Linear algebra for computational statistics I

Jong-June Jeon

September , 2022

Department of Statistics, University of Seoul

Things to know

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

Visualization of Vectors

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.

Different views of vectors

- Coding
- Math

Different views of vectors

$$a_1 = [1, 0], \quad a_2 = [2, 0]$$

What is $a_1 + a_2$? What is $3a_1$? Draw your idea!

Different views of vectors

$$a_1 = [1, 1], \quad a_2 = [2, 1]$$

What is $a_1 + a_2$? Draw your idea!

- On x-axis (another example)
- On y-axis (another example)

Different views of vectors

$$a_1 = [1,1] = [1,0] + [0,1] \quad a_2 = [2,1] = [2,0] + [0,1]$$

Let $a_{11} = [1,0]$, $a_{12} = [0,1]$, $a_{21} = [2,0]$, $a_{22} = [0,1]$ then $a_1 = a_{11} + a_{12}$ and $a_2 = a_{21} + a_{22}$. Thus,

$$a_1 + a_2 = (a_{11} + a_{12}) + (a_{21} + a_{22}) = (a_{11} + a_{21}) + (a_{12} + a_{22})$$

Draw the last term.

벡터의 분해를 이용하여 벡터 더하기 계산 과정을 분해하여 이해할 수 있다. 분해한 벡터를 길이가 1 벡터와 스칼라의 곱으로 더 세밀하게 분해해보자.

- $a_{11} = 1 \cdot [1, 0]$
- $a_{12} = 1 \cdot [0, 1]$
- $a_{21} = 2 \cdot [1, 0]$
- $a_{22} = 1 \cdot [0, 1]$

Visualize $a_1 + a_2$!

여기서 a_1과 a_2는 두 벡터 $[1,0]$와 $[0,1]$의 각 스칼라 곱에 합을 이용하면 항상 표현이 됨을 알 수 있다.

(잠깐!) 만약 다른 두 벡터로 $[1,1]/\sqrt{2}$와 $[-1,1]/\sqrt{2}$로 a_1과 a_2를 표현할 수 있을까? (길이를 1로 맞춰주기 위해서 $\sqrt{2}$로 나눈 것임!)

- (방향)벡터를 분해할 때 기준이 되는 벡터를 잡은 것이다.
- 기준이 되는 벡터는 적당한 조건을 만족하도록 잡아야 한다. (기저벡터 참고)

Vector and Matrix

in the view of computational perspective

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,

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

```python
import numpy as np
A = [[1,2,3],[4,5,6],[7,8,9],[10,11,12]]
print("list:\n", A)
# matrix
A = np.array([[1,2,3],[4,5,6],[7,8,9],[10,11,12]])
print('matrix:\n', A)
A[1,1]
A[1,:]
A[:,0]
# example
n = 100
p = 10
A = np.random.normal(size=(n, p))
A[0,]
A[:,3]
```

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,

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

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 $\mathbf{\Sigma} = \mathbb{E}(\mathbf{X} - \mu)^\top(\mathbf{X} - \mu)$ is a covariance matrix. For all $a \in \mathbb{R}^p$

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

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 $b \in \mathbb{R}^n$.