4-1. Determinant and Trace

행렬식(**Determinant**)은 선형대수학에서 매우 중요한 개념입니다.

행렬식은 오직 정사각 행렬 $A \in \mathbb{R}^{n \times n}$ 에서만 정의됩니다.

행렬식 표현 (해당 교재)

- $det(A)$
- $|A|$ (절대값 X)

정사각 행렬 A의 행렬식은 A를 실수로 mapping하는 함수로 볼 수 있습니다.

일반적인 $n \times n$ 행렬에 대한 행렬식을 정의하기에 앞서, 몇 가지 특정한 형태의 행렬에 대한 행렬식의 정의를 살펴보겠습니다.

예시 4.1 (Testing for Matrix Invertibility)

- 정사각 행렬 A가 역행렬이 존재한다고 가정

가장 크기가 작은 경우를 생각했을 때, 우리는 이미 A A

- $A\ (1 \times 1)$ 행렬 → scalar number
- $A = a \rightarrow A^{-1} = \frac{1}{a}$
- $\therefore a \neq 0,\ a\frac{1}{a} = 1$ 을 만족합니다.

2×2 행렬의 경우를 생각했을 때, 역행렬 정의에 의해 $AA^{-1} = I$ 라는 것을 알고 있습니다.

따라서 A의 역행렬 A^{-1}은

$$\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \quad (4.2)$$

그러므로 위 경우에만 A의 역행렬이 존재합니다.

$$a_{11}a_{22} - a_{12}a_{21} \neq 0.$$

해당 경우의 $A \in \mathbb{R}^{n \times n}$ 의 행렬식은 다음과 같습니다.

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \,. \tag{4.4}$$

해당 예시 4.1은 행렬식과 역행렬의 존재 유무의 관계를 보여주고 있습니다.

정리 4.1

모든 정사각 행렬 ($A \in \mathbb{R}^{n \times n}$)에 대해, $det(A) \neq 0$ 일 때, A의 역행렬이 존재합니다.
$\rightarrow$ invertible

- $n = 1$ 인 경우

$$\det(\boldsymbol{A}) = \det(a_{11}) = a_{11} \,. \tag{4.5}$$

- $n = 2$ 인 경우

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \,, \tag{4.6}$$

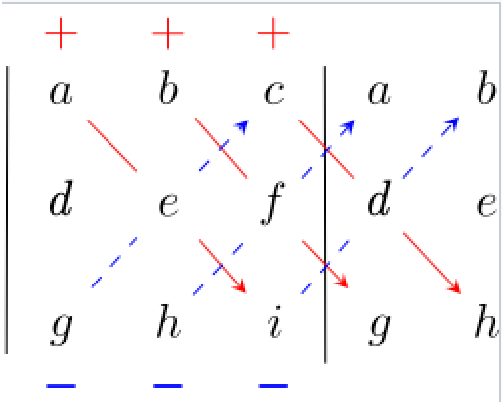
- $n = 3$ 인 경우 (Sarrus' rule)

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \tag{4.7}$$
$$- a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} \,.$$

보충 (Sarrus'rule)

- 행렬식을 계산하기 위한 기억 장치
- 행렬 안에서 대각선 방향으로 곱셈이 어떻게 이루어지는지 선을 따라가면, 쉽게 계산할 수 있습니다.

$$\det(M) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - gec - hfa - idb.$$



정사각 행렬 T를 상삼각행렬(upper-triangular matrix)라고 부르는 것은 $T_{ij} = 0$ 일 때 $(i > j)$, 대각선 아래쪽이 모두 0인 행렬을 뜻합니다.

이와 유사하게 하삼각행렬(lower-triangular matrix)은 대각선 위쪽이 모두 0인 행렬로 정의됩니다.

삼각행렬 $T \in \mathbb{R}^{n \times n}$에 대해, 행렬식은 대각 원소들의 곱으로 주어집니다.

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} T_{ii} .$$

- 상삼각행렬 (upper-triangular matrix)

$$\mathbf{U} = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n} \\ & u_{2,2} & u_{2,3} & \cdots & u_{2,n} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & u_{n-1,n} \\ 0 & & & & u_{n,n} \end{bmatrix}$$

- 하삼각행렬 (lower-triangular matri)

$$\mathbf{L} = \begin{bmatrix} l_{1,1} & & & & 0 \\ l_{2,1} & l_{2,2} & & & \\ l_{3,1} & l_{3,2} & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ l_{n,1} & l_{n,2} & \dots & l_{n,n-1} & l_{n,n} \end{bmatrix}$$

예시 4.2 (Determinants as Measure of Volume)

- 행렬식에 대한 기하학적 의미

행렬 A의 행렬식 $det(A)$은 A를 이루는 열벡터들이 $\mathbb{R}^n$ 공간에서 spanning 하며 이루는 평행육면체(parallelepiped) 형태의 부호 있는 부피(signed volume)를 나타냅니다.

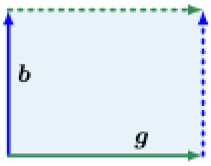
- $n = 2$인 경우 (2차원)

  행렬의 두 열벡터는 하나의 평행사변형(parallelogram)을 만듭니다.

  두 벡터 사이의 각도가 작아질수록, 평행사변형의 넓이는 줄어듭니다.

  두 벡터 b, g를 열벡터로 가지는 행렬 $A = [b, g]$을 생각해보면, 이때 $|det(A)|$는 두 벡터가 만드는 평행사변형의 넓이고, 이 도형의 꼭짓점은 $(0,\ b,\ g,\ b+g)$ 입니다.

Figure 4.2 The area of the parallelogram (shaded region) spanned by the vectors $\boldsymbol{b}$ and $\boldsymbol{g}$ is $|\det([\boldsymbol{b}, \boldsymbol{g}])|$.



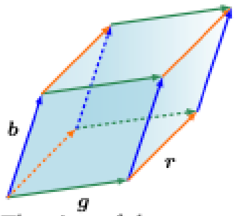
  만약 $b = \lambda g$ 처럼 두 벡터가 선형 종속이라면, 이들은 더 이상 2차원 도형을 만들지 못하고 한 직선만 span하게 됩니다. 즉, 넓이는 0이 됩니다.

반면 b, g가 선형 독립이고 표준기저 e_1, e_2 의 배수라면 → $b = \begin{bmatrix} b \\ 0 \end{bmatrix}$, $g = \begin{bmatrix} 0 \\ g \end{bmatrix} \rightarrow$
$\begin{bmatrix} b & 0 \\ 0 & g \end{bmatrix}$
이때 $det(A) = bg - 0 = bg$ 이 됩니다.

- $n = 3$ 인 경우 (3차원)

  벡터 $r, b, g \in \mathbb{R}^n$이 평행육면체의 세 모서리를 이루는 경우에는 행렬 $A = [r, b, g] \in \mathbb{R}^{3\times3}$에 대해 $|det(A)| =$ 평행육면체의 부피가 됩니다.

Figure 4.3 The volume of the parallelepiped (shaded volume) spanned by vectors $\boldsymbol{r}, \boldsymbol{b}, \boldsymbol{g}$ is $|\det([\boldsymbol{r}, \ \boldsymbol{b}, \ \boldsymbol{g}])|$.



The sign of the determinant indicates the orientation of the spanning vectors.

다음 벡터 $r, b, g \in \mathbb{R}^n$가 있을때,

$$\boldsymbol{r} = \begin{bmatrix} 2 \\ 0 \\ -8 \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix}. \tag{4.9}$$

$$\boldsymbol{A} = [\boldsymbol{r}, \ \boldsymbol{g}, \ \boldsymbol{b}] = \begin{bmatrix} 2 & 6 & 1 \\ 0 & 1 & 4 \\ -8 & 0 & -1 \end{bmatrix} \tag{4.10}$$

$$V = |\det(\boldsymbol{A})| = 186 .$$

- 부호가 의미하는 것 : 방향성

  행렬식의 부호는 b, g 벡터 쌍이 표준 기저 (e_1, e_2)를 기준으로 어떤 방향으로 도형을 구성하는지 나타냅니다.

  예를 들어, $A = [b, g]$ 에서 $A = [g, b]$ 로 열의 순서가 바뀌면 행렬식의 부호가 바뀌고, 평행사변형의 방향이 반대가 됩니다.

$n \times n$ 행렬$(n > 3)$에 대해 행렬식을 계산하는 것은 일반적인 알고리즘이 필요합니다.

정리 4.2는 $n \times n$ 행렬의 행렬식 계산 문제를 $(n-1) \times (n-1)$ 행렬식 계산 문제로 축소시켜 줍니다.

즉, Laplace 전개 정리를 재귀적으로 반복하면 결국 2×2 행렬의 행렬식만 계산하면 되므로, 모든 차원의 행렬식 계산이 가능해집니다.

정리 4.2 (Laplace Expansion)

Theorem 4.2 (Laplace Expansion). *Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Then, for all $j = 1, \ldots, n$:*

1. *Expansion along column j*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j}) . \tag{4.12}$$

2. *Expansion along row j*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k}) . \tag{4.13}$$

Here $\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the submatrix of $\boldsymbol{A}$ that we obtain when deleting row k and column j.

- 라플라스 전개는 여인자 전개(cofactor expansion)로 불리기도 합니다.
- 더 작은 두 행렬식과 그에 맞는 부호를 곱한 것들의 합으로 전개됩니다.

예시 4.3 (Laplace Expansion)

Example 4.3 (Laplace Expansion)
Let us compute the determinant of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.14)$$

using the Laplace expansion along the first row. Applying (4.13) yields

$$\begin{vmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} = (-1)^{1+1} \cdot 1 \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+2} \cdot 2 \begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3} \cdot 3 \begin{vmatrix} 3 & 1 \\ 0 & 0 \end{vmatrix} . \quad (4.15)$$

We use (4.6) to compute the determinants of all 2×2 matrices and obtain

$$\det(A) = 1(1-0) - 2(3-0) + 3(0-0) = -5 . \quad (4.16)$$

For completeness we can compare this result to computing the determinant using Sarrus' rule (4.7):

$$\det(A) = 1 \cdot 1 \cdot 1 + 3 \cdot 0 \cdot 3 + 0 \cdot 2 \cdot 2 - 0 \cdot 1 \cdot 3 - 1 \cdot 0 \cdot 2 - 3 \cdot 2 \cdot 1 = 1 - 6 = -5 . \quad (4.17)$$

$A \in \mathbb{R}^{n \times n}$ 행렬식에 대한 성질들

- $det(AB) = det(A)det(B)$
- $det(A) = det(A^T)$
- A가 가역행렬 (regular, invertible), $det(A^{-1}) = \frac{1}{det(A)}$
- 정의 2.22 유사행렬(similar matrices)은 동일한 행렬식을 가집니다.
    - linaer mapping $\Phi : V \to V$에 대해 Φ의 모든 변환 행렬 A_Φ는 같은 행렬식을 가집니다.
    - linear mapping의 행렬식은 기저 변환에도 변하지 않습니다.
- 한 행(또는 열)에 다른 행(또는 열)의 배수를 더하는 것은 행렬식 $det(A)$에 영향을 주지 않습니다.
- 행렬을 $\lambda \in \mathbb{R}$로 scaling하면, 그 행렬의 행렬식은 λ^n을 곱한 것과 같습니다.
    - $det(\lambda A) = \lambda^n det(A)$

- 두 행/열을 바꾸면 $det(A)$의 부호가 바뀝니다.

마지막 세 가지 성질 덕분에, 가우스 소거법을 사용하여 행렬 A를 REF(행 사다리꼴)로 변환함으로써 $det(A)$를 계산할 수 있습니다.

가우스 소거법은 대각선 아래의 원소들이 모두 0인 삼각행렬(upper-triangular matrix)형태가 될 때까지 수행하면 됩니다. 삼각행렬의 행렬식은 주대각원소들의 곱입니다.

정리 4.3

정사각행렬 $A \in \mathbb{R}^{n \times n}$에 대해 $det(A) \neq 0$인 필요충분조건은 $rk(A) = n$

즉, A가 가역(invertible) ⇔ A가 full rank 일 때 입니다.

과거에는 계산이 대부분 손으로 이루어졌기 때문에, 행렬이 가역인지 분석하기 위해서는 행렬식 계산이 매우 중요했습니다. 하지만 현대 머신러닝 분야에서는 직접적인 수치적 방법을 사용합니다.

예를들어, 2장에서 역행렬은 가우스 소거법으로 구할 수 있습니다. 따라서 가우스 소거법은 행렬식 계산에도 사용될 수 있습니다.

이후 배울 내용에서는 행렬식은 중요한 역할을 하게 됩니다. 특히 고유값(eigenvalue), 고유벡터(eigenvector)를 특성 다항식(characteristic polynominal)을 통해 배울 때 중요합니다.

Trace (대각합)

정의 4.4

정사삭행렬 $A \in \mathbb{R}^{n \times n}$ 에 대해 대각합(trace)은 다음과 같이 정의됩니다.

$$\mathrm{tr}(\boldsymbol{A}) := \sum_{i=1}^{n} a_{ii} \, , \tag{4.18}$$

대각합은 행렬 A의 대각 성분의 합입니다.

대각합(trace)는 다음 성질을 만족합니다.

- $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$
- $\mathrm{tr}(\alpha\boldsymbol{A}) = \alpha\mathrm{tr}(\boldsymbol{A})$, $\alpha \in \mathbb{R}$ for $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- $\mathrm{tr}(\boldsymbol{I}_n) = n$
- $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$ for $\boldsymbol{A} \in \mathbb{R}^{n \times k}, \boldsymbol{B} \in \mathbb{R}^{k \times n}$

이 4가지 성질을 만족하는 유일한 함수는 trace 하나 뿐이라는 것이 증명되었습니다. (Gohberg et al., 2012).

trace의 성질은 행렬 곱에 대해서 더 일반적입니다. 특히 **순환 순열(cycliv permutation)**에 대해 불변합니다.

$$\mathrm{tr}(\boldsymbol{AKL}) = \mathrm{tr}(\boldsymbol{KLA}) \tag{4.19}$$

- $A \in \mathbb{R}^{a \times \Bbbk}, \ K \in \mathbb{R}^{\Bbbk \times l}, \ L \in \mathbb{R}^{l \times a}$

이 성질은 임의의 개수의 행렬 곱으로 일반화 될 수 있습니다.

식 4.19의 특별한 경우로서, 두 벡터 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ 에 대해 다음이 성립합니다.

$$\mathrm{tr}(\boldsymbol{x}\boldsymbol{y}^\top) = \mathrm{tr}(\boldsymbol{y}^\top\boldsymbol{x}) = \boldsymbol{y}^\top\boldsymbol{x} \in \mathbb{R}\,. \tag{4.20}$$

벡터 공간 V에 대해 정의된 linear mapping $\Phi : V \rightarrow V$가 주어졌을 때, Φ의 대각합(trace)을 Φ의 행렬표현의 trace를 이용해 정의합니다.

어떤 기저가 주어지면, 우리는 linear mapping Φ 를 변환 행렬 A를 통해 나타낼 수 있습니다.

이때 Φ의 trace는 A의 trace로 정의됩니다.

> 쉽게말해, 선형 변환을 이해하기 쉽게 행렬로 바꿔 생각하고 그 행렬의 대각선 값을 더한 것을 그 변환(linear mapping)의 trace라고 부른다는 말입니다.

다른 기저를 선택하면, Φ에 대한 새로운 변환 행렬 B는 기저 변환을 통해 $B = S^{-1}AS$의 형태로 얻어집니다.

따라서 Φ의 trace에 대해 다음이 성립합니다.

$$\mathrm{tr}(\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}) \overset{(4.19)}{=} \mathrm{tr}(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}) = \mathrm{tr}(\boldsymbol{A})\,.$$

trace는 닮음 변환 (similarity transformation)에 대해 불변이므로 $tr(B) = tr(A)$ 입니다.

즉, trace는 기저 선택에 영향을 받지 않습니다.

예시

- 기존 선형 변환 행렬 A

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}, \quad tr(A) = 2 + 3 = 5$$

- 기저 변환 행렬 S

$$S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad S^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

- 새로운 기저에서의 행렬 $A^{'} = S^{-1}AS$

$$S^{-1}A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 0 & 3 \end{bmatrix}$$

$$A^{'} = \begin{bmatrix} 2 & -2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

$$tr(A^{'}) = 2 + 3 = 5$$

이번 절에서는 정사각행렬의 성질을 나타내는 **행렬식**과 **대각합**을 살펴보았습니다.

이제 이해한 행렬식과 대각합의 개념을 바탕으로 행렬 A를 **다항식(polynominal)** 형태로 나타내는 중요한 방정식을 정의할 수 있습니다. 이는 이후 여러 절에서 광범위하게 사용될 예정입니다.

Definition 4.5 (Characteristic Polynomial)**.** For $\lambda \in \mathbb{R}$ and a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

$$p_{\boldsymbol{A}}(\lambda) := \det(\boldsymbol{A} - \lambda \boldsymbol{I}) \tag{4.22a}$$
$$= c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n\,, \tag{4.22b}$$

$c_0, \ldots, c_{n-1} \in \mathbb{R}$, is the *characteristic polynomial* of $\boldsymbol{A}$. In particular,

$$c_0 = \det(\boldsymbol{A}), \tag{4.23}$$
$$c_{n-1} = (-1)^{n-1}\mathrm{tr}(\boldsymbol{A}). \tag{4.24}$$

The characteristic polynomial (4.22a) will allow us to compute eigenvalues and eigenvectors, covered in the next section.

이 특성 다항식(Characteristic Polynominal)은 다음 절에서 다룰 고유값(eigenvalue), 고유벡터(eigenvectors)를 계산하는 데 핵심적인 도구가 됩니다.

출처

- Mathmatics for Machine Learning (https://github.com/mml-book/mml-book.github.io)
- https://junstar92.github.io/mml-study-note/2022/07/12/ch4-1.html
- https://en.wikipedia.org/wiki/Rule_of_Sarrus
- https://ko.wikipedia.org/wiki/%EC%82%BC%EA%B0%81%ED%96%89%EB%A0%AC