Basic linear algebra
(with R practice)

Woojoo Lee

Basic vector operations

1) Addition

Scalar multiplication

→ Linear combination

$$a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k$$

→ Linearly independent

$$a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k = 0$$
$$\Longleftrightarrow a_1 = \cdots = a_k = 0$$

$$\text{Ex:}\quad \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$
$$\rightarrow \quad \text{linearly dependent}$$

Remark) Check the geometric meaning of the “addition” and “scalar multiplication”.

Definition: Any set of m linearly independent vectors is called a **basis** for the vector space of all m-tuples of real numbers.

$$\Downarrow$$

Every vector can be expressed as a unique linear combination of a fixed basis.

Ex: $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

2) Inner product

$$\mathbf{x}^T\mathbf{y} = \sum_i x_i y_i$$

$$\text{Length: } \sqrt{\mathbf{x}^T\mathbf{x}}$$

$$\text{Angle: } \cos(\theta) = \frac{\mathbf{x}^T\mathbf{y}}{\sqrt{\mathbf{x}^T\mathbf{x}}\sqrt{\mathbf{y}^T\mathbf{y}}} = \frac{\mathbf{x}^T}{\sqrt{\mathbf{x}^T\mathbf{x}}}\frac{\mathbf{y}}{\sqrt{\mathbf{y}^T\mathbf{y}}}$$

$\mathbf{x}$ and $\mathbf{y}$ are perpendicular

$$\Longleftrightarrow \mathbf{x}^T\mathbf{y} = 0$$

$$\text{Ex: } \mathbf{x}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

Remark) 1) Compare your results with R results.

$$\text{Ex: } \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

2) Make a R-function to compute the angle.

The projection of a vector $\mathbf{x}$ on a vector $\mathbf{y}$ is

$$\frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$$

This definition can be used for constructing perpendicular vectors.

Ex) Construct perpendicular vectors from

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Given linearly independent vectors $\mathbf{x}_1, \cdots, \mathbf{x}_k$,

how can we make mutually perpendicular vectors with the same linear span ?

Gram-Schmidt process

$$1) \quad \mathbf{u}_1 = \mathbf{x}_1$$

$$2) \quad \mathbf{u}_j = \mathbf{x}_j - \frac{\mathbf{x}_j^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1} \mathbf{u}_1 - \cdots - \frac{\mathbf{x}_j^T \mathbf{u}_{j-1}}{\mathbf{u}_{j-1}^T \mathbf{u}_{j-1}} \mathbf{u}_{j-1}$$

for $j = 2, \cdots, k$.

Remark 1) Do “?qr” in R.

Matrix : basic definitions and elementary operations

Matrix addition/ scalar multiplication

Transpose

Identity matrix

Matrix inverse

Remark) See how to input the matrix in R.

Remark) Matrix multiplication is, in general, not commutative.

Consider a $p \times q$ matrix A:

$$A = (a_{ij})(i = 1, \cdots, p, q = 1, \cdots, q)$$

1. A is a square matrix if $p = q$.
2. $(AB)^T = B^T A^T$ where T denotes the transpose.
3. A square matrix A is symmetric if $A = A^T$.
4. A square matrix A is diagonal if $a_{ij} = 0$ for all $i \neq j$
5. A square matrix A is orthogonal if $AA^T = A^T A = I_p$.
6. A square matrix A is idempotent if $AA = A$.

Ex: $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}, \mathbf{R} = \begin{pmatrix} cos(\alpha) & -sin(\alpha) \\ sin(\alpha) & cos(\alpha) \end{pmatrix}$

$I_n - \frac{1}{n} 1_n 1_n^T, A(A^T A)^{-1} A^T$

Some terminologies

- Row rank: the maximum number of linearly independent rows
- Column rank: the maximum number of linearly independent columns
- The row rank and the column rank of a matrix are equal.
- The rank of matrix is either the row rank or the column rank.
- A square matrix $\mathbf{A}$ is nonsingular if A has full rank.

See Result 2A.10. and 2A.11. for the details related to "inverse".

Remark) Do "?solve" in R.

The determinant of a square matrix A is denoted by $|A|$ or $det(A)$. i.e.

1. For a 2×2 matrix A, $|A| = a_{11}a_{22} - a_{12}a_{21}$.
2. For $p \times p$ matrix A, $|aA| = a^p|A|$.
3. For two $p \times p$ matrices A and B, $|AB| = |A||B|$.
4. If A is orthogonal matrix, $|A| = \pm 1$.
5. A square matrix A is nonsingular if $|A| \neq 0$.

Ex: $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$

The trace of a square matrix A is the sum of diagnoal elements.
i.e.

$$tr(A) = \sum_i A_{ii}$$

See Result 2A.12. for the details related to "trace".

Some useful formulas are

1. $tr(aA + bB) = atr(A) + btr(B)$ for scalars a and b (the trace is a linear function)
2. $tr(A) = tr(A^T)$
3. $tr(AB) = tr(BA)$
4. $tr(AA^T) = \sum_{i,j} A_{ij}^2$

Ex: $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$

Eigenvalue/ eigenvector

A square matrix A is said to have an eigenvalue λ and its associated eigenvector $\mathbf{x}(\neq 0)$, if

$$A\mathbf{x} = \lambda\mathbf{x}$$

Ordinarily, we normalize $\mathbf{x}$ to have unit length.

Ex: Compute eigenvalues and eigenvectors:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix}$$

Eigen-analysis

Spectral decomposition:

Let A be a $k \times k$ symmetric matrix.
Then A can be expressed in terms of its k eigenvalue-eigenvectors pairs (λ_i, e_i) as

$$A = \sum_{i=1}^{k} \lambda_i e_i e_i^T = P \Lambda P^T$$

where $P = [e_1, e_2, \cdots, e_k]$ and $\Lambda = diag(\lambda_i)$

Remark) P is an orthogonal matrix.

Why is the spectral decomposition important ?

For instance, consider A^d.

Some properties are

1. $|A| = \prod_i \lambda_i$
2. $tr(A) = \sum_i \lambda_i$

Note that the below is not always defined.

$$1) A^{-1}$$

$$2) A^{1/2}$$

$$3) A^{-1/2}$$

Now, do you understand why the spectral decomposition is important in statistics ? E.g.- standardize multivariate random vectors.

$$A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = I$$

Quadratic form

Let $\mathbf{x} = (x_1, \cdots, x_p)^T$.

$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for a symmetric $p \times p$ matrix A.

Another representation: $Q(\mathbf{x}) = \sum_{i=1}^{p} \sum_{j=1}^{p} A_{ij} x_i x_j$.

$\rightarrow Q(\mathbf{x})$ is called a quadratic form.

e.g.) $\mathbf{x} = (x_1, x_2, x_3)$ and $A = \begin{pmatrix} 1 & 3 & 0 \\ 3 & -1 & -2 \\ 0 & -2 & 2 \end{pmatrix}$. Compute $Q(\mathbf{x})$.

Positive-definite matrix (PD)

Suppose that A is a symmetric matrix.

A is positive definite if

$$\mathbf{x}^T A\mathbf{x} > 0$$

for all $\mathbf{x} \neq 0$.

Remark) The characteristic of PD matrix can be understood easily in terms of its spectral decomposition.

Remark) A is positive definite if and only if every eigenvalue of A is positive.

If A is PD, A^{-1} is also PD ?

Show that for all $(x_1, x_2) \neq (0, 0)$:

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2 > 0.$$

Why does PD matrix have a special status ?

Consider a quadratic form of $\mathbf{x}^T A\mathbf{x}$ where A is PD.

Let's check whether $\mathbf{x}^T A\mathbf{x}$ satisfies the three condition for the distance function.

Some useful techniques frequently used in multivariate analysis

Matrix inequalities (an extended version of Cauchy-Schwarz inequality.)

Let b and d be any two $p \times 1$ vectors.

1. $(b^T d)^2 \leq (b^T b)(d^T d)$

   with equality if and only if $b = cd$ for some constant c.

2. $(b^T d)^2 \leq (b^T B b)(d^T B^{-1} d)$

for a PD matrix B.

3. $$max_{x\neq 0}\frac{(x^T d)^2}{x^T B x} \quad = \quad d^T B^{-1} d$$

for $p \times p$ PD matrix B and d is a given vector.

4. $$\begin{aligned} max_{x\neq 0}\frac{x^TBx}{x^Tx} &= \lambda_1 \\ min_{x\neq 0}\frac{x^TBx}{x^Tx} &= \lambda_p \end{aligned}$$

for $p \times p$ PD matrix B

with eigenvalues $\lambda_1 \geq \cdots \lambda_p \geq 0.$

An application of the matrix inequalities

$$\max_{||a||=1} Var(a^T \mathbf{x})$$

Remark) Think of geometry of this problem !