

Mapping Database Concepts to Mathematical Terms

siwon yun

22nd July 2024

0.1 배경

DataBase (이하 DB)에서 Cartesian product는 두 table의 가능한 모든 조합을 생성하는 Join 방법입니다. Cross product는 Cartesian product와 동일한 개념인데, Cross product는 Cartesian product에 비하여 비교적 기능에 직관적인 이름을 가지고 있습니다. 그렇다면, 왜 Cartesian product라는 이름이 사용되는 것 일까요? 본 보고서는 해당 질문에서 시작합니다.

0.2 SQL에서 Cartesian product와 Cross product

```

1 -- Cross Join
2 SELECT * FROM Table1 CROSS JOIN Table2;
3
4 -- Cartesian Join
5 SELECT * FROM Table1, Table2;

```

위는 Cartesian product와 Cross product의 SQL 코드입니다. 현대에는 Cross product를 사용하는 것이 더욱 일반적이며, Cross product는 SQL 표준에서 사용되는 용어입니다. 반면, Cartesian product는 수학적 개념에서 유래한 용어입니다.

0.3 수학 문헌에서 Cartesian product

우연히, SOCP 문제를 해결하는 새로운 방법에 대한 poster에서 ‘Cartesian product’라는 용어를 발견하였습니다.

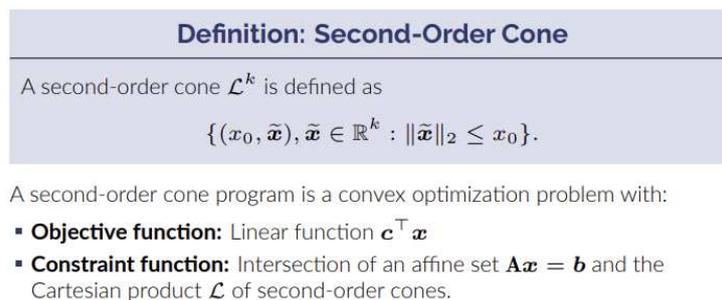


Figure 1: 수학 문헌에서의 Cartesian product
Credit: Michelle Wei

‘Cartesian product’라는 용어와 ‘Cartesian function’이라는 용어가 나온 것을 보니, ‘Cartesian’이 특정한 뜻이 있으리라 생각했었습니다. 비단, ‘Cartesian’은 프랑스의 수학자 René Descartes의 라틴어 이름 ‘Cartesius’에서 유래하였습니다. 작년 진로 시간에 제가 role model 발표 시간에 조사했던 descartes가 바로 ‘Cartesian’이었습니다. 그의 기하학자로서의 업적은 Descartes coordinate system을 제안한 것으로, 여기에는 좌표평면이 포함됩니다.

0.4 좌표계와 Cartesian product

좌표계와 Cartesian product는 서로 연관이 있습니다. 흔히, x 축과 y 축으로 이루어지는 2차원 좌표계는 실수 집합 \mathbb{R} 이 두개가 있음을¹ \mathbb{R}^2 로 표현합니다. 또한 좌표계에서 각 좌표는 순서쌍으로 표현되며, x 축과 y 축의 교차점은 순서쌍 (x, y) 로 표현됩니다. 좌표계의 모든 좌표는 \mathbb{R}^2 의 원소이며, 이는 Cartesian product로 표현됩니다.

$$X \times Y = \{(x, y) \mid x \in X, y \in Y\} \quad (1)$$

¹ x 축과 y 축

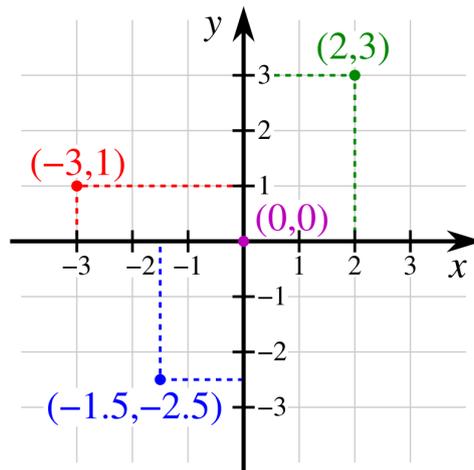


Figure 2: Descartes coordinate system
Credit: [Wikipedia](#)

Equation 1는 집합 X 와 Y 의 Cartesian product를 나타냅니다.² i.e., Descartes coordinate system을 표현하기 위해 사용되는 연산이 Cartesian product이며, 이는 DB에서의 Cross product와 동일한 연산입니다. 수학적 개념과 연관되기 때문에, 직관적이지 아니한 이름을 가진 Cartesian product라는 용어를 사용합니다.

0.5 Relation

DB에서 Table을 Relation이라고 부릅니다. 수학에서는 Relation을 Cartesian product의 임의의 부분집합으로 정의하며, 문자 R 로 표현합니다. 이는 DB에서의 relation과 동일한 개념으로, 모든 가능한 조합에서 특정한 경우를 relation이라 부릅니다.

0.6 Domain

DB에서 relation의 모든 가능한 조합을 구성할 때, 제약 조건이 있을 수 있습니다. 제가 생각하기에 DB의 domain은 수학 용어와 가장 직접적으로 연관된 용어로, 수학에서 domain은 함수의 정의역을 의미합니다. 앞서 언급한 $X \times Y$ 에 대하여는 모든 순서쌍의 x 값을 모아둔 집합을 뜻합니다. 따라서, DB에서 domain을 'int'로 정하는 것을 수학의 함수 표현으로 나타내면 $f: \mathbb{R} \rightarrow \mathbb{R}$ 로 표현됩니다.

0.7 Conclusion

관계형 DataBase는 영국의 수학자 Edgar F. Codd에 의해 제안 되었다고 합니다. 그가 수학자였던 만큼, DB 곳곳에는 수학적 용어가 사용되고 있는 것 같습니다. 따라서, DB를 배울 때 수학적 개념과 관련 지어 생각하면 더욱 깊이 있게 이해할 수 있을 것입니다.

²연산자 '×'