인공지능수학 개론

(저자 김종락)

(임시적인) 목록

Chapter 5. 선형변환, 고윳값, 고유벡터

두 개의 벡터 공간 V, W 가 주어졌을 때, 이 공간들 사이의 관계를 어떻게 이해할 수 있을까? 두 벡터 공간이 본질적으로 같다는 것은 무슨 뜻일까? 또한 두 벡터 공간이 유사하다는 것은 무슨 뜻일까? 이러한 질문에 답하기 위해선 선형변환의 개념이 필요하다.

선형변환은 하나의 벡터 공간 V에서 자신 또는 다른 벡터 공간 W 가는 함수 L로서 V의 연산들이 그대로 보존되는 성질을 갖고 있다. 따라서 V의 원소들 사이의 연산 값이 L에 의하여 이동된 원소들 사이의 연산 값과 같다. 예를 들어, 거울은 우리의 이미지를 좌우가 바뀐 상태로 그대로 반영하고 있다. 망원경이나 현미경은 작은 물체를 크게 확대하지만 그 모양을 그대로 유지하고 있다. 이러한 성질을 수학적으로 설명할 수가 있다.

고윳값은 선형대수의 중에 특별한 성질을 갖고 있다. 주어진 선형변환 L이 적당한 $v \in V$의 원소를 스칼라 곱으로 변환시키는 경우를 말한다. 즉 $L(v) = \lambda v$을 만족한다. 이때 나오는 λ를 고윳값이라고 하고 v를 고유벡터라고 한다. 고윳값과 고유벡터는 선형변환의 특징을 반영하는 중요한 값들이다. 머신러닝에서 등장하는 주성분분석(PCA)은 고윳값의 크기가 큰 순서로 나열한 후 그에 대응하는 고유벡터들을 새로운 축으로 변환하여 기존 데이터를 읽어 들이는 방법이다. 이 방법은 차원의 저주(curse of dimension)을 부분적으로 해결한 것으로 볼 수 있다.

이 장에서는 여러 개념들을 소개하고 서로 어떻게 연결되는지 살펴본다.

■ 선형변환

● (정의)

벡터공간 V에서 벡터공간 W로 가는 함수 $L: V \to W$가

(1) 모든 $u,v \in V$에 대하여 $L(u+v)=L(u)+L(v)$

(2) 모든 $v \in V$와 스칼라 c에 대하여 $L(cv)=cL(v)$

을 만족할 때 이 L을 선형변환(linear transformation, linear map)이라고 한다.

(그림)

● (예제) $V=R^n$이라고 하고 $L(x)=3x$라고 할 때 L은 선형변환이다.

(증명)

● (예제) 일반적으로 $V=R^n$이라고 하고 $L(x)=ax$ (단, a는 실수인 상수)라고 할 때 L은 선형변환이다. 만약 $a=0$이면 $L(x)=0$이고 이를 자명변환(trivial map)이라고 한다.

(증명)

● (예제) 일반적으로 $V=R^n$이라고 하고 $L(x)=ax+b$ (단 a,b는 상수이며 $b \neq 0$)일 때 L은 선형변환이 아니다.

(증명)

● (질문) 선형변환은 어떻게 만들면 될까?

$V=R^2$라고 하고 $L(x)=3x$인 선형변화을 좀 더 다르게 표현해 보자.

$$L\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \end{pmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

따라서 스칼라로 표현된 선형변환을 대각행렬로 표현할 수 있었다. 여기에 영감을 받아 대각행렬대신 임의의 행렬로 대신해도 과연 선형변환이 될까?

● (정리) 벡터공간 $V = R^n$이라고 하고 $W = R^m$이라고 하자. 행렬 A를 실수 위에서 정의된 $m \times n$ 행렬이라고 하자. 그러면 $L(v) = Av$라고 정의된 함수가 성형변환이 된다. (단 여기서 v는 열벡터로 이해한다.)
(증명)

놀라운 사실은 $V = R^n$에서 $W = R^m$로 가는 임의의 선형변환은 행렬로 표현가능하다는 것이다.

● (**중요 정리**) 만약 $L : R^n \to R^m$이 선형변환이라고 하면 $L(v) = Av$를 만족하는 $m \times n$ 행렬 A가 존재한다.
(증명) 임의의 벡터 $v \in V$는 $v = v_1 e_1 + \cdots + v_n e_n$ (단 $\{e_1, \ldots, e_n\}$는 표준기저)로 표현할 수 있다.

$$\begin{aligned} L(v) &= v_1 L(e_1) + \cdots + v_n L(e_n) \\ &= [L(e_1) \ \ldots \ L(e_n)] \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ &= Av \end{aligned}$$

● (**정의**) 위 정리에 나오는 A를 선형변환 L의 표준행렬(Standard matrix)라고 한다.

● (예제)
- 회전변환 rotation

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \qquad \begin{aligned} x_2 &= x_1 \cos\theta - y_1 \sin\theta \\ y_2 &= x_1 \sin\theta + y_1 \cos\theta \end{aligned}$$

- 반사변환 reflection

1. x축을 기준으로 반사하는 선형변환이 아래와 같을 때

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \quad \begin{matrix} x_2 = x_1 \\ y_2 = -y_1 \end{matrix}$$

표준행렬 A를 구하시오.

2. y축을 기준으로 반사하는 선형변환과 그의 표준행렬을 B를 구하시오.

선형변환를 합성해도 선형변환이 된다.

● (정리) 벡터공간 U, V, W에 대하여 두 개의 선형변환 $L_1 : U \to V$, $L_2 : V \to W$에 대하여 그의 합성인 $L_2 \circ L_1 : U \to V$도 선형변환이 된다.

(증명)

위 합성 선형변환을 행렬로 표현하면 어떤 것이 될까?

● (정리) 두 개의 선형변환 $L_1 : U \to V$, $L_2 : V \to W$ 의 표준행렬을 각각 A_1, A_2라고 할 때, $L_2 \circ L_1 : U \to V$의 표준행렬은 $A_2 A_1$이 된다.

(증명)

실수에서 곱셈의 항등원 1이 있듯이 항등원의 역할을 하는 선형변환이 있다.

● (정의) 벡터공간 V에 대하여 $I_V : V \to V$, $I_V(v) = v$ 를 만족할 때, I_V를 항등 선형변환(identity linear transformation)이라고 한다.

실수 a의 곱셈에 대한 역원 a^{-1}이 있듯이 역변환을 정의할 수 있다.

● (정의)

선형변환 $L_1 : U \to V$과 $L_2 : V \to U$ 에 대하여 $L_2 \circ L_1 = I_U$, $L_1 \circ L_2 = I_V$ 를 만족할 때 L_2를 L_1의 역변환(inverse linear transformation)이라고 하고 $L_2 = L_1^{-1}$이라고 쓴다.

만일 $U = V$이고 선형변환 $L : V \to V$의 역변환이 존재하고 표준행렬을 A라고 하면

역변환 L^{-1}의 표준행렬은 A^{-1}이 된다. 선형변환 $L: V \to V$을 선형연산자(linear operator)라고 한다.

● (정의) 선형연산자 $L: V \to V$가 주어졌을 때 모든 $u, v \in V$에 대하여 $L(u) \cdot L(v) = u \cdot v$을 만족할 때 L를 직교연산자(linear operator)라고 한다.

● (정리) 선형연산자 $L: R^n \to R^n$가 주어졌을 때,
모든 $u \in R^n$에 대하여 $\|L(u)\| = \|u\|$인 필요충분 조건은 모든 $u, v \in R^n$에 대하여 $L(u) \cdot L(v) = u \cdot v$을 만족하는 것이다.

따라서 직교연산자를 노름보존 선형연산자(norm-preserving linear operator)라고 한다.

■ 선형변환의 부분공간과 랭크

● (정의) A를 $m \times n$행렬이라고 할 때 다음과 같이 정의한다.

- 열공간(column space)

- 행공간(row space)

- 영공간(null space or kernel)

Ax= a linear comb. of the
columns of A
∈ Col(A)=Span(columns of A)
subspace of R^m

Row(A)=Span(rows of A) 행공간

L(x)= Ax라고 하자.
Nul(A)=Kernel(L)
= {x in R^n | L(x)=0 즉, Ax=0 }

● (정의) A를 $m \times n$행렬이라고 할 때 A의 계수(rank)를 rank(A)라고 쓰고
rank(A)=dim(Col(A))
로 계산된다.

● (예제)

● (정리) 영공간의 정의역 R^n의 부분공간이다.

(증명)

Nul(A) is a subspace of $\mathbb{R}^n$.

($A: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear transform.)

(pf. $0 \in Nul(A)$ since $A0 = 0$.

· Let $x, y \in Nul(A)$. $A(x+y) = \underset{\shortparallel \atop 0}{Ax} + \underset{\shortparallel \atop 0}{Ay} = 0$

$x+y \in Nul(A)$

· Let $x \in Nul(A)$. $A(cx) = c\underbrace{Ax}_{=0} = 0$.

$\therefore cx \in Nul(A)$.

따라서 영공간을 구하기 위해서는 $Ax = 0$의 해집합을 구하면 된다.

● (정의) 행렬 A의 영공간의 차원을 퇴화차수(nullity)라고 한다. 즉

nullity(A)=dim(Null(A)).

● (정리) (**차원의 정리 또는 랭크 정리**)

$m \times n$행렬 A에 대하여

$$\dim(\mathrm{Col}(A)) + \dim(\mathrm{Nul}(A)) = n$$

이 성립한다.

(증명)

pivot의 개수 $= \dim(\mathrm{Row}(A))$ $\quad A =$

$\| $

$\dim(\mathrm{Col}\, A)$ $\quad \mathrm{Col}(A) = \{ Ax \mid x \in \mathbb{R}^n \}$

$Ax = b$를 만족하는 x가 존재.

$\exists\, x \in \mathbb{R}^n$ s/t $Ax = b$.

$\Leftrightarrow$ $Bx = b'$ ($B = \mathrm{rref}(A)$, $A \sim B$.)

$\mathrm{Col}(B)$

■ 고윳값과 고유벡터

● (정의) n차 정방행렬 A에 대하여 $Ax=\lambda x$를 만족하는 $x\neq 0$가 존재할 때 λ를 A의 고윳값(eigenvalue)라고 하고 x를 고유벡터(eigenvector)라고 한다.

- 고유벡터는 영벡터가 될 수 없으나 고윳값은 0이 될 수가 있다.
- $\lambda=0$과 동치는 무엇일까?
- $Ax = \lambda x$
  $Ax - \lambda x = 0$
  $\Leftrightarrow (\lambda I - A)x = 0$
  $\Leftrightarrow \det(\lambda I - A) = 0$

따라서 행렬 A의 고윳값을 모두 구할 수 있다.

● (정의) n차 정방행렬 A에 대하여 $p_A(\lambda)=\det(A-\lambda I)$를 특성다항식 characteristic polynomial이라고 하고 $\det(A-\lambda I)=0$을 A의 특성방정식 characteristic equation이라고 한다.

- n차 정방행렬 A이 고윳값은 없을 수도 있고 중복이 될 수도 있다. 최대 n개를 갖을 수 있다.

- n차 정방행렬 A의 고윳값 중의 하나를 λ라고 하자. 그러면 이에 대응하는 고유벡터들은 벡터공간을 이룬다. 이 벡터공간을 E_λ라고 하고 고유공간이라고 한다. 즉 $E_\lambda = \{x \in R^n \mid Ax = \lambda x\}$. 즉 E_λ는 모든 고유벡터들과 영벡터를 합한 것이다.

● (정리) n차 정방행렬 A의 고윳값은 최대 n개다.
(증명)

- n차 정방행렬 A에 대한 특성방정식 $\det(\lambda I - A) = 0$은 n차 다항식
- n차 다항식은 복소수 범위에서 n개의 근을 가짐
- 동일한 고윳값이 나타나는 횟수인 중복도(multiplicity)와 복소수 근까지 고려하면, n개의 고윳값 존재

● (정리) n차 정방행렬 A이 가역행렬이고 λ이 고윳값이면 A^{-1}의 고윳값은 λ^{-1}이 된다.

(증명)

$$Ax = \lambda x \quad \Rightarrow \quad x = \lambda A^{-1} x$$

$$\Rightarrow \quad A^{-1}x = \frac{1}{\lambda} x$$

● (정리) n차 정방행렬 A에 대하여 n개의 고윳값 $\lambda_1, \dots, \lambda_n$이 존재할 때 다음이 성립한다.

(1) $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$

(2) $tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

(증명) $p_A(\lambda) = |\lambda I - A|$

$$= (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

$$= \lambda^n - (\lambda_1 + \lambda_2 + \cdots + \lambda_n)\lambda^{n-1} + \cdots + (-1)^n \lambda_1 \cdots \lambda_n$$

$$= \lambda^n - tr(A)\lambda^{n-1} + \cdots + (-1)^n \det(A)$$

(1) $\lambda = 0$을 대입하면 $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$가 성립한다.

(2) $|\lambda I - A| = \begin{bmatrix} \lambda - a_{11} & \lambda - a_{12} & \cdots & \lambda - a_{1n} \\ \lambda - a_{21} & \lambda - a_{22} & \cdots & \lambda - a_{2n} \\ \cdots & & & \\ \lambda - a_{n1} & \lambda - a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix}$

$$= \lambda^n - (a_{11} + \cdots + a_{nn})\lambda^{n-1} + r(\lambda)$$

(단, $\deg(r(\lambda)) \le n-2$)

이므로 $tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$가 성립한다.

● (정리)(**케일리-해밀턴 정리** Cayley-Hamilton theorem) n차 정방행렬 A의 특성방정식 $p_A(\lambda) = |\lambda I - A| = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots a_1 \lambda + a_0 = 0$일 때 λ대신에 A를 대입하면 $p_A(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = O$ 을 만족한다.

(증명) Omit.

■ 주성분 분석(Principal Component Analysis, PCA)

● (정의) n차원의 데이터 $\{x_1, x_2, \ldots, x_k\}$에 대하여 평균벡터 (mean vector)는

$$m = \frac{1}{k}\sum_{i=1}^{k} x_i$$

로 정의된다. 공분산 행렬(covariance matrix)는

$$C = \frac{1}{k}\sum_{i=1}^{k} (x_i - m)(x_i - m)^T$$

로서 $n \times n$ 행렬이다. $m = 0$이면 $C = \frac{1}{k}\sum_{i=1}^{k} x_i x_i^T$가 된다.

● (PCA) [참고, <응용이 보이는 선형대수학>]

$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m\}$을 새로운 기저 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n\}$ 의 좌표계로 선형변환

$$\boldsymbol{x} = \sum_{i=1}^{n} (\boldsymbol{x}^\top \boldsymbol{u}_i)\boldsymbol{u}_i \qquad u_j (i = 1, 2, \cdots, n)\text{는 직교하는 단위벡터}$$

$\boldsymbol{x}^\top \boldsymbol{u}_i = \boldsymbol{x} \cdot \boldsymbol{u}_i$는 $\boldsymbol{x}$를 $\boldsymbol{u}_i$ 방향으로 정사영한 벡터의 크기에 해당

처음 K개의 기저벡터만을 사용하여 $\boldsymbol{x}$를 $\hat{\boldsymbol{x}}$으로 근사

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{K} (\boldsymbol{x}^\top \boldsymbol{u}_i)\boldsymbol{u}_i$$

- K차원으로 근사할 때 발생하는 오차를 J 라고 하면 아래가 성립한다.

$$
\begin{aligned}
J &= \frac{1}{n}\sum_{i=1}^{n}\left\| \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i \right\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left\| \sum_{j=1}^{n}(\boldsymbol{x}_i^\top \boldsymbol{u}_j)\boldsymbol{u}_j - \sum_{j=1}^{K}(\boldsymbol{x}_i^\top \boldsymbol{u}_j)\boldsymbol{u}_j \right\|^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\| \sum_{j=K+1}^{n}(\boldsymbol{x}_i^\top \boldsymbol{u}_j)\boldsymbol{u}_j \right\|^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=K+1}^{n}(\boldsymbol{x}_i^\top \boldsymbol{u}_j)^2 \qquad (\because \boldsymbol{x}_i^\top \boldsymbol{u}_j\text{는 스칼라, } \boldsymbol{u}_j\text{는 단위벡터}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=K+1}^{n}(\boldsymbol{x}_i^\top \boldsymbol{u}_j)^\top(\boldsymbol{x}_i^\top \boldsymbol{u}_j) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=K+1}^{n}\boldsymbol{u}_j^\top \boldsymbol{x}_i\boldsymbol{x}_i^\top \boldsymbol{u}_j \\
&= \sum_{j=K+1}^{n}\boldsymbol{u}_j^\top\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)\boldsymbol{u}_j \\
&= \sum_{j=K+1}^{n}\boldsymbol{u}_j^\top C\boldsymbol{u}_j \qquad (\because \text{평균벡터가 영벡터이므로, } \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top = C)
\end{aligned}
$$

한편 $u_j^T C u_j$값은 아래와 같이 계산하면 σ_j^2이 된다.

$$
\begin{aligned}
\boldsymbol{u}_j^\top C\boldsymbol{u}_j &= \boldsymbol{u}_j^\top\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)\boldsymbol{u}_j = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{u}_j^\top \boldsymbol{x}_i\boldsymbol{x}_i^\top \boldsymbol{u}_j \\
&= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i^\top \boldsymbol{u}_j)^\top(\boldsymbol{x}_i^\top \boldsymbol{u}_j) = \sigma_j^2
\end{aligned}
$$

즉

$$
J = \sum_{j=K+1}^{n}\sigma_j^2
$$

- 분산을 가장 크게 하는 기저벡터 u를 찾는 문제를 풀어야 한다. 라그랑지 승수 (Lagrange multiplier)를 이용하여 다음의 문제를 해결한다.

$$\boxed{\begin{array}{c}\boldsymbol{u}^\top C\boldsymbol{u}\text{를 최대화하는 }\boldsymbol{u}\text{를 찾으라.}\\ (\text{제한 조건 : }\|\boldsymbol{u}\|^2 = 1)\end{array}}$$

$$\tilde{J} = \boldsymbol{u}^\top C\boldsymbol{u} - \lambda(\boldsymbol{u}^\top \boldsymbol{u} - 1)$$

$$C\boldsymbol{u} - \lambda\boldsymbol{u} = 0 \qquad \Longrightarrow \qquad C\boldsymbol{u} = \lambda\boldsymbol{u}$$

$$\boldsymbol{u}^\top C\boldsymbol{u} = \lambda\boldsymbol{u}^\top \boldsymbol{u} = \lambda = \sigma^2$$

- 따라서 C로부터 고윳값의 크기가 큰 순서대로 K개의 고유벡터 $\{u_1, u_2, \dots, u_K\}$를 선택한다. 보통 $K=2$를 많이 선택한다.
-