{B})$

4.5 Gram-Schmidt process

- Sometimes it is useful/necessary to find orthogonal vectors in some vector space.
- Given a set of n linearly independent vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, the Gram-Schmidt process provides an algorithm for forming a set of n orthogonal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ spanning the same vector space.
- Therefore, any vector $\boldsymbol{v} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n$ can be written as $\boldsymbol{v} = b_1\boldsymbol{q}_1 + b_2\boldsymbol{q}_2 + \cdots + b_n\boldsymbol{q}_n$.
- Geometrically, this can be seen as the projection of the vectors onto an orthonormal basis.

4.5 Gram-Schmidt process

The Gram-Schmidt process is as follows:

Step 1

$$\boldsymbol{q}_1 = \boldsymbol{x}_1$$

Step 2

$$\boldsymbol{q}_2 = \boldsymbol{x}_2 - \left(\frac{\boldsymbol{q}_1'\boldsymbol{x}_2}{\boldsymbol{q}_1'\boldsymbol{q}_1}\right)\boldsymbol{q}_1$$

Step 3

$$\boldsymbol{q}_3 = \boldsymbol{x}_3 - \left(\frac{\boldsymbol{q}_1'\boldsymbol{x}_3}{\boldsymbol{q}_1'\boldsymbol{q}_1}\right)\boldsymbol{q}_1 - \left(\frac{\boldsymbol{q}_2'\boldsymbol{x}_3}{\boldsymbol{q}_2'\boldsymbol{q}_2}\right)\boldsymbol{q}_2$$

$\vdots$

Step n

$$\boldsymbol{q}_n = \boldsymbol{x}_n - \left(\frac{\boldsymbol{q}_1'\boldsymbol{x}_n}{\boldsymbol{q}_1'\boldsymbol{q}_1}\right)\boldsymbol{q}_1 - \left(\frac{\boldsymbol{q}_2'\boldsymbol{x}_n}{\boldsymbol{q}_2'\boldsymbol{q}_2}\right)\boldsymbol{q}_2 - \left(\frac{\boldsymbol{q}_{n-1}'\boldsymbol{x}_n}{\boldsymbol{q}_{n-1}'\boldsymbol{q}_{n-1}}\right)\boldsymbol{q}_{n-1}$$

Exercises

Exercise 4.2

(a) Compute $\boldsymbol{q}_1'\boldsymbol{q}_2$ above to show that they are orthogonal vectors.

(b) Consider the two vectors $\boldsymbol{x}_1 = \begin{bmatrix}1 & 0 & 2\end{bmatrix}'$ and $\boldsymbol{x}_2 = \begin{bmatrix}1 & -1 & 1\end{bmatrix}'$. Find a set of two orthogonal vectors spanning the same vector space as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

(c) Find a set of two orthonormal vectors spanning the same vector space as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in (b).