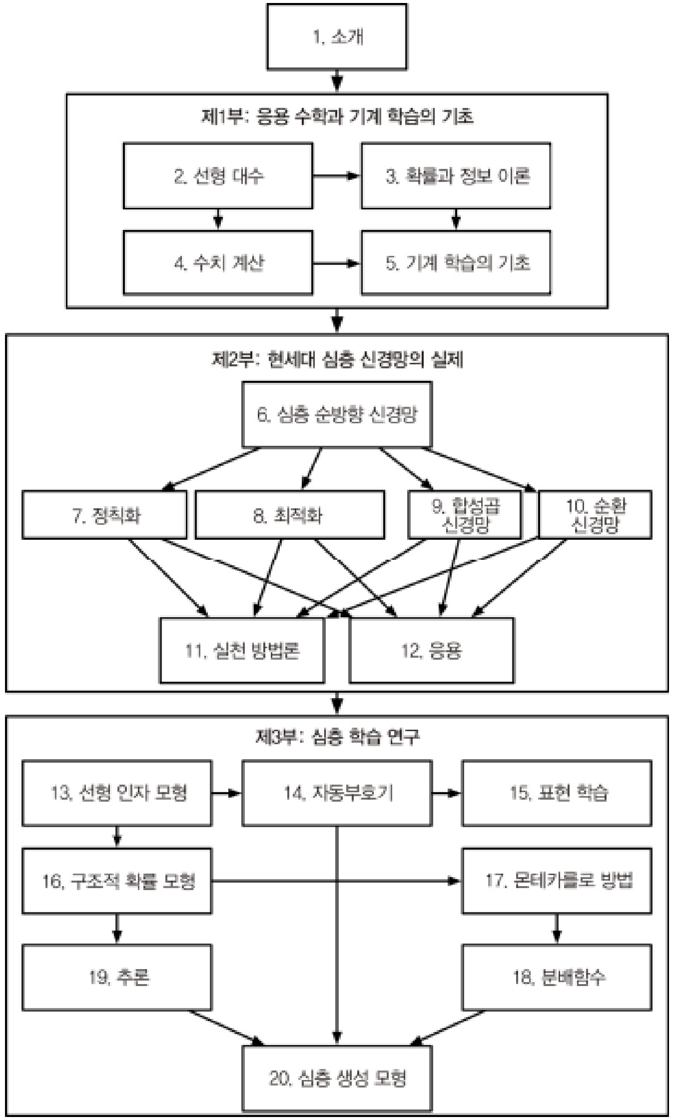
1장 소개

윤예준

심층학습
저자: 이안 굿펠로, 요슈아 벤지오, 에런 쿠빌

책의 구성


1. 소개
제1부: 응용 수학과 기계 학습의 기초
2. 선형 대수
3. 확률과 정보 이론
4. 수치 계산
5. 기계 학습의 기초
제2부: 현세대 심층 신경망의 실제
6. 심층 순방향 신경망
7. 정칙화
8. 최적화
9. 합성곱 신경망
10. 순환 신경망
11. 실천 방법론
12. 응용
제3부: 심층 학습 연구
13. 선형 인자 모형
14. 자동부호기
15. 표현 학습
16. 구조적 확률 모형
17. 몬테카를로 방법
18. 분배함수
19. 추론
20. 심층 생성 모형

다루는 내용

- 직관적인 문제에 대한 하나의 해결책
  - 컴퓨터가 개념들의 계층구조(hierarchy of concepts)를 이용해서 경험으로부터 배우고 세상을 이해하게 만든다는 것

인공지능 접근 방식

- 심층 학습 (Deep learning)
- 지식 베이스(knowledge base)
- 기계 학습(machine learning)
- 표현 학습(representation learning)

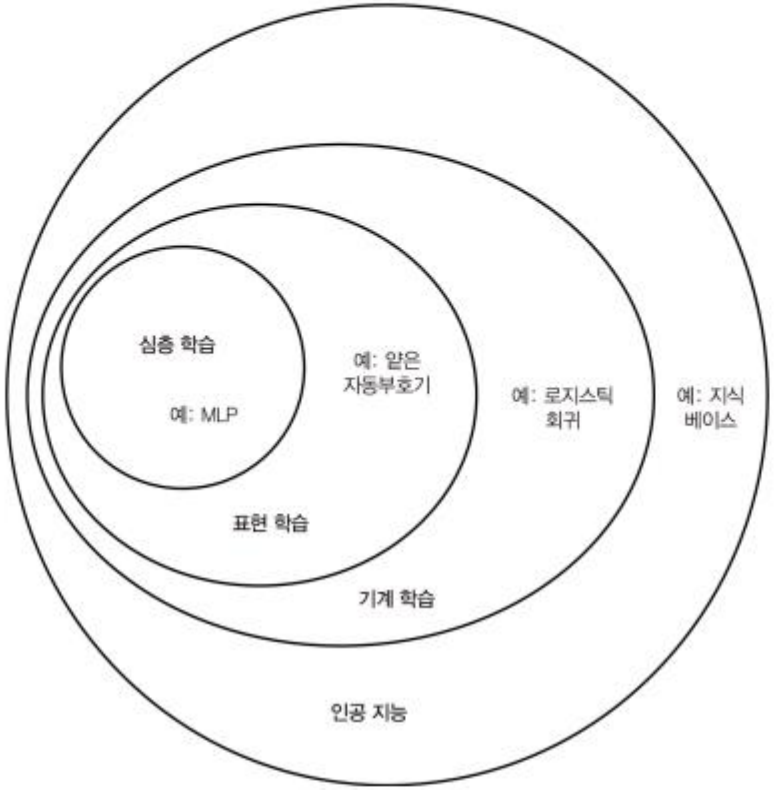


그림 1.4: 여러 인공지능 분야의 관계와 각 분야의 대표적인 예를 보여주는 벤 다이어그램. 심층 학습이 표현 학습의 일종이고, 표현 학습은 기계 학습의 일종임을 알 수 있으며, 기계 학습 접근 방식이 인공지능의 여러 분야에 쓰이긴 하지만 모든 분야에 쓰이는 것은 아니라는 점도 알 수 있다.

인공지능 접근 방식

- 심층 학습 (Deep learning)
- 지식 베이스(knowledge base)
- 기계 학습(machine learning)
- 표현 학습(representation learning)

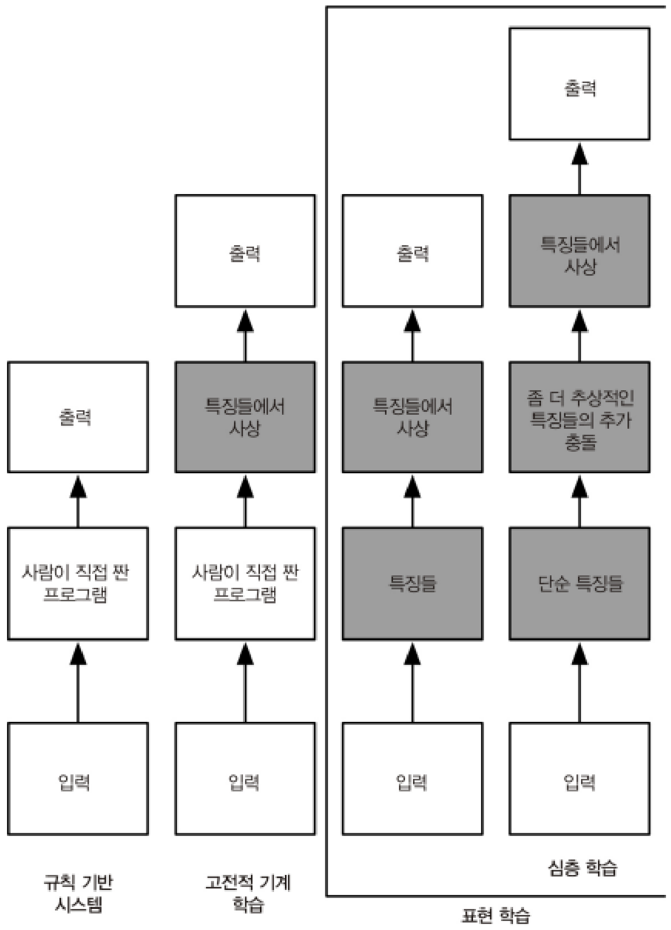


그림 1.5: 여러 인공지능 접근 방식에서 AI 시스템을 구성하는 부분들과 그 관계. 짙은 색 상자는 해당 구성 요소가 자료로부터 학습할 수 있음을 나타낸다.

심층학습 모형과 계산 그래프

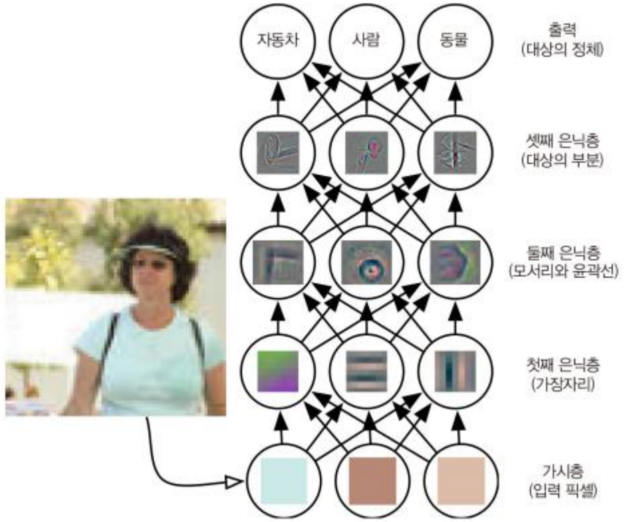


그림 1.2: 심층 학습 모형의 예. 일단의 픽셀값들로 표현된 이 이미지 같은 원본 지각(sensory) 입력 자료의 의미를 컴퓨터가 이해하기는 어렵다. 픽셀 집합을 대상의 정체(identity; 또는 신원)로 사상하는 함수는 아주 복잡하다. 이러한 사상을 배우거나 평가하는 과제를 직접 공략하는 것은 현실적으로 불가능한 일로 보인다. 심층 학습은 원하는 복잡한 사상을 일련의 내포된 단순한 사상들로 분할함으로써 그러한 어려움을 해결한다. 이때 분할된 각각의 사상은 모형의 개별 층으로 서술된다. 원본 입력 이미지는 **가시층**(visible layer)으로 입력된다. 가시층이라는 이름은 이 층이 우리가 실제로 관측할 수 있는 변수들을 담고 있다는 점에서 유래한 것이다. 가시층 밑에는 일련의 **은닉층**(hidden layer)들이 있다. 은닉층들은 이미지로부터 점점 더 추상적인 특징들을 추출한다. 은닉층의 '은닉'은 이 층의 값들이 원본 자료에 주어지지 않은 것이라는 점을 반영한 것이다. 대신 모형은 관측된 자료의 관계들을 서술하는 데 유용한 개념들을 스스로 결정해야 한다. 오른쪽 이미지들은 각 은닉층이 표현하는 특징들의 종류를 보여준다. 첫째 은닉층은 이미지에서 가장자리들을 추출해서 둘째 은닉층에 제공한다. 둘째 은닉층은 그로부터 손쉽게 모퉁이와 윤곽선(둘 다 일단의 가장자리들로 구성된다)을 찾아낸다. 둘째 은닉층의 모퉁이들과 윤곽선들에 기초해서 셋째 은닉층은 대상의 특정 부분들을 식별한다. 각 부분은 특정한 모퉁이들과 윤곽선들로 구성된다. 마지막으로, 대상의 부분들에 대한 서술에 기초해서 모형은 주어진 이미지에 담긴 대상의 정체를 결정한다. [Zeiler & Fergus, 2014]의 그림들을 허락하에 전재했다.

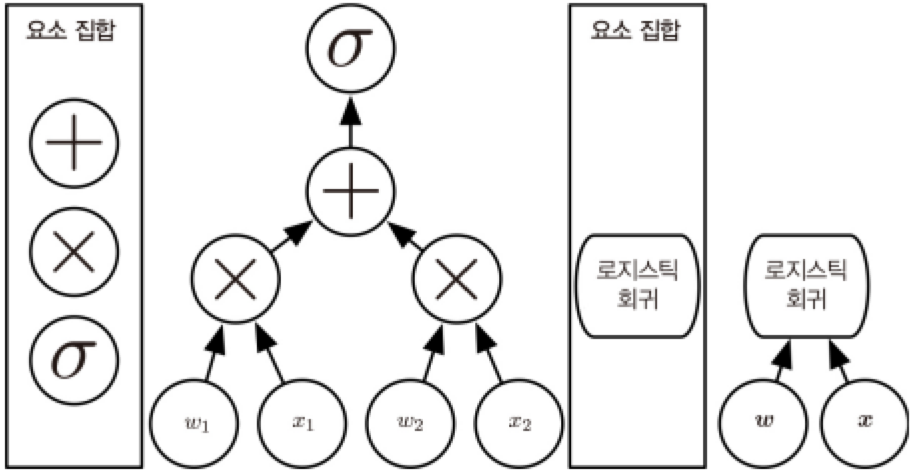
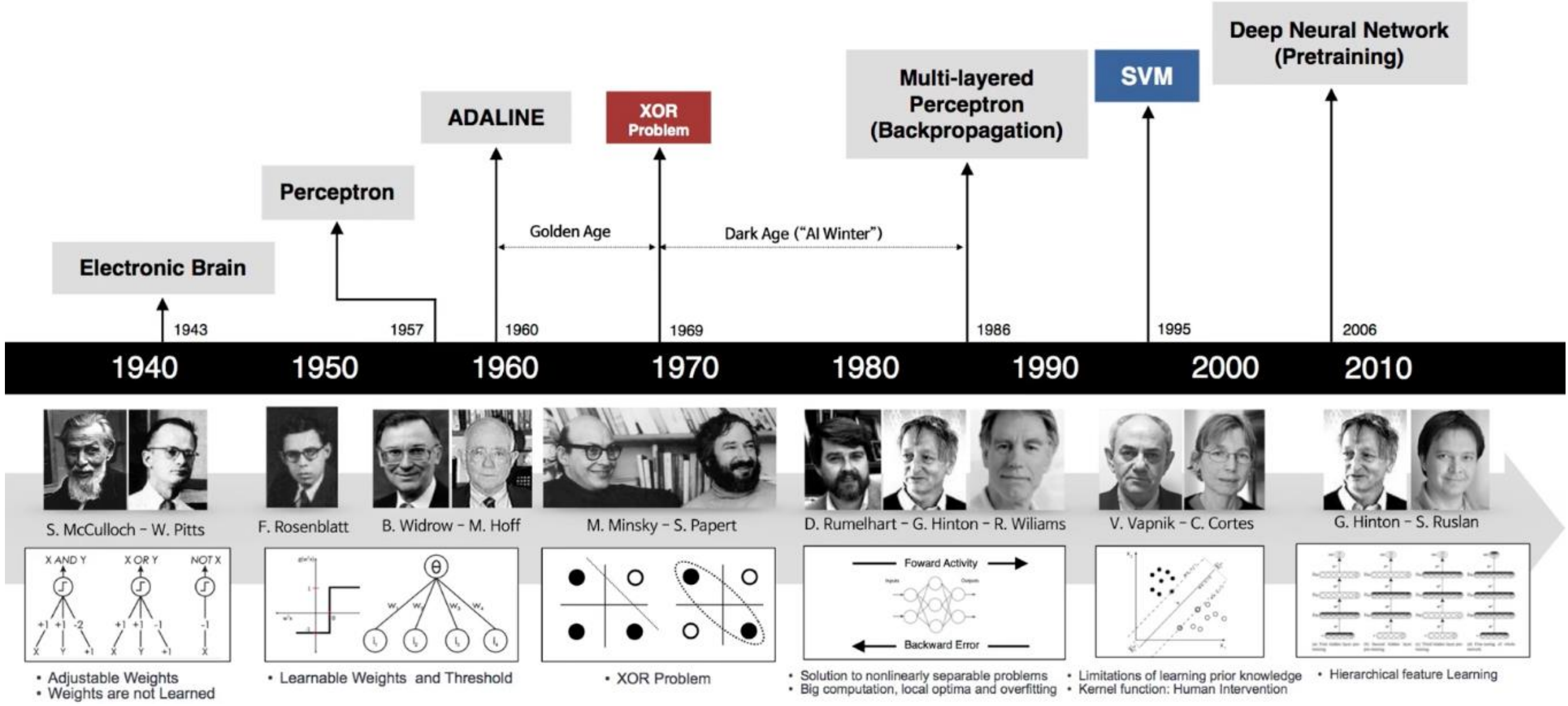


그림 1.3: 한 입력을 한 출력으로 사상하는 계산 그래프. 그래프의 각 노드는 하나의 연산을 수행한다. 깊이는 입력에서 출력으로의 가장 긴 경로의 길이인데, 그 길이는 가능한 계산 단계를 구성하는 요소들이 어떤 것인가에 따라 달라진다. 그림의 그래프들이 나타내는 계산은 로지스틱 회귀 모형 $\sigma(\boldsymbol{w}^T\boldsymbol{x})$ 의 출력인데, 여기서 σ 는 로그 S자형(logistic sigmoid) 함수이다. 왼쪽 그래프처럼 덧셈과 곱셈, 로그 S자형 함수를 컴퓨터 언어의 요소들로 사용한다면 이 모형의 깊이는 3이다. 그러나 로지스틱 회귀 자체를 하나의 요소로 간주한다면 모형의 깊이는 1이다.

심층 학습의 역사적 추세

- 1940 ~ 1960 사이버네틱스
- (1차 침체기)
- 1980 ~ 1990 연결주의
- (2차 침체기)
- 2006 ~ 현재 심층학습



2장 선형대수

2.1 스칼라, 벡터, 행렬, 텐서

- 스칼라: 하나의 수
- 벡터: 여러 개의 수를 특정한 순서로 나열한 것

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

- 행렬: 수들을 2차원으로 배열한 것
  - 전치(transpose)

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}. \qquad (\boldsymbol{A}^\top)_{i,j} = A_{j,i}.$$

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^\top = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

- 텐서: 2차원 행렬을 일반화해서, 수들의 배열을 임의의 개수의 축을 가진 정규 격자 형태로 배치한 것

2.2 행렬과 벡터의 곱셈

Matrix product 표현 방법

$$C = AB. \tag{2.4}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}. \tag{2.5}$$

Hadamard product: 성분끼리 곱하는 것

$$A \odot B$$

$$A(B + C) = AB + AC. \tag{2.6}$$

$$A(BC) = (AB)C. \tag{2.7}$$

$$x^\top y = y^\top x. \tag{2.8}$$

$$(AB)^\top = B^\top A^\top. \tag{2.9}$$

$$x^\top y = \left(x^\top y\right)^\top = y^\top x. \tag{2.10}$$

일차연립방적식 표현 방법

$$Ax = b \tag{2.11}$$

$$A_{1,:} x = b_1 \tag{2.12}$$

$$A_{2,:} x = b_2 \tag{2.13}$$

$$\dots \tag{2.14}$$

$$A_{m,:} x = b_m \tag{2.15}$$

$$A_{1,1} x_1 + A_{1,2} x_2 + \dots + A_{1,n} x_n = b_1 \tag{2.16}$$

$$A_{2,1} x_1 + A_{2,2} x_2 + \dots + A_{2,n} x_n = b_2 \tag{2.17}$$

$$\dots \tag{2.18}$$

$$A_{m,1} x_1 + A_{m,2} x_2 + \dots + A_{m,n} x_n = b_m. \tag{2.19}$$

2.2 행렬과 벡터의 곱셈

Matrix product 표현 방법

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}. \tag{2.4}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}. \tag{2.5}$$

Hadamard product: 성분끼리 곱하는 것

$$A \odot B$$

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}\boldsymbol{C}. \tag{2.6}$$

$$\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}. \tag{2.7}$$

$$\boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x}. \tag{2.8}$$

$$(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top. \tag{2.9}$$

$$\boldsymbol{x}^\top \boldsymbol{y} = \left(\boldsymbol{x}^\top \boldsymbol{y}\right)^\top = \boldsymbol{y}^\top \boldsymbol{x}. \tag{2.10}$$

일차연립방적식 표현 방법

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.11}$$

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1 \tag{2.12}$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2 \tag{2.13}$$

$$\dots \tag{2.14}$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m \tag{2.15}$$

$$\boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \dots + \boldsymbol{A}_{1,n}x_n = b_1 \tag{2.16}$$

$$\boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \dots + \boldsymbol{A}_{2,n}x_n = b_2 \tag{2.17}$$

$$\dots \tag{2.18}$$

$$\boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \dots + \boldsymbol{A}_{m,n}x_n = b_m. \tag{2.19}$$

2.2 행렬과 벡터의 곱셈

Matrix product 표현 방법

$$C = AB. \tag{2.4}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}. \tag{2.5}$$

Hadamard product: 성분끼리 곱하는 것

$$A \odot B$$

$$A(B + C) = AB + AC. \tag{2.6}$$

$$A(BC) = (AB)C. \tag{2.7}$$

$$x^\top y = y^\top x. \tag{2.8}$$

$$(AB)^\top = B^\top A^\top. \tag{2.9}$$

$$x^\top y = \left(x^\top y\right)^\top = y^\top x. \tag{2.10}$$

일차연립방적식 표현 방법

$$Ax = b \tag{2.11}$$

$$A_{1,:}x = b_1 \tag{2.12}$$

$$A_{2,:}x = b_2 \tag{2.13}$$

$$\dots \tag{2.14}$$

$$A_{m,:}x = b_m \tag{2.15}$$

$$A_{1,1}x_1 + A_{1,2}x_2 + \dots + A_{1,n}x_n = b_1 \tag{2.16}$$

$$A_{2,1}x_1 + A_{2,2}x_2 + \dots + A_{2,n}x_n = b_2 \tag{2.17}$$

$$\dots \tag{2.18}$$

$$A_{m,1}x_1 + A_{m,2}x_2 + \dots + A_{m,n}x_n = b_m. \tag{2.19}$$

2.2 행렬과 벡터의 곱셈

Matrix product 표현 방법

$$\boldsymbol{C} = \boldsymbol{AB}. \tag{2.4}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}. \tag{2.5}$$

Hadamard product: 성분끼리 곱하는 것

$$A \odot B$$

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}. \tag{2.6}$$

$$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}. \tag{2.7}$$

$$\boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x}. \tag{2.8}$$

$$(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top. \tag{2.9}$$

$$\boldsymbol{x}^\top \boldsymbol{y} = \left(\boldsymbol{x}^\top \boldsymbol{y}\right)^\top = \boldsymbol{y}^\top \boldsymbol{x}. \tag{2.10}$$

일차연립방적식 표현 방법

$$\boldsymbol{Ax} = \boldsymbol{b} \tag{2.11}$$

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1 \tag{2.12}$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2 \tag{2.13}$$

$$\dots \tag{2.14}$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m \tag{2.15}$$

$$\boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \dots + \boldsymbol{A}_{1,n}x_n = b_1 \tag{2.16}$$

$$\boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \dots + \boldsymbol{A}_{2,n}x_n = b_2 \tag{2.17}$$

$$\dots \tag{2.18}$$

$$\boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \dots + \boldsymbol{A}_{m,n}x_n = b_m. \tag{2.19}$$

2.3 단위행렬과 역행렬

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x} \tag{2.20}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n. \tag{2.21}$$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.22}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{2.23}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{2.24}$$

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{2.25}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

그림 2.2: 단위행렬의 예. 이것은 $\boldsymbol{I}_3$이다.

2.3 단위행렬과 역행렬

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x} \tag{2.20}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n. \tag{2.21}$$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.22}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{2.23}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{2.24}$$

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{2.25}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

그림 2.2: 단위행렬의 예. 이것은 $\boldsymbol{I}_3$이다.

2.4 일차종속과 생성공간

- 역행렬이 존재하려면 , 모든 b값에 대해 식 $Ax = b$에 정확히 하나의 해가 있어야한다.

$$z = \alpha x + (1 - \alpha) y \qquad (2.26)$$

- 선형결합과 생성공간

$$Ax = \sum_i x_i A_{:,i}. \qquad (2.27)$$

$$\sum_i c_i v^{(i)}. \qquad (2.28)$$

- 선형종속과 선형독립
  $b \in \mathbb{R}^m$에 대해 $Ax = b$에 하나의 해가 존재하려면, A의 열공간이 $\mathbb{R}^m$ 전체와 같아야 한다.
  즉, $n \geq m$ (모든 점에 대해 해가 있을 필요조건)

  열공간의 중복과 같은 종류의 중복성, 선형결합이 하나도 없는 벡터 집합

$$AA^{-1} = I. \qquad (2.29)$$

2.5 노름

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \tag{2.30}$$

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|. \tag{2.31}$$

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|. \tag{2.32}$$

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}, \tag{2.33}$$

$$\boldsymbol{x}^\top \boldsymbol{y} = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \cos\theta. \tag{2.34}$$

$$f(x) = 0 \Rightarrow x = 0$$
$$f(x+y) \le f(x) + f(y) \text{ (삼각부등식)}$$
$$\forall \alpha \in \mathbb{R}, f(\alpha x) = |\alpha| f(x)$$