인공지능수학 개론

(저자 김종락)

(임시적인) 목록

Chapter 4. 벡터와 공간, 행렬과 사상

벡터는 선형대수에서 행렬만큼 중요한 주제이다. n차원 벡터는 보통 $a=(a_1, a_2, \ldots, a_n)$로 표현된다. 우리가 알고 있는 실수는 1차원 벡터, 평면 위의 좌표는 2차원 벡터, 공간 위의 좌표는 3차원 벡터로 생각할 수 있다. 시간을 표현하여 3차원 위치를 나타낸다면, 4차원 벡터가 필요할 것이다. 하나의 데이터 혹은 샘플은 3차원 이상의 성분으로 이루어져 있다. 이때 각 성분을 피쳐(feature)라고도 한다. 따라서 데이터 분석을 한다는 것은 여러 개의 벡터들을 하나의 공통 알고리즘으로 이해하고자 한다는 것이다.

벡터들로 이루어진 집합이 벡터의 합과 스칼라 곱에 닫혀있을 때 이 집합 V을 **벡터 공간**이라고 한다. 두 개의 벡터 공간 V, W가 있을 때 이 두 개를 연결하는 함수map를 **선형 변환**(linear transformation)이라고 한다. 이 선형 변환이 행렬로 표현된다는 것은 중요한 사실이다. 따라서 행렬은 단순히 격자로 이루어진 배열이 아니고 함수로도 볼 수 있다. 왜냐하면, $m \times n$ 행렬 A에다가 n차원 (열)벡터 x를 곱하면 Ax는 m차원 벡터가 된다.

이 장에서는 여러 개념들을 소개하고 서로 어떻게 연결되는지 살펴본다.

■ 벡터와 벡터공간

● 벡터의 합과 차

$u=(u_1,\ \ldots, u_n), v=(v_1,\ \ldots, v_n)$라고 할 때

합은$u+v=(u_1+v_1,\ \ldots,u_n+v_n)$, 차는 $u-v=(u_1-v_1,\ \ldots,u_n-v_n)$

로 표현된다.

● 벡터의 스칼라 곱

α가 스칼라이고 $u=(u_1,\ \ldots,u_n)$라고 할 때

스칼라 곱은 $\alpha u=(\alpha u_1,\ \ldots,\alpha u_n)$로 표현된다.

- 벡터의 차는 $u-v=u+(-1)v$ 이므로 합과 스칼라 곱만 있으면 된다.

● 벡터 공간

V를 공집합이 아니라고 하고 u, v, w를 V에 속하는 n차원 벡터라고 하자.

α, β가 스칼라라고 하자. 보통 스칼라는 실수 집합 R 혹은 복소수 집합 C를 의미하나, 일반적으로는 체(field)에서 성립한다. 그러면 다음이 성립하면

(1) $\boldsymbol{u}, \boldsymbol{v} \in V$이면, $\boldsymbol{u}+\boldsymbol{v} \in V$이다.

(2) $\alpha\boldsymbol{u} \in V$

(3) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$

(4) $\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})=(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$

(5) $\boldsymbol{u}+\boldsymbol{0}=\boldsymbol{u}$인 영벡터 $\boldsymbol{0}$가 V에 단 하나 존재한다.

(6) V의 모든 벡터 $\boldsymbol{u}$에 대해,

$\boldsymbol{u}+(-\boldsymbol{u})=\boldsymbol{0}$를 만족하는 $-\boldsymbol{u}$가 존재한다.

(7) $\alpha(\boldsymbol{u}+\boldsymbol{v})=\alpha\boldsymbol{u}+\alpha\boldsymbol{v}$

(8) $(\alpha+\beta)\boldsymbol{u}=\alpha\boldsymbol{u}+\beta\boldsymbol{u}$

(9) $\alpha(\beta\boldsymbol{u})=(\alpha\beta)\boldsymbol{u}$

(10) $1\boldsymbol{u}=\boldsymbol{u}$

이것을 만족하는 집합을 **벡터공간(vector space)**라고 한다. 이 원소들을 벡터라고 한다.

(1), (4), (5), (6)을 만족하는 것을 군(**group**)이라고 한다. (3)까지 만족할 때 이 군을 가환군(**Abelian group**)이라고 한다.

● 벡터 공간의 예

R^n= n차원 실수 벡터들로 이루어진 집합

R^n의 부분 공간= R^n의 부분집합으로서 그 자체가 벡터 공간을 이루는 집합.

- (예) In R^2에서 $S=\{(x,y)\,|\,x+y=0\}$인 부분집합을 생각한다. S가 벡터공간인가?

  Yes

  $x+y=0$은 $y=-x$와 동치. 즉 원점을 지나고 기울기가 -1인 직선을 의미한다. 따라서 S의 원소는 $(x,-x)$형태이다. $(x_1,-x_1)+(x_2,-x_2)=(x_1+x_2,-(x_1+x_2))\in S$ 이고 $\alpha(x_1,-x_1)=(\alpha x_1,-\alpha x_1)\in S$. 또한 0=(0,0)=(0, -0) 이므로 S에 있다. $(x,-x)$의 역원은 $-(x_1,-x_1)=(-x_1,x_1)=(-x_1,-(-x_1))\in S$. 따라서 S는 벡터공간이 된다.

- In R^2에서 $T=\{(x,y)\,|\,x+y=1\}$인 부분집합을 생각한다. T가 벡터공간인가?

  No

  왜냐하면 0=(0,0)이 T에 들어가 있지 않기 때문에.

- 실제로는 R^2의 부분 공간은 {0}, 원점을 지나는 직선들, R^2뿐이다.

P_n= 차수(degree)가 최대 n인 실수 계수를 갖는 다항식의 모임(벡터공간)

 $=\left\{a_0+a_1x+a_2x^2+\cdots+a_nx^n\,|\,a_i\in R\ \forall i\right\}$

 =span$\{1,x,x^2,\ \ldots\ ,x^n\}$

 =$<1,x,x^2,\ \ldots\ ,x^n>$

 따라서 차원 dimP_n= n+1이라고 한다.

$M_{m\times n}(R)$= 실수 위에서 정의된 크기가 $m\times n$인 행렬들의 모임(벡터공간)

(두 행렬의 합과 스칼라 곱이 정의되어 있기 때문)

- 만약 행렬의 곱셈까지 생각한다면 $M_{m\times n}(R)$은 환(ring)이라는 수학적 구조가 된다. 자세한 내용은 생략.

● 부분 공간의 정의

벡터공간 V의 부분집합 S로서, S는 V의 연산을 그대로 이용하면서 자신이 벡터공간이 될 때, S를 부분 공간(subspace)이라고 한다.

- 부분공간 테스트(subspace test)

  S가 V의 부분공간의 필요충분 조건은

  (i) $0 \in S$

  (ii) $x, y \in S$이면 $x + y \in S$

  (iii) $x \in S$일 때, $\alpha x \in S \; \forall \alpha$

● 선형결합(linear combination)

S={$v_1, ..., v_k$}이라고 할 때, $a_1 v_1 + ... + a_k v_k$를 선형결합이라고 한다.

● 선형종속(linearly dependent)와 선형독립(linearly independent)

벡터공간 V의 부분집합 S={$v_1, ..., v_k$}가 linearly independent라는 말은 $a_1 v_1 + ... + a_k v_k = 0$ 을 성립하는 경우는 $a_i = 0 \; \forall \, i$뿐이다. 그 이외의 경우를 linearly dependent라고 한다. 즉 $a_1 v_1 + ... + a_k v_k = 0$을 만족하는 영이 아닌 a_i가 존재한다.

● 예제

In R^3, $S = \{e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)\}$는 linearly independent하다. 왜냐하면 $a_1 e_1 + a_2 e_2 + a_3 e_3 = 0, i.e.,$

$$(a_1, a_2, a_3) = 0, i.e.,$$

$$a_1 = 0 = a_2 = a_3$$

● 생성 집합(spanning set)

V의 임의의 원소를 주어진 집합 S의 선형결합으로 표현가능할 때 S가 V를 span(생성, generate)한다고 한다.

기호로는 V=span(S)

● 기저(basis)

V의 부분집합 S가 V를 생성하고 동시에 S가 linearly independent할 때 S를 V의 basis라고 한다.

● 표준기저(standard basis)

R^n의 표준기저는 {$e_i = (0,0, ..., 1, 0, ..., 0)$| i=1.., n}이다. 직교 좌표계의 근본을 이루는 집합이 바로 표준기저였다.

● 정리
벡터공간 V의 기저가 주어져 있을 때, V의 임의의 원소는 기저들의 선형결합으로 유일하게 표현된다.

● 차원(dimension of a vector space)

● 벡터의 크기(norm)

● 벡터의 내적

(1) $\| \boldsymbol{u} \| \geq 0$

(2) $\| \alpha \boldsymbol{u} \| = |\alpha| \, \| \boldsymbol{u} \|$

(3) $\| \boldsymbol{u} + \boldsymbol{v} \| \leq \| \boldsymbol{u} \| + \| \boldsymbol{v} \|$

(4) $\| \boldsymbol{u} \| = 0$인 경우는 $\boldsymbol{u} = \mathbf{0}$일 때뿐이다.

● 내적의 성질

(1) $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$

(2) $(\boldsymbol{x} + \boldsymbol{y}) \cdot \boldsymbol{z} = \boldsymbol{x} \cdot \boldsymbol{z} + \boldsymbol{y} \cdot \boldsymbol{z}$

(3) $\boldsymbol{z} \cdot (\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{z} \cdot \boldsymbol{x} + \boldsymbol{z} \cdot \boldsymbol{y}$

(4) $c(\boldsymbol{x} \cdot \boldsymbol{y}) = (c\boldsymbol{x}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot (c\boldsymbol{y})$

(5) $\boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2 \geq 0$

(6) $\boldsymbol{x} = \mathbf{0}$일 때만 $\boldsymbol{x} \cdot \boldsymbol{x} = 0$이다.

● 내적과 사잇각

$$\mathbf{x}\cdot\mathbf{y}=\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$$

● 코시-슈바르츠 부등식 Cauchy-Schwarz inequality

$$|\mathbf{x}\cdot\mathbf{y}|\le\|\mathbf{x}\|\|\mathbf{y}\|$$

● 삼각부등식 Triangle inequality

$$\|\mathbf{x}+\mathbf{y}\|\le\|\mathbf{x}\|+\|\mathbf{y}\|$$

● 직교 orthogonal

$x\cdot y=0$

● 피타고라스의 정리 Pythagorean theorem

R^n에 있는 임의의 두 벡터 x, y가 직교하는 것과 동치는

$\|x+y\|^2=\|x\|^2+\|y\|^2$

(증명) $\|x+y\|^2=(x+y)\cdot(x+y)$를 계산해 본다.

● 정사영

R^n에 있는 임의의 두 벡터 x,y에 대하여, 벡터 x의 y방향의 벡터를 x의 y위로의 **정사영**(orthogonal projection)이라고 한다. 영어로는 projection of x onto y라고 읽고

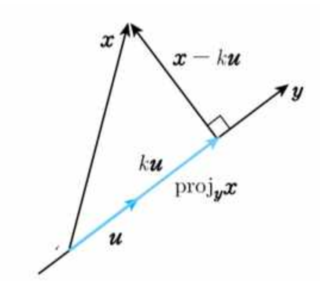
$$\mathrm{proj}_{\boldsymbol{y}}\boldsymbol{x}$$

라고 쓴다. 기학학적 의미는?

● 정사영을 dot product로 표현하면 다음과 같다.

$$\mathrm{Pro}j_x^y = \frac{x \cdot y}{y \cdot y} y$$

(증명)



$u = \dfrac{y}{\|y\|}$라고 하자. 즉 u는 y방향의 단위벡터. 그러면

$\mathrm{Pro}j_x^y = ku$ (단, k는 어떤 실수). 정사영의 정의에 의하여

$(x - ku) \cdot ku = 0$이 된다. k관하여 풀면

$k = \dfrac{x \cdot u}{u \cdot u}$가 된다. 따라서

$$\mathrm{Pro}j_x^y = ku = \frac{x \cdot u}{u \cdot u} u = \frac{x \cdot (\frac{y}{\|y\|})}{\frac{y}{\|y\|} \cdot \frac{y}{\|y\|}} \frac{y}{\|y\|} = \frac{x \cdot y}{y \cdot y} y$$

가 된다.

● 예제

벡터 $x = (1,1)$의 벡터 $y = (2,0)$으로의 정사영을 구하시오.

- 정답은 (1,0)

● (정의) 내적공간 inner product space

실벡터공간 V에 속하는 임의의 두 벡터 x, y 에 대하여 다음의 세가지 조건을 만족하는 함수를 x와 y의 내적이라고 하고 $\langle x, y \rangle$ 또는 $x \cdot y$라고도 쓴다.

(1) 임의의 $\boldsymbol{x} \in V$에 대해, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$이다. 특히, $\boldsymbol{x} = \mathbf{0}$일 때만 $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$이다.

(2) 임의의 $\boldsymbol{x}, \boldsymbol{y} \in V$에 대해, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$이다.

(3) 임의의 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, 스칼라 α, β에 대해, $\langle \alpha\boldsymbol{x} + \beta\boldsymbol{y}, \boldsymbol{z} \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{z} \rangle + \beta\langle \boldsymbol{y}, \boldsymbol{z} \rangle$이다.

● (예제) $V = R^n$라고 할 때 $x = (x_1, \cdots, x_n), y = (y_1, \cdots, y_n)$에 대하여 내적공간을 $x \cdot y = \sum_{i=1}^{n} x_i y_i$으로 정의한다. 이것을 보통의 내적(usual inner product) 또는 유클리디언 내적(Euclidean inner product)이라고 한다.

● (예제) $V = M_{m \times n}(R)$=실수 위에서 정의된 $m \times n$행렬의 모임라고 할 때 $A, B \in M_{m \times n}(R)$에 대하여 내적공간을 아래와 같이 정의할 수 있다.

$$\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

이 내적의 의미는?

● (예제) 구간 $[0,1]$위에서 정의된 연속 함수 $f(x), g(x)$들의 집합은 벡터공간 V을 이룬다. 이러한 $f(x), g(x)$에 대하여 아래와 같이 내적공간을 정의할 수 있다.

$$\langle f, g \rangle = \int_0^1 f(x) g(x) dx$$

● (정의) 벡터공간 V가 주어지고, 임의의 v에 대하여 실수 값 $\|v\|$가 정의되고 다음을 만족할 때, 이 벡터공간을 노름공간 normed space라고 한다.

(1) 임의의 $\boldsymbol{v} \in V$에 대해 $\|\boldsymbol{v}\| \geq 0$이다. 특히, $\boldsymbol{v} = \boldsymbol{0}$일 때만 $\|\boldsymbol{v}\| = 0$이다.
(2) 임의의 $\boldsymbol{v} \in V$와 스칼라 α에 대해, $\|\alpha\boldsymbol{v}\| = |\alpha|\|\boldsymbol{v}\|$이다.
(3) 임의의 $\boldsymbol{u}$, $\boldsymbol{v} \in V$에 대해, $\|\boldsymbol{u}+\boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$이다.

● (예제) V를 내적이 정의된 실수 벡터 공간이라고 하고 $\|x\| = \sqrt{\langle x, x \rangle}$라고 하면 V는 노름공간이 된다.

● (예제) 그러나 R^n이 $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$는 노름을 정의하지만 내적을 정의할 수는 없다. 따라서 내적공간은 노름공간이 되나 그 역은 참이 아니다.

■ 벡터와 미분

이제 벡터와 미분을 연결해 보자. 즉 다변수함수에 대하여 벡터에 관하여 미분을 할 수가 있다.

● (정의) 다변수함수 $f(x_1, x_2, \dots, x_n)$에 대하여 벡터 $x = (x_1, x_2, \dots, x_n)$에 대한 미분의 정의는 $\nabla f = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$인 벡터이고 f의 **그래디언트**gradient라고 한다.

● (예제)
$f(x, y, z) = xyz$라고 할 때 ∇f를 구하시오.

● (정의) 벡터함수 $F(x_1, \dots, x_n) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_n(x_1, \dots, x_n) \end{bmatrix}$에 대하여 각 변수 x_i에 대한 F의 미분은 아래와 같다.

$$\frac{\partial F}{\partial x_i} = \left[\frac{\partial f_1}{\partial x_i} \ \frac{\partial f_2}{\partial x_i} \ \cdots \ \frac{\partial f_n}{\partial x_i} \right]$$

이다. $x = (x_1, \dots, x_n)$이라고 할 때

$\frac{\partial F}{\partial x} = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix}$를 자코비안 행렬 Jacobian matrix라고 하고 J_F라고 쓴다.

● (정의) 다변수함수 $f(x_1, x_2, \dots, x_n)$에 대하여 벡터 $x = (x_1, x_2, \dots, x_n)$에 대한 2차 미분 $\frac{\partial^2 f}{\partial x^2}$을 헤시안 행렬 Hessian matrix라고 하고 $H(f)$라고 쓴다.

$$\frac{\partial^2 f}{\partial \boldsymbol{x}^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

● (정의) 다변수함수 $f(x_1, x_2, \dots, x_n)$의 라플라시안 Laplacian은

$$\nabla^2 f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}$$

이다.

● (예제) $f(x,y) = x^3 + 2x + y^2$이라고 할 때 $\nabla^2 f$를 구하시오.
x에 대하여 f를 두 번 미분하면 $6x$가 된다. y에 대하여 f를 두 번 미분하면 2가 된다. 따라서 $\nabla^2 f = 6x + 2$가 된다.