Linear algebra for computational statistics I

Jong-June Jeon

September, 2025

Department of Statistics, University of Seoul

Things to know

- basic operation of matrix
- spanning space, null space
- projection and geometry
- linear map and matrix

Vector and Matrix

in the view of computational perspective

Step 1

계산에서 행렬과 벡터를 사용하는 중요한 이유 중 하나가 간결한 표현이다. 복잡한 식의 형태를 행렬과 벡터를 도입함으로써 간단한 표현형을 얻을 수 있고, 그것을 이용하여 식의 변형과 계산에 대한 insight를 얻을 수 있다. 여기서는 행렬과 벡터의 기본 연산과 최적화에서 자주 사용되는 간단한 등식에 대해서 배운다.

Notation

- Denote a 2-dimensional data array ($n \times p$ matrix) by $\mathbf{X}$.
- Denote the element in the ith row and the jth column of $\mathbf{X}$ by x_{ij} or $(\mathbf{X})_{ij}$.
- Denote by X_j the jth column vector of $\mathbf{X}$.
- Denote the ith data(observation or record) by $\mathbf{x}_i$ (column vector). Thus,

$$\mathbf{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_p \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}.$$

Multiplication

Let $\mathbf{A}$ be $n \times p$ matrix, and $\mathbf{C}$ be $p \times m$ matrix. The $\mathbf{AC}$ is $n \times m$ matrix, and $(\mathbf{AC})_{ij} = \sum_{k=1}^{p} (\mathbf{A})_{ik} (\mathbf{C})_{kj}$.

- $AB_1C + AB_2C = A(B_1 + B_2)C$
- $B_1AC + AB_2C \neq A(B_1 + B_2)C$

Multiplication of block matrix

Suppose that $A_{ij}B_{jk}$s are well defined. Then,

$$\begin{aligned}
&\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\
= \quad &\begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}
\end{aligned}$$

Transpose Transpose is an operation defined on matrix. We denote the transpose of $\mathbf{A}$ by $\mathbf{A}^\top$. Image of transpose of $n \times p$ matrix is $p \times n$ matrix with $(\mathbf{A}^\top)_{ij} = \mathbf{A}_{ji}$

- $(AB)^\top = B^\top A^\top$
- $(A_1 A_2 \cdots A_k)^\top = A_k^\top \cdots A_2^\top A_1^\top$

Transpose of block matrix

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^\top = \begin{pmatrix} A_{11}^\top & A_{21}^\top \\ A_{12}^\top & A_{22}^\top \end{pmatrix}$$

Example

Let X be

$$\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix},$$

then $X^\top X$ is given by

$$\begin{aligned} & \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}^\top \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} = \begin{pmatrix} X_{11}^\top & X_{21}^\top \\ X_{12}^\top & X_{22}^\top \end{pmatrix} \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \\ = & \begin{pmatrix} X_{11}^\top X_{11} + X_{21}^\top X_{21} & X_{11}^\top X_{12} + X_{21}^\top X_{22} \\ X_{12}^\top X_{11} + X_{22}^\top X_{21} & X_{12}^\top X_{12} + X_{22}^\top X_{22} \end{pmatrix} \end{aligned}$$

Trace Trace is an operation defined on squared matrix.

$$tr : A \in \mathbb{R}^{p\times p} \mapsto \sum_j (A)_{jj} \in \mathbb{R}$$

.

- $tr(A + B) = tr(A) + tr(B)$
- $tr(kA) = ktr(A)$ (k is a constant)
- Let $\mathbf{A} \in \mathbb{R}^{n\times p}$ and $\mathbf{C} \in \mathbb{R}^{p\times n}$. Then,

$$tr(\mathbf{AC}) = tr(\mathbf{CA})$$

- $tr(A^\top A) = \sum_{i,j} (A)_{ij}^2$

Let $x \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p \times p}$.

$$\exp(x^\top Ax) = \exp(tr(Axx^\top))?$$

(example) $\mathbf{x} \in \mathbb{R}^p$, and let $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$.

- $\exp(-\mathbf{x}^\top \mathbf{\Sigma} \mathbf{x}) = \exp(-tr(\mathbf{x}^\top \mathbf{\Sigma} \mathbf{x}))$
- $\exp(-tr(\mathbf{x}^\top (\mathbf{\Sigma} \mathbf{x})) = \exp(-tr(\mathbf{\Sigma} \mathbf{x} \mathbf{x}^\top)) = \exp(-tr(\mathbf{x} \mathbf{x}^\top \mathbf{\Sigma}))$

Inverse matrix

Let $A, B \in \mathbb{R}^{p \times p}$. If

$$AB = BA = I$$

then B is inverse of A and we denote $B = A^{-1}$.

If the inverse matrices exist,

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

Schur's lemma*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix},$$

provided that A^{-1} and $(D - CA^{-1}B)^{-1}$ are exist.

Orthogonal matrix

$U \in \mathbb{R}^{p\times p}$ is orthogonal if $U^\top U = UU^\top = I$.

Denote the jth column and ith row of U by U_j and $\mathbf{u}_i$, respectively. Check that

- $U^\top = U^{-1}$.
- $U_j^\top U_j = 0$ for $j \neq k$.
- $\mathbf{u}_j^\top \mathbf{u}_k = 0$ for $j \neq k$.

Positive definite matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$. If $a^\top \mathbf{A} a > 0$ for all $a \in \mathbb{R}^p$ $(a \neq 0 \in \mathbb{R}^p)$, then $\mathbf{A}$ is positive definite.

Nonnegative definite matrix If $a^\top \mathbf{A} a \geq 0$ for all $a \in \mathbb{R}^p$ $(a \neq 0 \in \mathbb{R}^p)$, then $\mathbf{A}$ is nonnegative definite.

(Ian) Can we measure a certain amount of positive definiteness?

(Louise) How about this? $\max_a a^\top Aa$ and $\min_a a^\top Aa$.

(Ian) Hm, reasonable. But, we have to worry about the scaling problem.

(Louise) Right. For a fixed A, $a^\top Aa$ can be arbitrary large as $(ka)^\top A(ka) > a^\top Aa$ for all $k > 1$.

(Ian) It'd be better fix it as $\max_{a:\|a\|=1} a^\top Aa$ and $\min_{a:\|a\|=1} a^\top Aa$

Note that every covariance matrix is nonnegative definite.

(proof) Let $\mathbf{X}$ be a random vector and $\mu = \mathrm{E}(\mathbf{X})$, then $\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{X} - \mu)^\top (\mathbf{X} - \mu)$ is a covariance matrix. For all $a \in \mathbb{R}^p$

$$
\begin{aligned}
a^\top \boldsymbol{\Sigma} a &= \mathbb{E} a^\top (\mathbf{X} - \mu)^\top (\mathbf{X} - \mu) a \\
&= \mathbb{E}((\mathbf{X} - \mu) a)^\top ((\mathbf{X} - \mu) a) \\
&= \mathbb{E} \|(\mathbf{X} - \mu) a\|^2 \geq 0
\end{aligned}
$$

Linear equations

Let $x = (x_1, \cdots, x_p)$ be a variable and a_{ij}s and b_js are constants.

$$
\begin{aligned}
a_{11}x_1 + \cdots + a_{1p}x_p &= b_1 \\
a_{21}x_1 + \cdots + a_{2p}x_p &= b_2 \\
&\vdots \\
a_{n1}x_1 + \cdots + a_{np}x_p &= b_n
\end{aligned}
$$

These n equations are simply written by matrix and vector.

$$Ax = b$$

where $A \in \mathbb{R}^{n \times p}$, $x \in \mathbb{R}^p$ and $b \in \mathbb{R}^n$.

Matrix norms measure the size or magnitude of a matrix. They play a crucial role in numerical analysis and matrix computations.

Commonly used matrix norms include:

- Operator Norm (Induced Norm)
- Frobenius Norm

The operator norm (also called the induced norm) of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as:

$$\|\mathbf{A}\|_{\mathsf{op}} = \sup_{x \neq \mathbf{0}} \frac{\|\mathbf{A}x\|_2}{\|x\|_2} = \sup_{\|\mathbf{x}\|_2 = 1} \|\mathbf{A}x\|_2$$

- Measures how much $\mathbf{A}$ stretches a vector.
- Equivalent to the largest singular value (i.e. σ_1 in SVD) of $\mathbf{A}$.
- Sub-multiplicative: $\|\mathbf{AB}\|_{\mathsf{op}} \leq \|\mathbf{A}\|_{\mathsf{op}} \|\mathbf{B}\|_{\mathsf{op}}$

The Frobenius norm of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as:

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |(A)_{ij}|^2 \right)^{1/2}$$

Alternatively,

$$\|A\|_F = \sqrt{\mathrm{Tr}(A^\top A)} = \left(\sum_{i=1}^{\min(m,n)} \sigma_i^2 \right)^{1/2}$$

, where σ_is are singular value of A. (See the SVD in LA03.)

- Equivalent to the Euclidean norm of the matrix as a vector.
- Easy to compute and differentiable.
- Unitary invariant: $\|\mathbf{UAV}\|_F = \|\mathbf{A}\|_F$ for orthogonal matrices $\mathbf{U}$, $\mathbf{V}$ (why?)