기계학습과 수학

선형대수

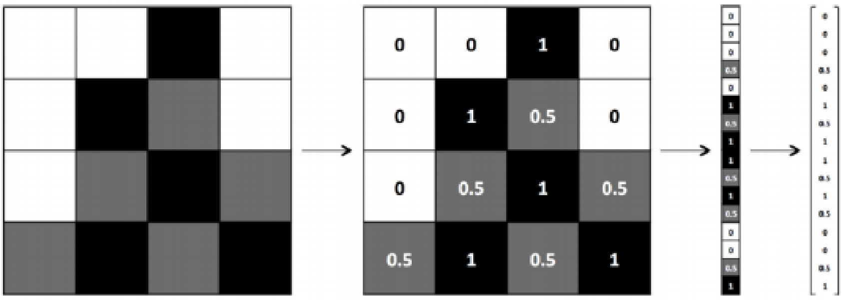
벡터와 행렬

- 벡터
  - 샘플을 **특징 벡터**로 표현

    예) Iris 데이터에서 꽃받침의 길이, 꽃받침의 너비, 꽃잎의 길이, 꽃잎의 너비라는 4개의 특징이 각각 5.1, 3.5, 1.4, 0.2인 샘플

    $$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 5.1 \\ 3.5 \\ 1.4 \\ 0.2 \end{pmatrix}$$

    예) 사진image을 벡터화하여 표현하는 과정

  - 여러 개의 특징 벡터를 첨자로 구분 (X_i)

    $$\mathbf{x}_1 = \begin{pmatrix} 5.1 \\ 3.5 \\ 1.4 \\ 0.2 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 4.9 \\ 3.0 \\ 1.4 \\ 0.2 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 4.7 \\ 3.2 \\ 1.3 \\ 0.2 \end{pmatrix}, \cdots, \mathbf{x}_{150} = \begin{pmatrix} 5.9 \\ 3.0 \\ 5.1 \\ 1.8 \end{pmatrix}$$

  - 종류와 크기 표현의 예) x ∈ R^n
- 행렬
  - 여러 개의 벡터를 담음

    요소: $X_{i,j}$, i번째 행: $X_{i,:}$, j번째 열: $X_{:,j}$

  - 훈련집합을 담은 행렬 : 설계행렬

    $$\mathbf{X} = \begin{pmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 4.7 & 3.2 & 1.3 & 0.2 \\ 4.6 & 3.1 & 1.5 & 0.2 \\ \vdots & \vdots & \vdots & \vdots \\ 6.2 & 3.4 & 5.4 & 2.3 \\ 5.9 & 3.0 & 5.1 & 1.8 \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{149,1} & x_{149,2} & x_{149,3} & x_{149,4} \\ x_{150,1} & x_{150,2} & x_{150,3} & x_{150,4} \end{pmatrix}$$

    행row

    열column

- 전치 행렬 : A (i,j) 를 A' (j,i)로 바꾸는 것
- 행렬을 이용하면 방정식을 간결하게 표현 가능하다

  예) 다항식의 행렬 표현

$$f(\mathbf{x}) = f(x_1, x_2, x_3)$$

$$= 2x_1x_1 - 4x_1x_2 + 3x_1x_3 + x_2x_1 + 2x_2x_2 + 6x_2x_3 - 2x_3x_1 + 3x_3x_2 + 2x_3x_3 + 2x_1 + 3x_2 - 4x_3 + 5$$

$$= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 2 & -4 & 3 \\ 1 & 2 & 6 \\ -2 & 3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 2 & 3 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 5$$

$$= \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + c$$

- 특수한 행렬들
  - 정사각행렬(정방행렬)
  - 대각행렬
  - 단위행렬
  - 대칭행렬

$$\text{정사각행렬} \begin{pmatrix} 2 & 0 & 1 \\ 1 & 21 & 5 \\ 4 & 5 & 12 \end{pmatrix}, \quad \text{대각행렬} \begin{pmatrix} 50 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 8 \end{pmatrix},$$

$$\text{단위행렬} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{대칭행렬} \begin{pmatrix} 1 & 2 & 11 \\ 2 & 21 & 5 \\ 11 & 5 & 1 \end{pmatrix}$$

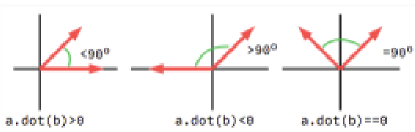
- 행렬 연산
  - 행렬 곱셈

$$\mathbf{C} = \mathbf{AB}, \text{ 이때 } c_{ij} = \sum_{k=1,s} a_{ik} b_{kj}$$

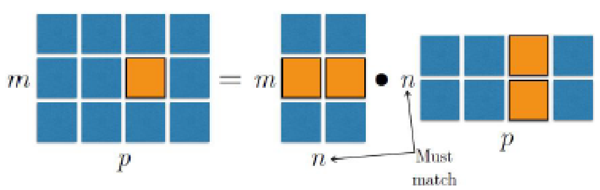
    - 교환법칙 성립 X : AB ≠ BA
    - 분배법칙 / 결합법칙 성립 : A(B+C) = AB + AC , A(BC) = (AB)C
  - 벡터의 내적(Inner Product)

$$\text{벡터의 내적 } \mathbf{a} \cdot \mathbf{b} = \mathbf{a}^{\mathrm{T}}\mathbf{b} = \sum_{k=1,d} a_k b_k \qquad (2.2)$$

$$\mathbf{x}_1 = \begin{pmatrix} 5.1 \\ 3.5 \\ 1.4 \\ 0.2 \end{pmatrix} \text{와 } \mathbf{x}_2 = \begin{pmatrix} 4.9 \\ 3.0 \\ 1.4 \\ 0.2 \end{pmatrix} \text{의 내적 } \mathbf{x}_1 \cdot \mathbf{x}_2 \text{는 } 37.49$$

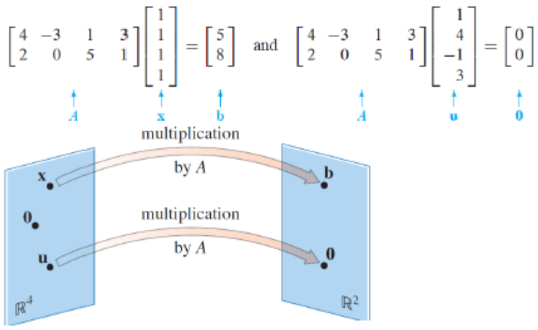


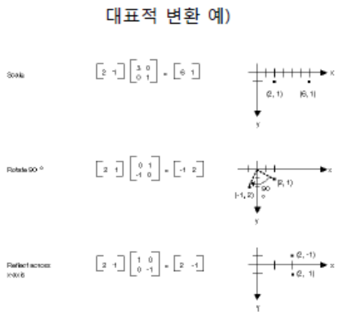
  - C = AB 예시



- $C_{i,j} = \sum_k A_{i,k} B_{k,j}$
- A: $m \times n$, B: $n \times p$, C: $m \times p$

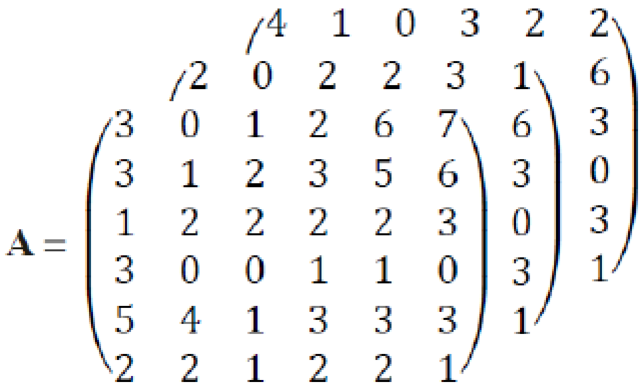
  - 행렬 곱셈을 통한 벡터의 변환(Function / Mapping) 예시





- 텐서
  - 3차원 이상의 구조를 가진 숫자 배열
    - 0차 : 수(Scalar)
    - 1차 : 벡터
    - 2차 : 행렬
    - 고차원 ...

예) 3차원 구조의 RGB 컬러 영상



놈과 유사도

- 벡터와 행렬의 크기를 놈(Norm)으로 측정
  - 벡터의 p차 놈

$$p\text{차 놈: } \|\mathbf{x}\|_p = \left(\sum_{i=1,d} |x_i|^p\right)^{\frac{1}{p}} \qquad (2.3)$$

1차 $(p=1)$ 놈, 2차 $(p=2)$ 놈 Euclidean norm, 최대 $(p = infinite)$ 놈 max norm

$$\text{최대 놈: } \|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \cdots, |x_d|) \qquad (2.4)$$

예) $\mathbf{x} = (3\ \ -4\ \ 1)$ 일 때, 2차 놈은 $\|\mathbf{x}\|_2 = (3^2 + (-4)^2 + 1^2)^{1/2} = 5.099$

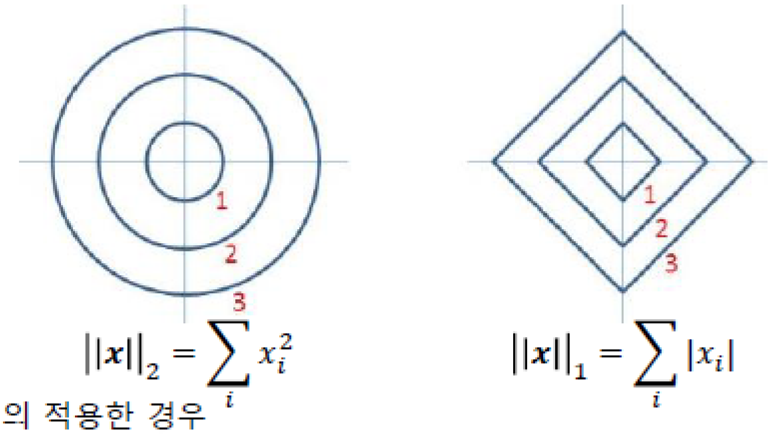
  - 행렬의 프로베니우스(Frobenius Norm) : 행렬의 크기를 측정

$$\text{프로베니우스놈: } \|\mathbf{A}\|_F = \left(\sum_{i=1,n} \sum_{j=1,m} a_{ij}^2 \right)^{\frac{1}{2}} \tag{2.6}$$

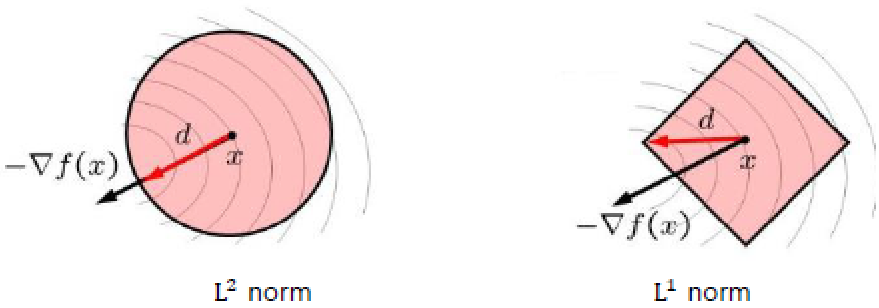
$$\text{예를 들어, } \left\| \begin{pmatrix} 2 & 1 \\ 6 & 4 \end{pmatrix} \right\|_F = \sqrt{2^2 + 1^2 + 6^2 + 4^2} = 7.550$$

- 1차 놈(Manhattan Distance, L1)과 2차 놈(Euclidean Distance, L2) 비교

크기의 차이



하강 기울기의 적용한 경우



- 유사도(Similarity)와 거리(Distance)
  - 벡터를 기하학적으로 해석

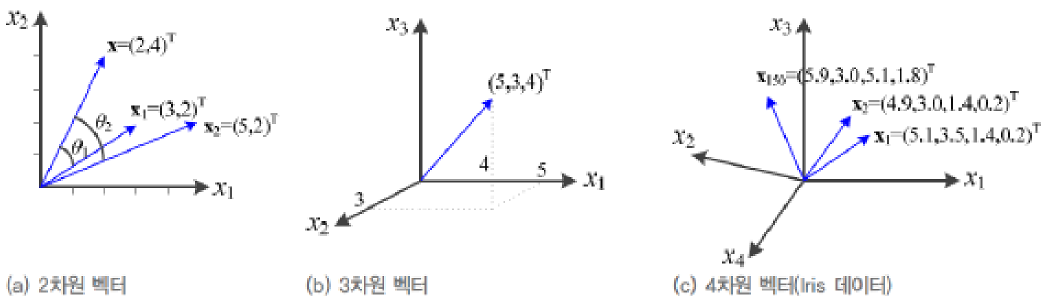


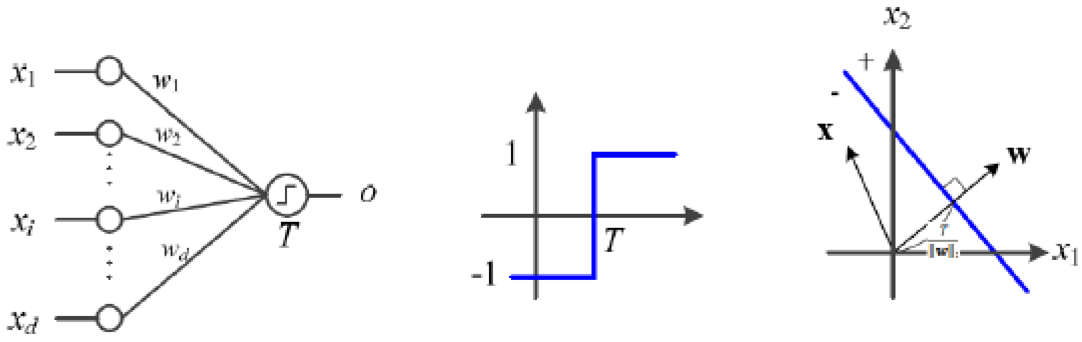
그림 2-2 벡터를 기하학적으로 해석

  - 코사인 유사도(Cosine Similarity)

$$cosine\_similarity(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}}{\|\mathbf{a}\|} \cdot \frac{\mathbf{b}}{\|\mathbf{b}\|} = cos(\theta) \tag{2.7}$$

퍼셉트론의 해석

- 퍼셉트론 : 1958년 로젠블라트가 고안한 **분류기(Classifier) 모델**



(a) 퍼셉트론 구조 (b) 계단형 활성함수(비선형) (c) 퍼셉트론의 공간 분할

그림 2-3 퍼셉트론의 구조와 동작

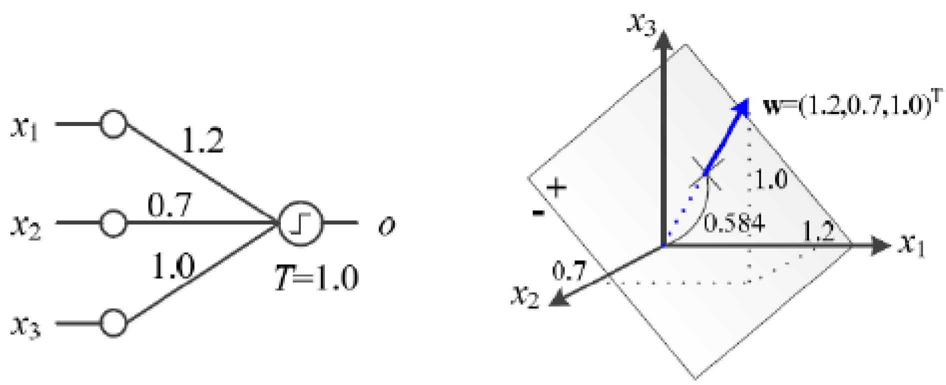
- 그림 2-3의 c의 파란 직선은 두 개의 부분공간을 나누는 결정직선(Decision Line)

$$\mathbf{w}\text{에 수직이고 원점으로부터 } \frac{T}{\|\mathbf{w}\|_2}\text{만큼 떨어져 있음}$$

- 동작을 수식으로 표현하면

$$o = \tau(\mathbf{w} \cdot \mathbf{x}), \quad \text{이때} \quad \tau(a) = \begin{cases} 1, & a \geq T \\ -1, & a < T \end{cases} \qquad (2.8)$$

  - 활성 함수(Activation Function)으로는 계단함수(Step Function) 사용
- 3차원 특징공간은 결정평면(Decision Plane), 4차원 이상은 결정 초평면(Decision Hyperplane)
  - 3차원 특징공간을 위한 퍼셉트론



(a) 퍼셉트론 (b) 공간 분할(2부류 분류)

그림 2-4 퍼셉트론의 예(3차원)

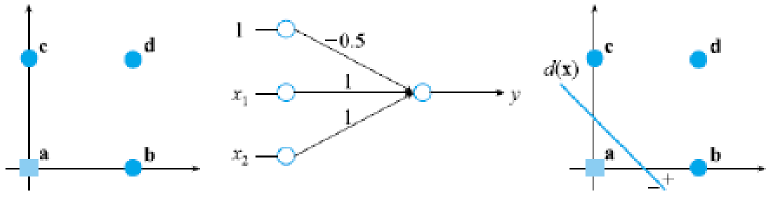
- 예제

$\mathbf{a}=(0,0)^{\mathrm{T}},\ t_{\mathbf{a}}=-1$

$\mathbf{b}=(1,0)^{\mathrm{T}},\ t_{\mathbf{b}}=1$

$\mathbf{c}=(0,1)^{\mathrm{T}},\ t_{\mathbf{c}}=1$

$\mathbf{d}=(1,1)^{\mathrm{T}},\ t_{\mathbf{d}}=1$

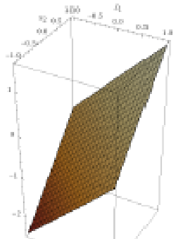


(a) OR 분류 문제 (b) OR 분류기로서 퍼셉트론 (c) 퍼셉트론은 선형 분류기

그림 4.3 퍼셉트론의 예

이 퍼셉트론은 $\mathbf{w}=(1,1)^{\mathrm{T}}$, b=-0.5

따라서 결정 직선은 $d(\mathbf{x})=x_1+x_2-0.5$

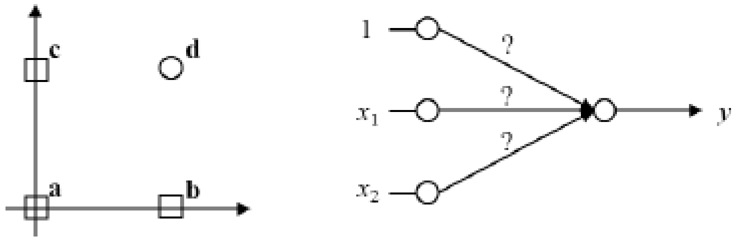


- 샘플 a를 맞추나 보자? $y=\tau(\mathbf{w}^{\mathrm{T}}\mathbf{c}+b)=\tau((1,1)\begin{pmatrix}0\\1\end{pmatrix}-0.5)=\tau(0.5)=1$
- 나머지 샘플 b, c, d도 맞추는가?

- AND 분류 문제는?

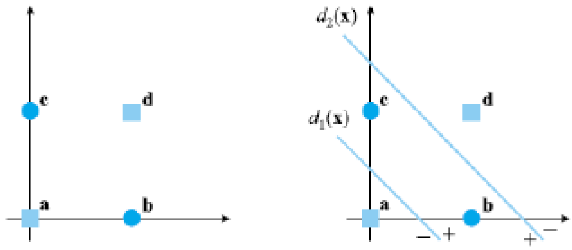
$\mathbf{a}=(0,0)^{\mathrm{T}}$ $\mathbf{b}=(1,0)^{\mathrm{T}}$ $\mathbf{c}=(0,1)^{\mathrm{T}}$ $\mathbf{d}=(1,1)^{\mathrm{T}}$

t_a= -1 t_b= -1 t_c= -1 t_d=1



- XOR 분류 문제는?
  - 퍼셉트론은 75% 분류 한계
  - 이 한계를 어떻게 극복?

  → 두 개의 퍼셉트론 (결정 직선) 사용



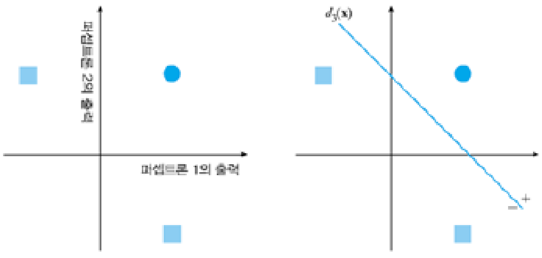
(a) XOR 분류 문제 (b) 두 개의 직선으로 해결

- 다층 퍼셉트론(**M**ulti-**L**ayer **P**erceptron)
  - 두 단계에 걸쳐 문제해결
    - 단계 1 : 원래 특징 공간을 새로운 공간으로 매핑
    - 단계 2 : 새로운 공간에서 분류
  - XOR의 경우, 다음과 같은 조건을 활용

$$\left.\begin{array}{l}\mathbf{w}_1^{\mathrm{T}}\mathbf{x}+b_1>0\text{이고 }\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}+b_2>0\text{이면, }\ \mathbf{x}\in\omega_1\\ \mathbf{w}_1^{\mathrm{T}}\mathbf{x}+b_1<0\text{이거나 }\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}+b_2<0\text{이면, }\ \mathbf{x}\in\omega_2\end{array}\right\}$$

두 단계로 XOR 문제 해결

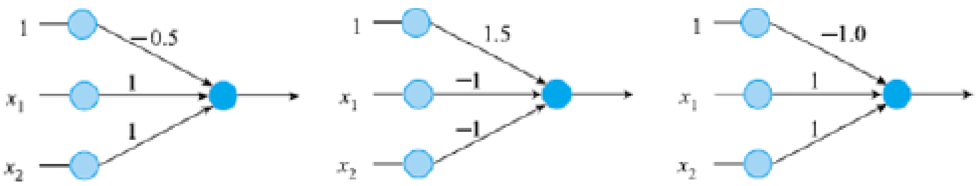
샘플	특징 벡터 (x)		첫 번째 단계		두 번째 단계
	x_1	x_2	퍼셉트론1	퍼셉트론2	퍼셉트론3
a	0	0	−1	+1	−1
b	1	0	+1	+1	+1
c	0	1	+1	+1	+1
d	1	1	+1	−1	−1



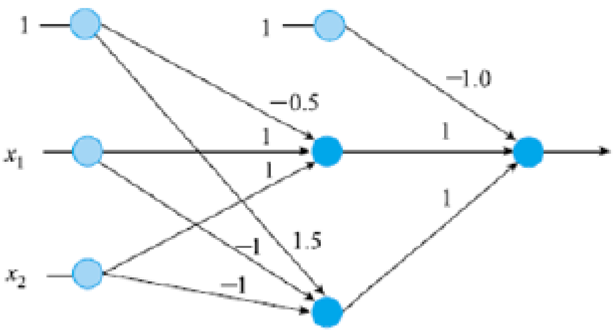
(a) 매핑 공간에서의 샘플 분포 (b) 퍼셉트론3에 의한 영역 분할

새로운 공간에서의 샘플 분포와 영역 분할

- 다층 퍼셉트론을 활용한 XOR 분류 문제



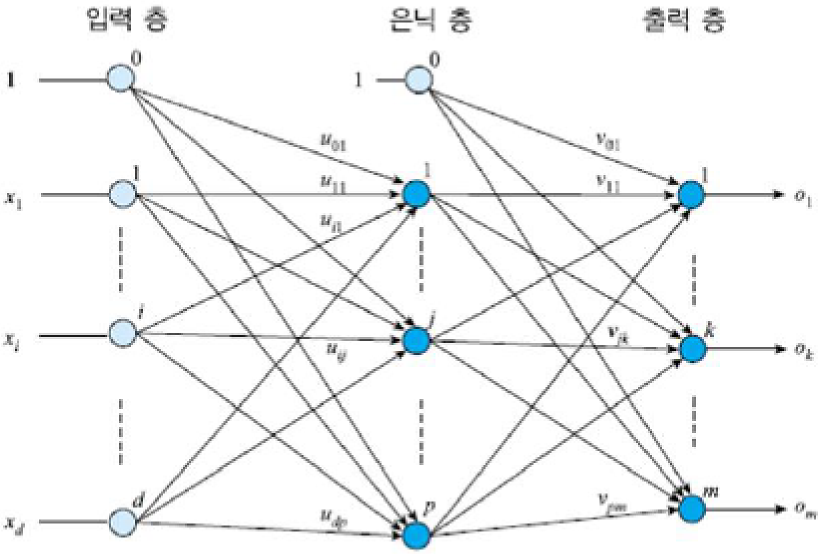
(a) 퍼셉트론1 (b) 퍼셉트론2 (c) 퍼셉트론3



(d) 다층 퍼셉트론

세 개의 퍼셉트론과 이들을 연결하여 만든 다층 퍼셉트론

- 다층 퍼셉트론의 구조(입력층, 은닉층, 출력층)



다층 퍼셉트론의 구조와 표기

- 출력이 여러 개인 퍼셉트론의 표현

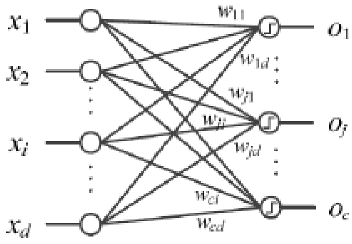


그림 2-5 출력이 여러 개인 퍼셉트론

출력은 벡터 $\mathbf{o} = (o_1, o_2, \cdots, o_c)^{\mathrm{T}}$로 표기

j번째 퍼셉트론의 가중치 벡터를 $\mathbf{w}_j = (w_{j1}, w_{j2}, \cdots, w_{jd})^{\mathrm{T}}$와 같이 표기

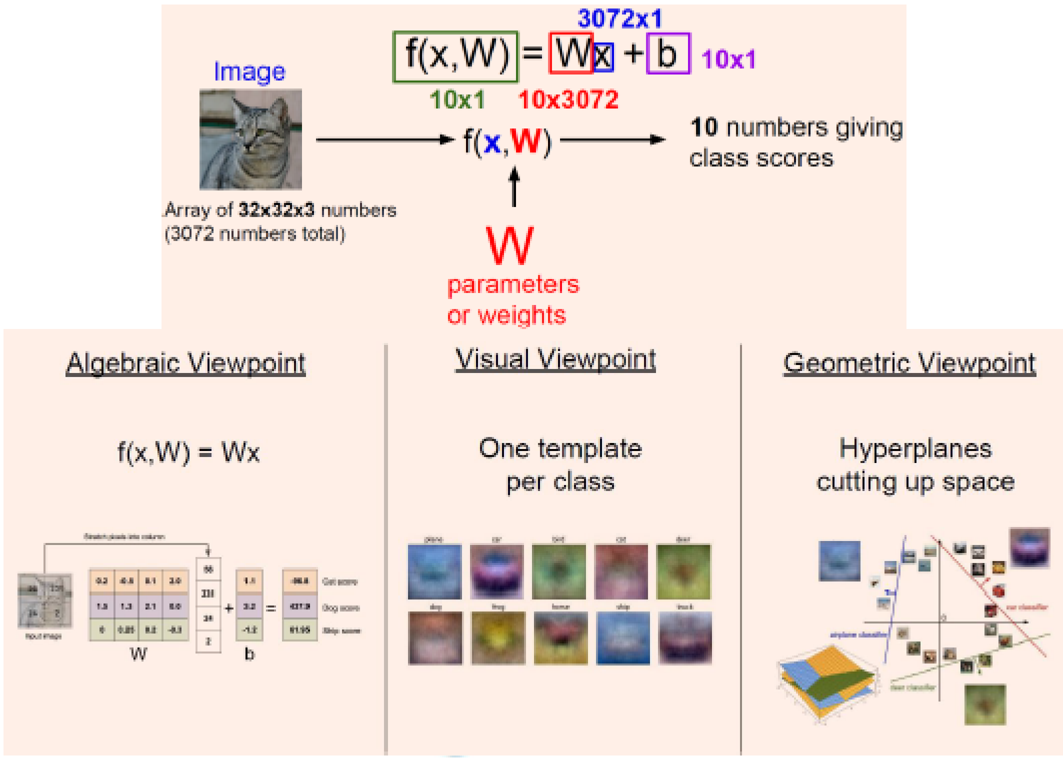
동작을 수식으로 표현하면,

$$\mathbf{o} = \boldsymbol{\tau}\begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{x} \\ \mathbf{w}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{w}_c \cdot \mathbf{x} \end{pmatrix} \longrightarrow \text{행렬로 간결하게 쓰면 } \mathbf{o} = \boldsymbol{\tau}(\mathbf{W}\mathbf{x})$$

$$\text{이때 } \mathbf{W} = \begin{pmatrix} \mathbf{w}_1^{\mathrm{T}} \\ \mathbf{w}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{w}_c^{\mathrm{T}} \end{pmatrix}$$

가중치 벡터를 각 부류의 기준 벡터로 간주하면, c개 부류의 유사도를 계산하는 셈

- 선형 분류기 이해



- 학습의 정의
  - 추론(Inferring) : 식(2,10)은 학습을 마친 알고리즘을 현장의 새로운 데이터에 적용했을 때 일어나는 과정

$$\text{분류라는 과업: } \overset{?}{\mathbf{o}} = \boldsymbol{\tau}(\overset{\text{앎}}{\mathbf{W}}\,\overset{\text{앎}}{\mathbf{x}}) \tag{2.10}$$

  - 학습(Learning) : 훈련집합의 샘플에 대해 식(2.11)을 가장 잘 만족하는 W를 찾아내는 과정

$$\text{학습이라는 과업: } \overset{\text{앎}}{\mathbf{o}} = \boldsymbol{\tau}(\overset{?}{\mathbf{W}}\,\overset{\text{앎}}{\mathbf{x}}) \tag{2.11}$$

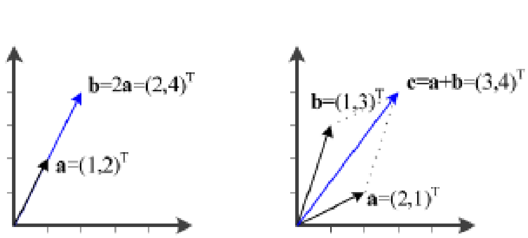
- 현대 기계 학습에서 퍼셉트론의 중요성
  - 딥러닝은 퍼셉트론을 여러 층으로 확장하여 만듦

선형결합과 벡터공간

- 벡터 : 공간상의 한 점으로 화살표 끝이 벡터의 좌표에 해당
- 선형결합이 만드는 벡터공간
  - 기저(basis)벡터 a와 b의 선형 결합(Linear Combination)

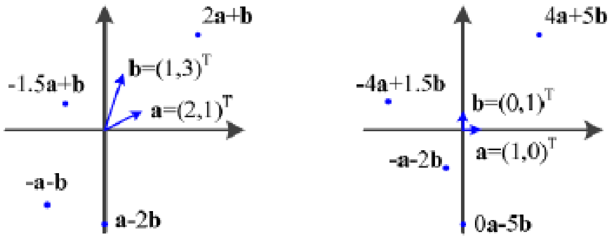
$$\mathbf{c} = \alpha_1 \mathbf{a} + \alpha_2 \mathbf{b} \tag{2.12}$$

  - 선형결합으로 만들어지는 공간을 벡터공간(Vector Space)이라 부름



(a) 벡터에 스칼라 곱 (b) 두 벡터의 덧셈

그림 2-6 벡터의 연산

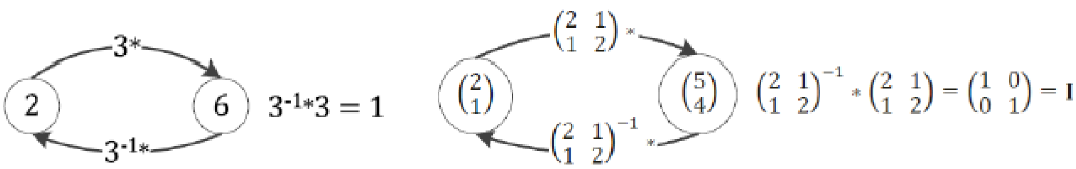


(a) 기저 벡터와 벡터공간 (b) 정규직교 기저 벡터

그림 2-7 벡터공간

역행렬

- 원리



(a) 역수의 원리 (b) 역행렬의 원리

그림 2-9 역행렬

  - 정사각행렬 $\mathbf{A}$의 역행렬 $\mathbf{A}^{-1}$

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

  - 예를 들어, $\begin{pmatrix} 2 & 1 \\ 6 & 4 \end{pmatrix}$의 역행렬은 $\begin{pmatrix} 2 & -0.5 \\ -3 & 1 \end{pmatrix}$

- 역행렬을 활용한 방정식 표현과 해

- 방정식: $\mathbf{Ax} = \mathbf{b}$의 확장
  - $\mathbf{A} \in \mathbb{R}^{m \times n}$: 알고 있는 행렬
  - $\mathbf{b} \in \mathbb{R}^{n}$: 알고 있는 벡터
  - $\mathbf{x} \in \mathbb{R}^{m}$: 알고 싶은 알지 못한 벡터

$$A_{1,:}\mathbf{x} = A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n = b_1$$
$$A_{2,:}\mathbf{x} = A_{2,1}x_1 + A_{22}x_2 + \cdots + A_{2,n}x_n = b_2$$
$$\cdots$$
$$A_{m,:}\mathbf{x} = A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n = b_m$$

- 선형 방정식의 경우
  - 불능: 해 없음
  - 부정: 다수의 해 존재
  - 유일해 존재 → 역행렬을 이용하여 해를 구함 →

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b} \\ \mathbf{A}^{-1}\mathbf{Ax} &= \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{I_n x} &= \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \end{aligned}$$

- 정리

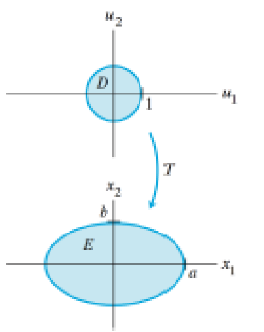
정리 2-1 다음 성질은 서로 필요충분조건이다.

- **A**는 역행렬을 가진다. 즉, 특이행렬이 아니다.
- **A**는 최대계수를 가진다.
- **A**의 모든 행이 선형독립이다.
- **A**의 모든 열이 선형독립이다.
- **A**의 행렬식은 0이 아니다.
- $\mathbf{A}^T\mathbf{A}$는 양의 정부호positive definite 대칭 행렬이다.
- **A**의 고윳값은 모두 0이 아니다.

- 행렬식(**Determinant**)

$$\left.\begin{aligned} det\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= ad - bc \\ det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} &= aei + bfg + cdh - ceg - bdi - afh \end{aligned}\right\} \quad (2.15)$$

예를 들어 $\begin{pmatrix} 2 & 1 \\ 6 & 4 \end{pmatrix}$의 행렬식은 2*4−1*6=2



{변환된 E 공간 부피}
= |det(**A**)|·{원래 D 공간 부피}

- 기하학적 의미
  - 행렬식의 절대값은 주어진 행렬의 곱에 의한 공간의 확장 또는 축소로 볼 수 있음
    - If det(A) = 0, 하나의 차원을 따라 축소되어 공간의 부피를 잃게 됨

- If det(A) = 1, 공간의 부피를 유지한 변환
  - 차원에서의 기하학적인 예시
    - 2차원에서는 2개의 행 벡터가 이루는 평행사변형 넓이
    - 3차원에서는 3개의 행 벡터가 이루는 평행사각기둥의 부피

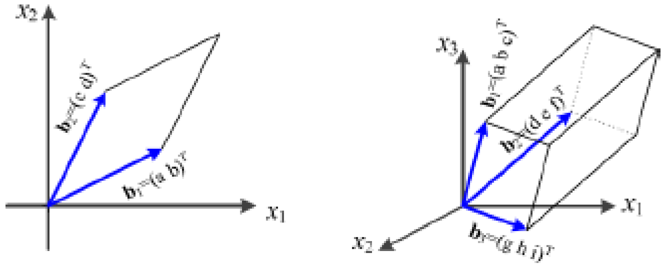


그림 2-10 행렬식의 기하학적 해석

- 정부호(**Definiteness**) 행렬

양의 정부호 행렬 : **0**이 아닌 모든 벡터 **x**에 대해, $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}>0$

- 예를 들어, $\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}$는 $(x_1 \quad x_2)\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} = x_1^2 + 2x_2^2$ 이므로

$\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}$는 양의 정부호 행렬

  - 종류

양의 준정부호positive semi-definite 행렬: **0**이 아닌 모든 벡터 **x**에 대해, $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\geq 0$

음의 정부호negative definite 행렬: **0**이 아닌 모든 벡터 **x**에 대해, $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}<0$

음의 준정부호negative semi-definite 행렬:**0**이 아닌 모든 벡터 **x**에 대해, $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\leq 0$

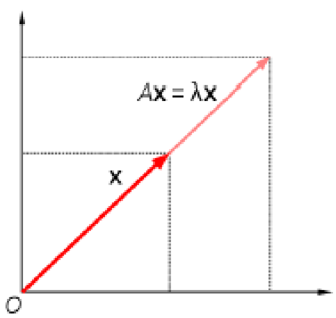
행렬 분해

- **분해**(Decomposition)란?
  - 정수 3717은 특성이 보이지 않지만, 3*3*7*59로 소인수분해하면 특성이 보이듯 분해하는 것
- 고윳값(**Eigenvalue**)과 고유 벡터(**Eigenvector**)

고유 벡터 **v**와 고윳값 λ

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \qquad (2.20)$$

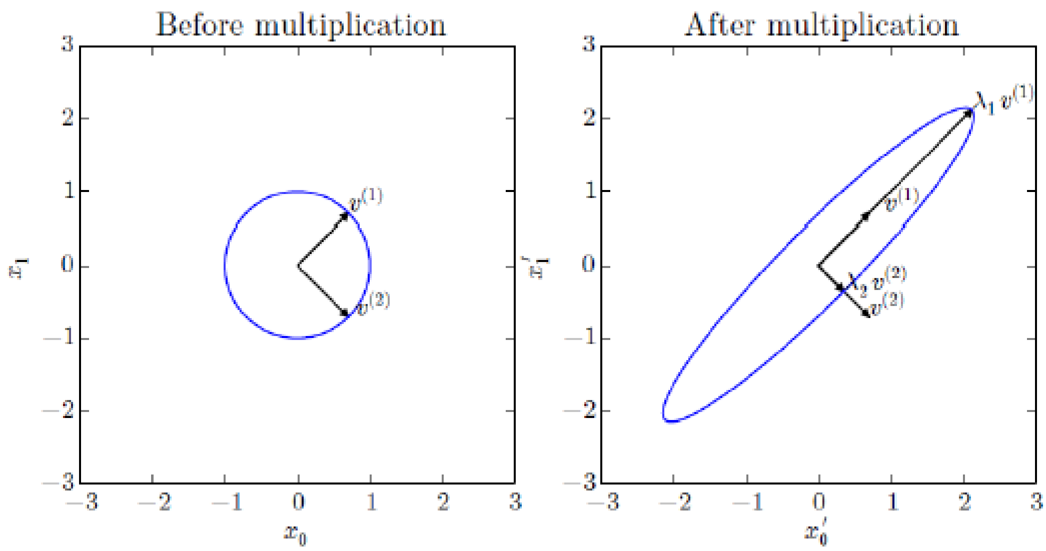
- 2차원 공간에서의 고윳값과 고유 벡터의 기하학적 해석



예를 들어, $\begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = 3\begin{pmatrix}1\\1\end{pmatrix}$이고 $\begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix} = 1\begin{pmatrix}1\\-1\end{pmatrix}$이므로,

$\lambda_1 = 3, \lambda_2 = 1$이고 $\mathbf{v}_1 = \begin{pmatrix}1\\1\end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix}1\\-1\end{pmatrix}$

- 고윳값의 효과



- 왼쪽: 원으로 표현된 단위 벡터 $\mathbf{u} \in \mathbb{R}^2$의 모든 점
- 오른쪽: 행렬 $\mathbf{A}$의 곱에 의한 $\mathbf{Au}$ 모든 점, $\mathbf{A}$는 원을 고유 벡터 방향으로는 고유값만큼 크기 변환만 시킴

- 고윳값과 고유 벡터의 기하학적 해석

예제 2-5

[그림 2-12]의 반지름이 1인 원 위에 있는 4개의 벡터 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$가 $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$에 의해 어떻게 변환되는지 살펴보자. 변환 후의 벡터를 각각 $\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3', \mathbf{x}_4'$로 표기한다.

$$\mathbf{x}_1' = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 3/\sqrt{2} \\ 3/\sqrt{2} \end{pmatrix}$$
$$\mathbf{x}_2' = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$
$$\mathbf{x}_3' = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$
$$\mathbf{x}_4' = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

눈 여겨 볼 점은 $\mathbf{A}$의 고유 벡터 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}$과 방향이 같은 $\mathbf{x}_1$과 $\mathbf{x}_3$이다. 이들은 변환 때문에 길이가 달라지더라도 방향은 그대로 유지한다. 식 (2.20)을 충실히 따르고 있다. 이때 길이의 변화는 고윳값 λ에 따른다. 즉, $\mathbf{x}_1$은 3배만큼, $\mathbf{x}_3$은 1배만큼 길이가 변한다. 나머지 $\mathbf{x}_2$와 $\mathbf{x}_4$는 길이와 방향이 모두 변한다. 파란 원 위에 있는 모든 점을 변환하면 빨간색의 타원이 된다. 파란 원 위에 존재하는 무수히 많은 점(벡터) 중에 방향이 바뀌지 않는 것은 고유 벡터에 해당하는 $\mathbf{x}_1$과 $\mathbf{x}_3$뿐이다.

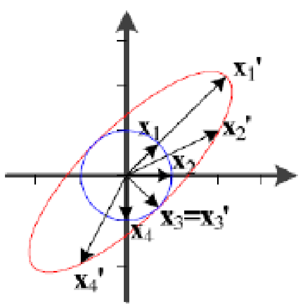


그림 2-12 고유 벡터의 공간 변환

- 고유 분해(**Eigen-decomposition**)

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1} \qquad (2.21)$$

- **Q**는 **A**의 고유 벡터를 열에 배치한 행렬이고 **Λ**는 고윳값을 대각선에 배치한 대각행렬

$$\begin{aligned} A &= Q\Lambda Q^T \\ &= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T + \cdots + \lambda_n v_n v_n^T \end{aligned}$$

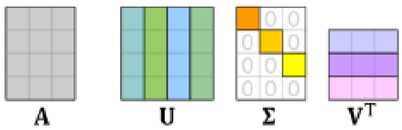
- 예를 들어, $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$
- 고유 분해는 고유값과 해당 고유 벡터가 존재하는 정사각행렬에만 적용 가능
- 하지만, 기계 학습에서는 정사각행렬이 아닌 경우의 분해도 필요하므로

  고유 분해는 한계를 가짐

- n*m 행렬 A의 **특잇값 분해**(**S**ingular **V**alue **D**ecomposition)

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}} \qquad (2.22)$$

- 왼쪽 특이행렬 **U**는 $\mathbf{AA}^{\mathbf{T}}$의 고유 벡터를 열에 배치한 n*n 행렬
- 오른쪽 특이행렬 **V**는 $\mathbf{A}^{\mathbf{T}}\mathbf{A}$의 고유 벡터를 열에 배치한 m*m 행렬
- **Σ**는 $\mathbf{AA}^{\mathbf{T}}$의 고윳값의 제곱근을 대각선에 배치한 n*m 대각행렬

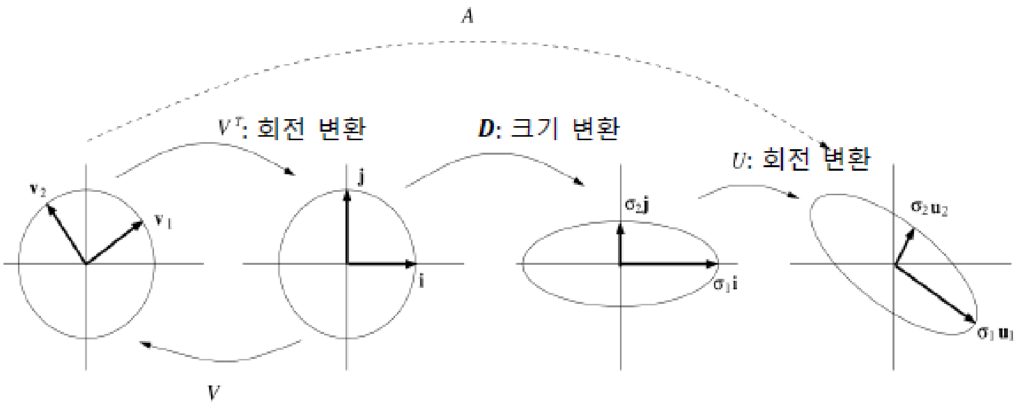


예를 들어, **A**를 4*3 행렬이라고 했을 때 다음과 같이 특잇값 분해가 된다.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 3 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} = \begin{pmatrix} -0.1914 & -0.2412 & 0.1195 & -0.9439 \\ -0.5144 & 0.6990 & -0.4781 & -0.1348 \\ -0.6946 & -0.6226 & -0.2390 & 0.2697 \\ -0.4651 & 0.2560 & 0.8367 & 0.1348 \end{pmatrix}$$

$$\begin{pmatrix} 3.7837 & 0 & 0 \\ 0 & 2.7719 & 0 \\ 0 & 0 & 1.4142 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -0.7242 & -0.4555 & -0.5177 \\ -0.6685 & 0.2797 & 0.6891 \\ 0.1690 & -0.8452 & 0.5071 \end{pmatrix}$$

- 기하학적 해석



- 정사각행렬이 아닌 행렬의 역행렬을 구하는데 사용됨

SVD Pseudoinverse

$$A = U \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_s \\ & & 0 \end{pmatrix} V^T \longrightarrow A^+ = V \begin{pmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_s & 0 \end{pmatrix} U^T$$