2022-02-17, Boostcamp RECsys team 12

박정호

Boostcamp

Abstract

벡터공간에 대해서 이해하고, 기저와 선형결합, 선형변환에 대해서 공부한다.

1. 벡터공간 (Vector space)

벡터공간은 벡터공간 위에서 정의된 하나의 연산과 스칼라 곱 연산이 잘 호환되는 대수적 공간을 의미한다. 스칼라 곱연산은 벡터공간 외에 다른 집합의 도움을 받아 정의되는데 그 집합은 주로 사칙연산이 잘 정의된 체(field) 를 이용한다. 딥러닝을 공부하면서 만나는 가장 대표적인 벡터공간은 유클리드 공간이다.

Definition 1.1 (Euclidean space)**.** For some $n \in \mathbb{N}$, $\mathbb{R}^n$ is a $\mathbb{R}$-vector space.

The addition is well-defined on $\mathbb{R}^n$:

1. for any $v, w \in \mathbb{R}^n$, $v + w \in \mathbb{R}^n$,
2. there is $0 \in \mathbb{R}^n$ such that $0 + v = v + 0 = v$ for all $v \in \mathbb{R}^n$,
3. for any $v \in \mathbb{R}^n$, there is $-v \in \mathbb{R}^n$ such that $v + (-v) = (-v) + v = 0$.

The scalar multiplication is well-defined:

1. for any $c \in \mathbb{R}$ and any $v \in \mathbb{R}^n$, $c \cdot v \in \mathbb{R}^n$,
2. for any c, $c' \in \mathbb{R}$ and any $v \in \mathbb{R}^n$, $(cc') \cdot v = c(c'v)$,
3. there is $1 \in \mathbb{R}$ such that $1 \cdot v = v$ for any $v \in \mathbb{R}^n$.

Moreover, both operations are compatible: for any c, $c' \in \mathbb{R}$, and any v, $w \in \mathbb{R}^n$,

$$c \cdot (v + w) = c \cdot v + c \cdot w \quad \text{and} \quad (c + c') \cdot v = c \cdot v + c' \cdot v. \tag{1}$$

벡터공간은 우리가 생각하는 불편함이 없이 계산을 할 수 있는 최소한의 공간이기 때문에 중요하다. 벡터공간은 무한집합이다. 그런데 유한개의 원소로 모든 벡터공간의 원소를 표현할 수 있다. 이러한 원소들의 집합을 기저라고 부른다.

2. 기저 (Basis)

벡터공간의 정의로부터 우리는 다음과 같은 사실을 알 수 있다. 임의의 $c_1, c_2, \cdots, c_n \in \mathbb{R}$, 임의의 $v_1, v_2, \cdots, v_n \in \mathbb{R}^n$ 에 대하여

$$\sum_{i=1}^{n} c_i v_i = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n \in \mathbb{R}^n \tag{2}$$

이 성립한다. 위와 같은 꼴을 우리는 선형결합 (linear combination) 이라고 부른다. $\{v_1, \cdots, v_m\}$이 선형독립 (linear independent)라는 표현은

$$0 = \sum_{i=1}^{m} c_i v_i \quad \Longrightarrow \quad c_1 = c_2 = \cdots = c_m = 0 \tag{3}$$

이 성립할 때 사용한다. 이 수식을 읽는 방법은 임의의 v_i를 나머지 다른 벡터들의 일차결합으로 표현할 수 없다는 것을 의미한다. 반대로 저 명제의 반대 경우를 우리는 선형종속 (linear dependent) 이라고 부른다. 이때 scalar 값들을 우리는 계수 (coefficient) 라고 부르고 좌표 (coordinate) 로 이해할 수 있다. 즉, 기저는 좌표계를 결정한다.

Proposition 2.1. n 차원 벡터 공간은 최대 n 개의 선형독립 집합을 가진다.

Proposition 2.2. n 개의 벡터로 이뤄진 선형독립 집합을 우리는 기저라고 부른다. 그리고 기저를 이용해서 벡터공간의 모든 벡터를 선형결합으로 표현할 수 있다.

Proof. $\mathcal{B} = \{v_1, \cdots, v_n\}$을 벡터공간 $\mathbb{R}^n$의 기저라고 가정하자. 임의의 $w \in \mathbb{R}^n$에 대해서, $w \in \mathcal{B}$ 이면 해당 벡터의 스칼라만 1 나머지를 0으로 둠으로서 일차결합으로 표현 가능하다. $w \notin \mathcal{B}$ 이면, 전의 proposition 2.1에 의해서 $\mathcal{B} \cup \{w\}$는 linear dependent 이고 $w = \sum_i c_i v_i$로 표현할 수 있다. □

3. 선형변환 (Linear transform)

수학에서 함수를 부르는 명칭을 함수 (function), 맵 (map), 변환 (transform), 연산자 (operator) 중 하나로 상황 혹은 뉘앙스에 따라서 다르게 사용한다. 우리가 이번 단원에서 다룰 선형변환이 어떤 뉘앙스에서 변환이라는 이름을 갖게 되었는지 확인해보자.

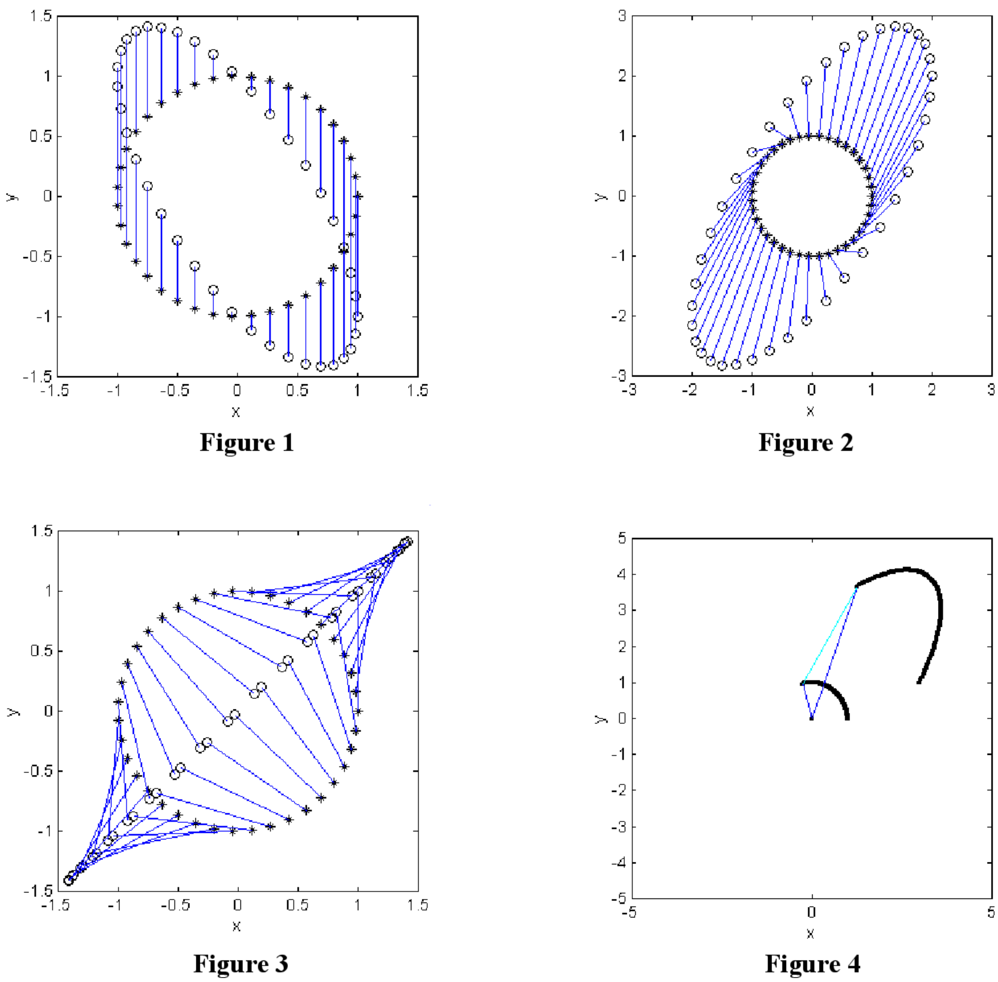

Figure 1: Transforms

선형변환이란 벡터공간 사이에서 정의된 함수로 선형성을 가지고 있다.

Definition 3.1 (선형변환)**.** For some vector spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ where $n,\ m \in \mathbb{N}$, a linear transform $L : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$L(v + cw) = L(v) + cL(w) \qquad \text{for any } v, w \in \mathbb{R}^n, c \in \mathbb{R}. \tag{4}$$

위의 정의가 추상적으로 보이지만 겁먹지 말자, 벡터공간이 가지는 연산을 잘 보존될 수 있도록 디자인 된 물건이다. 선형변환에는 좋은 정리가 하나 있다. 오직 기저가 어떻게 변환됐는지만 관찰해도 모든 벡터의 변환을 특정할 수 있다.

Theorem 3.1 (Linear extension theorem)**.** $\mathcal{B} = \{v_1, \cdots, v_n\}$ is a basis of $\mathbb{R}^n$. A map $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\tilde{L}(v_1) = \sum_{i=1}^{m} c_{1i} w_i, \quad \tilde{L}(v_2) = \sum_{i=1}^{m} c_{2i} w_i, \quad \cdots \quad , \quad \tilde{L}(v_n) = \sum_{i=1}^{m} c_{ni} w_i. \tag{5}$$

Then there is a unique linear transform L such that $L|_{\mathcal{B}} = \tilde{L}$.

Proof. 임의의 벡터를 $v = \sum_{i=1}^{n} c_i v_i$ 라고 표현하자. L은 선형성을 가져야 하기 때문에

$$L(v) = L\left(\sum_{i=1}^{n} c_i v_i\right) = \sum_{i=1}^{n} c_i L(v_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_i c_{ij} w_j \tag{6}$$

로 구체적으로 표현할 수 있다. □

이인석 교수님의 선대군에 등장하는 선형대수의 기본정리는 *선형변환과 행렬은 사실 같은 것*이라는 사실을 알려준다.

Definition 3.2 (선형대수의 기본정리)**.** 선형변환과 행렬은 사실 같은 것이다. 구체적인 관계는 다음과 같다.

$$A(L) = \begin{pmatrix} L(v_1) & L(v_2) & \cdots & L(v_n) \end{pmatrix} \in \mathbb{R}^{m \times n}. \tag{7}$$

즉,

$$L(v) = A(L)v \qquad \text{for any} \quad v \in \mathbb{R}^n. \tag{8}$$

반대로 주어진 행렬으로부터 얻어지는 선형변환은 theorem 3.1의 방법을 따른다.

이로부터 우리가 weight matrix를 곱하는 과정이 사실 선형변환을 의미한다는 것을 알 수 있었고, nonlinearity를 가지는 activation function을 이용해서 딥러닝 모델에 nonlinearity를 준다는 것을 알 수 있다. 선형변환의 합성은 행렬의 곱이 되는 것을 이해하면, 선형변환의 합성은 여전히 선형변환이라는 것을 알 수 있다. 그리고 선형변환의 역변환이 역행렬을 곱해주는 것이라는 것도 알 수 있다.

행렬의 사이즈가 다른 경우를 생각해보자. 이 경우 역행렬이 정의되지 않는다. 우리는 이럴 때 Moore-Penrose pseudoinverse matrix를 계산한다. 이는 특이값 분해를 통해서 얻을 수 있다. Moore-Penrose pseudoinverse matrix를 계산하는데 사용되고 다른 곳에서도 많이 사용되고 중요한 선형대수의 정리하나를 소개한다.

Theorem 3.2 (Spectral Theorem)**.** $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix. Then, there are orthogonal matrices $U \in \mathbb{R}^{n \times n}$ and a diagonal matrix D such that

$$A = UDU^t. \tag{9}$$

Example 3.1. PCA 는 SVD 를 활용한 예제이다. 센터를 0으로 맞춘 데이터들의 행렬 $A \in \mathbb{R}^{n \times m}$ 를 이용해서 공분산행렬을 다음과 같이 만들 수 있다.

$$\operatorname{cov}(A) = \frac{1}{n-1} AA^t \tag{10}$$

AA^t 는 symmetric 행렬이므로 spectral theorem 에 의해서 $AA^t = UDU^t$ 로 분해할 수 있다. 이 때, $D = \operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$라고 가정하면 (필요하면 rearrange한다.), σ_i에 대응하는 singular vector의 방향이 i 번째 주성분 방향을 의미한다.

4. 행렬의 분해 (Decomposition)

많은 선형대수학의 교재에서 우리는 다양한 분해를 배우게 된다. 이는 계산하기 쉬운 혹은 의미를 파악하기 쉬운 변환들로 분해해서 복잡한 변환을 이해할 수 있게 도와준다.

- LU decomposition
- LDU decomposition
- Rank factorization
- Cholesky decomposition
- QR decomposition
- **Eigendecomposition** or spectral decomposition
- **Singular value decomposition**