Interpretation of Number Set

소프트웨어 꼰대 강의

노기섭 교수

(kafa46@cju.ac.kr)

그리고 이게 도대체 뭔가요?

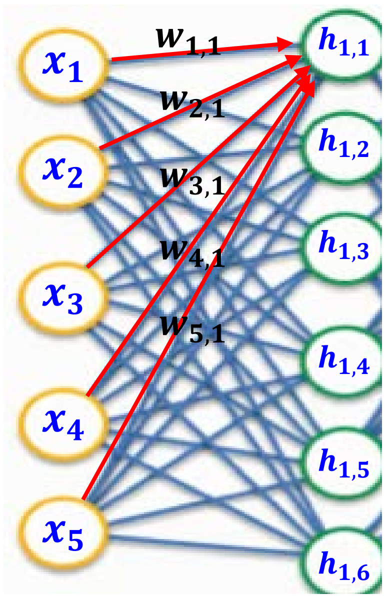
■ 다음 표현은 어떤 차이가 있나요?

$$\boldsymbol{R}_{2,3} \quad \text{vs.} \quad \mathbb{R}^{2\times 3}$$

요거는 행렬 표시

요거는 집합 표시
(오늘 공부할 내용)

머신러닝 처음 공부하면 당황하는 상황...



$$h_{1,1} = w_{1,1}x_1 + w_{2,1}x_2 + w_{3,1}x_3 + w_{4,1}x_4 + w_{5,1}x_5$$
$$h_{2,1} = w_{1,2}x_1 + w_{2,2}x_2 + w_{3,2}x_3 + w_{4,2}x_4 + w_{5,2}x_5$$
$$\vdots \qquad\qquad \vdots$$

$$\begin{pmatrix} h_{1,1} \\ h_{1,2} \\ \vdots \\ h_{1,6} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \begin{pmatrix} w_{1,1} & \cdots & w_{1,6} \\ \vdots & \ddots & \vdots \\ w_{5,1} & \cdots & w_{5,6} \end{pmatrix} \qquad h_1 = x \cdot W, \quad \text{where} \quad x \in \mathbb{R}^5, W \in \mathbb{R}^{5\times6}$$

(같은 결과)

$$\begin{pmatrix} h_{1,1} \\ h_{1,2} \\ \vdots \\ h_{1,6} \end{pmatrix} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,6} \\ \vdots & \ddots & \vdots \\ w_{5,1} & \cdots & w_{5,6} \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \qquad h_1 = W^T \cdot x, \quad \text{where} \quad x \in \mathbb{R}^5, W \in \mathbb{R}^{5\times6}$$

이건 또 뭐냐?

이번 강의를 들으면 수 집합 표현을
이해할 수 있게 됩니다 ^^

Recap: Number Set Notation

Notation for Number Sets

Symbol	Meaning	Remark
$\mathbb{N}$	Natural number	자연수
$\mathbb{Z}$	Integer (Zahl in German)	정수
$\mathbb{Q}$	Quotient(German) or ratio	유리수
$\mathbb{R}$	Real number	실수
$\mathbb{C}$	Complex number	복소수

Recap: Types of Sets

구분	영어 표기	기호 표기
전체집합	Universal set	U
공집합	Empty set	$\emptyset$ 또는 $\{\,\}$
중복집합 (중복원소 허용)	Duplicate set	M
부분집합	Subset	$A \subseteq B$, where $\forall x \in A, x \in B$
진부분집합	Proper subset	$A \subseteq B \wedge A \neq B$ 또는 $A \cap B$
합집합	Union	$A \cup B = \{x \mid x \in A \;\vee\; x \in B\}$ $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$
교집합	Intersection	$A \cap B = \{x \mid x \in A \;\wedge\; x \in B\}$ $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$
여집합	Complementary set	$A^c = \{x \mid x \in U \;\wedge\; x \notin A\}$ 또는 $\{x \in U \mid x \notin A\}$
차집합	Difference of sets	$A - B = \{x \mid x \in A \;\wedge\; x \notin B\}$ 또는 $A \cap B^c$

Recap: Set Theory

Definition

- **Set (집합): 명확한 기준에 의해 주어진 서로 다른 대상이 모여서 만들어지는 새로운 대상**
  - **"서로 다른 대상" → 중복은 의미 없음 + 순서는 상관 없음, 일반적으로 알파벳 대문자로 표기 (A, B, C, …)**
- **Element (원소): Set (집합)을 구성하는 객체, 일반적으로 알파벳 소문자로 표기 (a, b, c, …)**

원소와 집합

- $a \in A$: a는 집합 A의 원소이다 (영어로 "a in A" 라고 읽음)

집합의 표현

- 원소 나열 (원소가 몇 개 안될 경우 → 모두 써준다)

  $A = \{1, 2, 3, 4, 5\}$

- 조건 제시 (원소가 많을 경우 → 원소의 조건을 써준다)

  $A = \{x \mid 1 \leq x \leq 500, x \in \mathbb{N}\}$

  *참고: bar (|) 기호는 "such that"으로, 콤마(,)는 "and"라고 읽음

집합의 집합 → 집합족 (Family of Sets)

Family of Sets (집합족)

- 집합을 원소로 갖는 집합
- 대문자 $\mathcal{F}$로 표기
- $\mathcal{F}$ 원소는 Member(멤버)라고 부름

  (예) 집합 A 의 멱집합(Power Set) → 멱집합은 $P(A)$ 또는 2^A로 표현

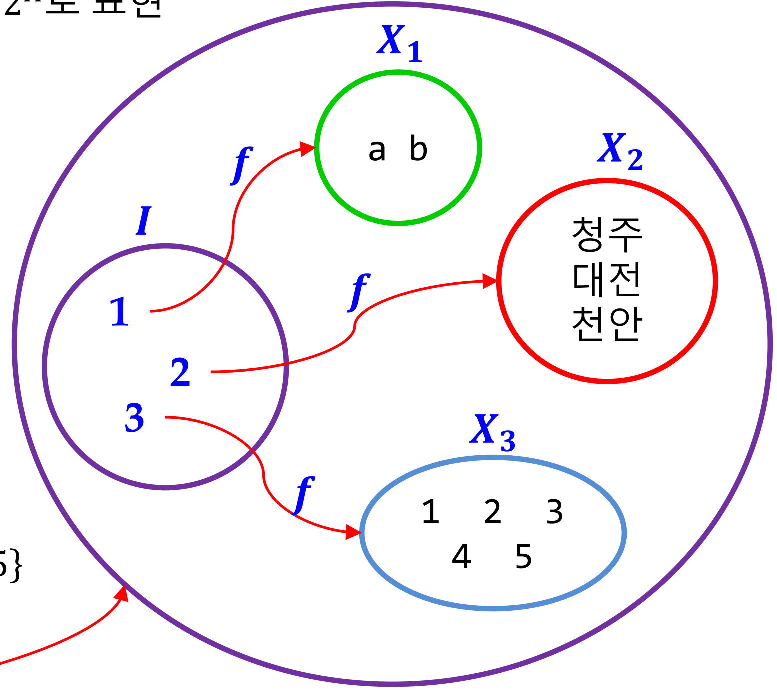
  - $A = \{1, 2\}$
  - $P(A) = 2^A = \{\phi,\ \{1\},\ \{2\},\ \{1,2\}\}$

Indexed Family(첨수족)

- 번호가 부여된 대상들로 이루어진 집합

  (예)

  - $X_1 = \{a, b\}, \quad X_2 = \{청주, 대전, 천안\}, \quad X_3 = \{1, 2, 3, 4, 5\}$
  - Index set (첨수집합): $I = \{1, 2, 3\}$
  - Index Family: $\mathcal{F} = \{X_i \mid i \in I\}$



$$\boldsymbol{f: I \to X}$$

$$\boldsymbol{i \longmapsto x_i = f(i)}$$

Operation of Family Sets

Family set 합집합

$$\cup \mathcal{F} = \bigcup_{A \in \mathcal{F}} A = A_1 \cup A_2 \cup \cdots$$

Family set 교집합

$$\cap \mathcal{F} = \bigcap_{A \in \mathcal{F}} A = A_1 \cap A_2 \cap \cdots$$

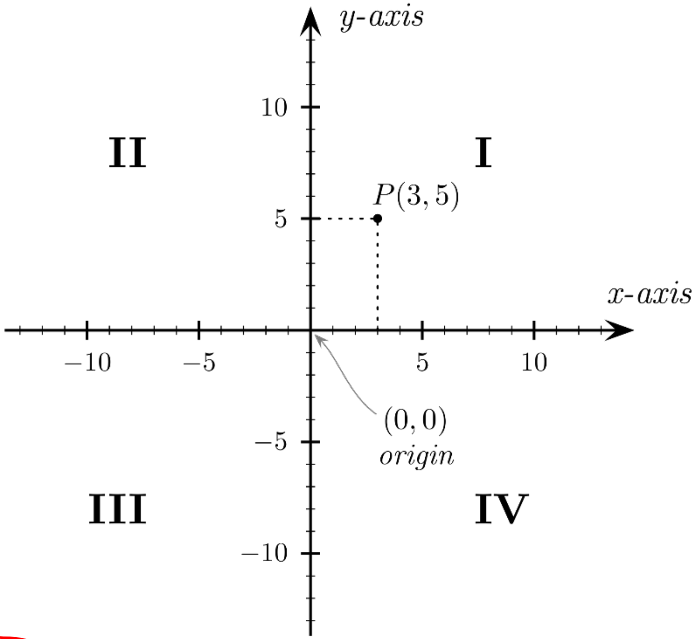
르네 데카르트

1596. 3. 31. ~ 1650. 2. 11.

프랑스의 철학자, 수학자, 과학자

'나는 생각한다 고로 존재한다.'

René Descarte



해석기하학 창시:

직교좌표(데카르트좌표)계 발명

곱집합도 발명

(그래서 Cartecian 곱이라고 부름)

Cartesian Product (곱집합) - Definition

Definition

- 온라인 위키: https://en.wikipedia.org/wiki/Cartesian_product
- 집합론에서 곱집합은 각 집합의 원소를 각 성분으로 하는 튜플들의 집합이다.
- 데카르트 곱(Descartes product), 카티지언곱(Cartesian product) 으로 부르기도 함
- 정의
  - $A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$

- 참고: 튜플(Tuple)
  - "Ordered pair", 순서가 있는 숫자의 열거
  - 소괄호로 묶고 Comma로 구분 ← 예: (1, 2, 3)

Cartesian Product (곱집합)

Cartesian product

- Cartesian product

$$\boldsymbol{A \times B = \{(x, y) \mid x \in A \land y \in B\}}$$

- n-ary Cartesian product

$$\boldsymbol{X_1 \times X_2 \times \cdots \times X_n = \left\{(x_1, x_2, \cdots, x_n) \mid x_i \in X_i, \textbf{for every } i \in \{1, 2, \cdots, n\}\right\}}$$

- n-ary Cartesian Power (거듭제곱 표현)

$$\boldsymbol{X^n = X \times X \times \cdots \times X = \left\{\overbrace{(x_1, x_2, \cdots, x_n)}^{n} \mid x_i \in X, \textbf{for every } i \in \{1, 2, \cdots, n\}\right\}}$$

Cartesian Product (곱집합) - Example

- **Simple Example**

$$\boldsymbol{A} \times \boldsymbol{B} = \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \in \boldsymbol{A} \ \wedge \boldsymbol{y} \in \boldsymbol{B}\}$$

$$A = \{1,2\}, \qquad B = \{3,4\}$$

$$A \times B = \{(1,3),(1,4),(2,3),(2,4)\}$$

$$A \times A = A^2 = \{(1,1),(1,2),(2,1),(2,2)\}$$

$$A \times A \times A = A^3 = \{(1,1,1),(1,1,2),(1,2,1),(1,2,2),(2,1,1),(2,1,2),(2,2,1),(2,2,2)\}$$

- **Family set 곱집합**

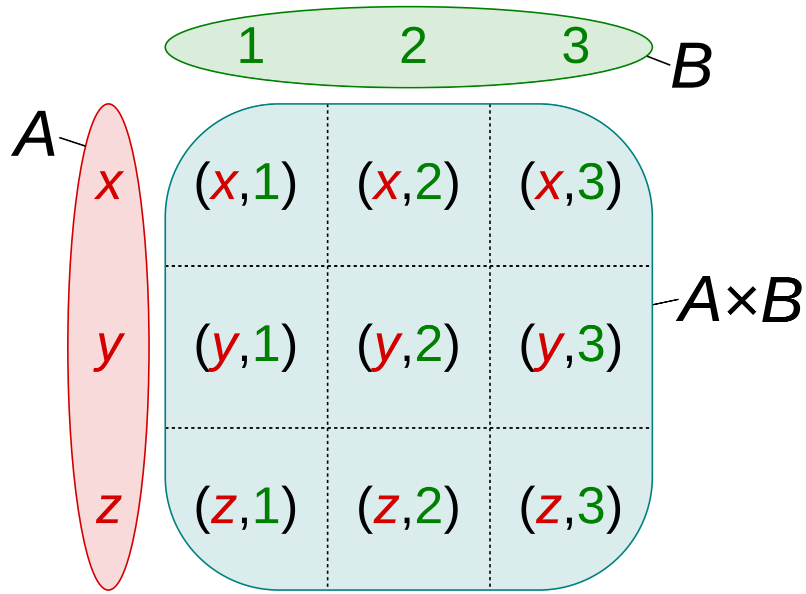
$$\mathcal{F} = \{\boldsymbol{A}_i \mid \boldsymbol{i} \in \boldsymbol{I}\}$$

$$\prod \boldsymbol{A}_i = \{(\boldsymbol{\alpha}_i)_{i \in \boldsymbol{I}} \mid \boldsymbol{\alpha}_i \in \boldsymbol{A}_i\}$$

Cartesian Product (곱집합) - Example

집합 $A = \{x, y, z\}$

집합 $B = \{1, 2, 3\}$



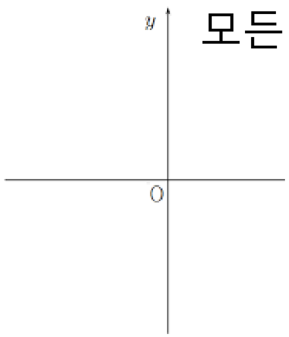
곱집합

$$A \times B = \{(x,1),(x,2),(x,3),(y,1),(y,2),(y,3),(z,1),(x,2),(z,3)\}$$

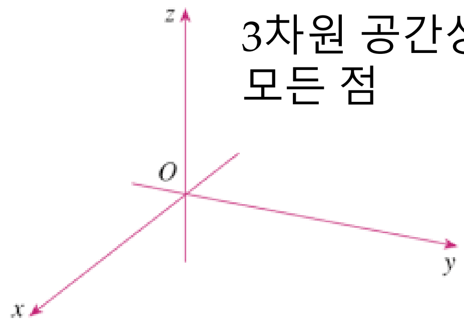
ML에서 주로 쓰는 집합론 지식만 추리면...

- **교환법칙은 성립하지 않음**
  - $A \times B \neq B \times A$
  - 왜냐하면... 순서가 있는 튜플이니까

2차원 평면상의
모든 점 (데카르트 평면)



3차원 공간상의
모든 점



- $\boldsymbol{n - ary}$ **Catecian Product in Euclidean Space**

$$X^2 = X \times X$$

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$$

실수 2개로 만들 수 있는 순서쌍 집합

$$\mathbb{R}^2 = \{(x_1, x_2) \mid x_i \in \mathbb{R}, \textbf{for every } i \in \{1, 2\}\}$$

$$\ddots$$

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$$

실수 3개로 만들 수 있는 순서쌍 집합

$$\mathbb{R}^2 = \{(x_1, x_2, x_3) \mid x_i \in \mathbb{R}, \textbf{for every } i \in \{1, 2, 3\}\}$$

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n}$$

실수 n개로 만들 수 있는 순서쌍 집합

$$\mathbb{R}^n = \{(x_1, x_2, \cdots, x_n) \mid x_i \in \mathbb{R}, \textbf{for every } i \in \{1, 2, \cdots, n\}\}$$

처음으로 돌아와서....

■ $\mathbb{R}^{n\times m}$, 조금 더 해석해 보면...

$$\mathbb{R}^{2\times 3} = \left(\mathbb{R}^2\right)^3 = \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$$

$$A \times B = \{(a,b) \mid a \in A \text{ and } b \in B\}$$

$$\mathbb{R}^2 = \left\{ \boxed{(x_1, x_2)} \mid x_i \in \mathbb{R}, \textbf{for every } i \in \{1,2\}\right\}$$

$$\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 = \{(x_1,x_2)\} \times \{(x_1,x_2)\} \times \{(x_1,x_2)\}$$

$$\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 = \left\{\left((x_1,x_2),(x_1,x_2),(x_1,x_2)\right) \mid x_i \in \mathbb{R}, \textbf{for every } i \in \{1,2\}\right\}$$

$$\mathbb{R}^6 = \left\{(x_1,x_2,x_3,x_4,x_5,x_6) \mid x_i \in \mathbb{R}, \textbf{for every } i \in \{1,2,\cdots,6\}\right\}$$

Matrix vs. Cartesian Product

- 3 ×2 Matrix

$$R_{3,2} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

결국 $R_{3,2}$는 순서가 의미있는
순서쌍 집합 중 하나의 원소

3 × 2로 구성된 순서쌍 원소는
수없이 많을 것임

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}$$

Cartesian Product로
표현하면 간단히 해결

내가 사용할 행렬의 크기가
3 × 2이고, 모든 행렬의 원소가
Real Number (실수)라는
것을 간단히 표기하면?

$$\boldsymbol{A} \in \mathbb{R}^{3\times 2}$$

$n \times m$ 경우로 일반화 시키면?

$$\boldsymbol{X} \in \mathbb{R}^{n\times m}$$

요런건 어떻게 해석하나요?

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^{n} w_i \, x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \, x_i \, x_j \qquad (1)$$

where the model parameters that have to be estimated are:

$$\boxed{w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{R}^{n \times k}} \qquad (2)$$

And $\langle \cdot, \cdot \rangle$ is the dot product of two vectors of size k:

Paper: Steffen Rendle, 'Factorization Machines', 2010
link: https://cseweb.ucsd.edu/classes/fa17/cse291-b/reading/Rendle2010FM.pdf

딥러닝 처음 공부할 때 겪는 어려움

요런건 어떻게 해석해요?

Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions. With a single attention head, averaging inhibits this.

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head_1}, ..., \mathrm{head_h})W^O$$
$$\text{where } \mathrm{head_i} = \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

In this work we employ $h = 8$ parallel attention layers, or heads. For each of these we use $d_k = d_v = d_{\text{model}}/h = 64$. Due to the reduced dimension of each head, the total computational cost is similar to that of single-head attention with full dimensionality.

Paper: A. Vaswani et al., 'Attention is all you need', 2017
link: https://arxiv.org/pdf/1706.03762.pdf

수고하셨습니다 ..^^..