기계학습 (Machine Learning)

- Machine Learning and Python/Mathematics

정보통신공학과

Prof. Jinkyu Kang

수업의 목차

- **기계학습과 Python**
  - 라이브러리 및 도구들
  - 구글 코랩 (Colab)

- **기계학습과 수학**
  - 선형 대수
  - 확률과 통계
  - 최적화 이론

기계학습과 Python |

- 파이썬(Python)은 데이터 과학 분야를 위한 표준 프로그래밍 언어
  - 범용 프로그래밍 언어의 장점 + MATLAB과 R 같은 특정 분야를 위한 스크립팅 언어의 편리함
  - 다양한 라이브러리
    - 데이터 적재, 시각화, 통계, 자연어 처리, 이미지 처리 등에 필요한 라이브러리 존재
  - 터미널이나 주피터 노트북(Jupyter Notebook) 같은 도구로 대화하듯 프로그래밍
  - 머신러닝과 데이터 분석은 데이터 주도 분석이라는 점에서 근본적으로 반복 작업, 따라서 반복 작업을 빠르게 처리하고 손쉽게 조작할 수 있는 도구가 필수
  - 범용 프로그래밍 언어로서 파이썬은 복잡한 그래픽 사용자 인터페이스(GUI)나 웹 서비스도 만들 수 있으며 기존 시스템과 통합하기도 좋음

기계학습과 Python | 라이브러리 및 도구들

- 오픈 소스인 사이킷런(scikit-learn)
  - 다양한 머신러닝 알고리즘 + 알고리즘을 설명한 풍부한 문서 제공
    - http://scikit-learn.org/stable/documentation
  - 매우 인기가 높고 독보적인 파이썬 머신러닝 라이브러리
  - 산업 현장이나 학계에도 널리 사용되고 많은 튜토리얼과 예제 코드
  - 사이킷런은 다른 파이썬의 과학 패키지들과도 잘 연동

- 사이킷런 설치
  - scikit-learn은 두 개의 다른 파이썬 패키지인 넘파이(NumPy)와 사이파이(SciPy)를 사용
  - 그래프를 그리려면 맷플롯립(matplotlib)을, 대화식으로 개발하려면 아이파이썬(Ipython)과 주피터 노트북도 설치해야 함
  - 필요한 패키지들을 모아 놓은 파이썬 배포판을 설치하는 방법을 권장
    - Anaconda: 대용량 데이터 처리, 예측 분석, 과학 계산용 파이썬 배포판

기계학습과 Python | 라이브러리 및 도구들

- 주피터 노트북
  - 주피터 노트북은 프로그램 코드를 브라우저에서 실행해주는 대화식 환경을 제공
- NumPy
  - 파이썬으로 과학 계산을 하려면 꼭 필요한 패키지임. 다차원 배열을 위한 기능과 선형 대수 연산과 푸리에 변환 같은 고수준 수학 함수와 유사(pseudo) 난수 생성기를 포함
- SciPy
  - 과학 계산용 함수를 모아놓은 파이썬 패키지임. SciPy는 고성능 선형 대수, 함수 최적화, 신호 처리, 특수한 수학 함수와 통계 분포 등을 포함한 많은 기능을 제공
- matplotlib
  - 파이썬의 대표적인 과학 계산용 그래프 라이브러리임. 선 그래프, 히스토그램, 산점도 등을 지원하며 출판에 쓸 수 있을 만큼의 고품질 그래프를 그려줌
- pandas
  - 데이터 처리와 분석을 위한 파이썬 라이브러리임

기계학습과 Python | 구글 코랩 (Colab)

- **Google Colab (colaboratory)**
  - 브라우저에서 python을 작성하고 실행 가능
  - 클라우드 기반으로 주피터 노트북 개발환경
  - https://colab.research.google.com/notebooks/intro.ipynb

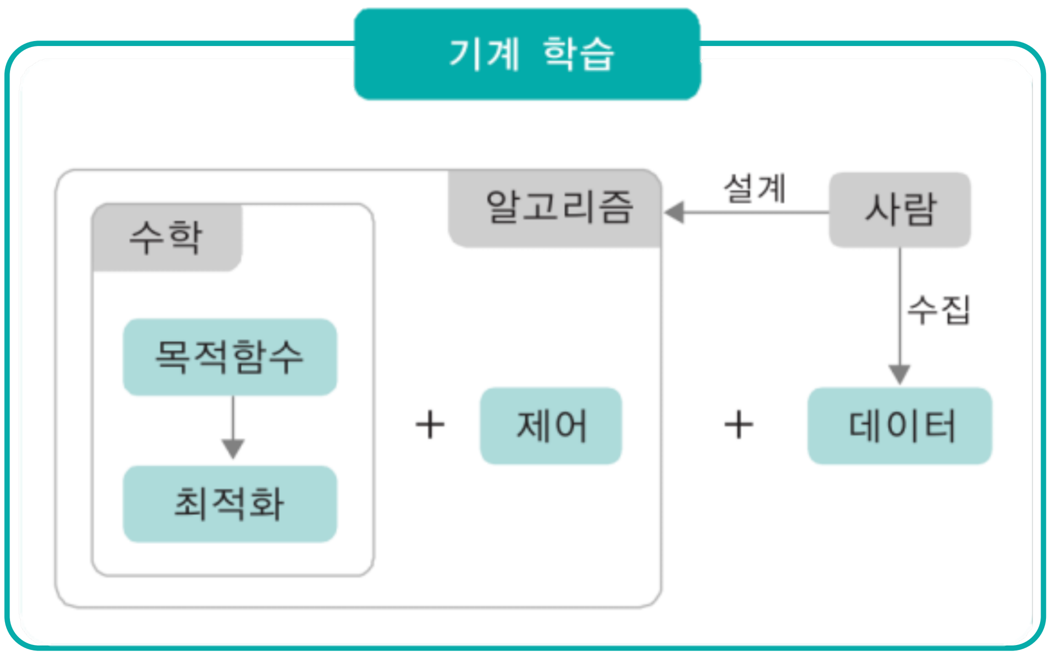
- **Colab 특징**
  - 별도의 파이썬 설치가 필요 없음
  - 데이터에 분석 사용되는 Tensor Flow, Keras, mataplotlib, scikit-learn, pandas와 같은 라이브러리가 기본적으로 설치
  - GPU를 무료로 사용 가능
  - Jupyter 노트북과 비슷하지만 더 좋은 기능을 제공
  - 깃과 연동이 가능하여 사람들과 협업하여 코딩이 가능

수업의 목차

- **기계학습과 수학**
  - 선형 대수
  - 확률과 통계
  - 최적화 이론

기계학습과 수학

- **기계 학습에서 수학의 역할**
  - 수학은 이론적인 배경과 알고리즘을 구성하는 기본적인 근간
  - 알고리즘의 원리를 이해하기 위해 필요
  - 사람은 알고리즘을 설계하고 데이터를 수집
  - 기계학습은 수학/알고리즘/사람이 수집하는 데이터 로 이루어진다.



기계학습과 수학

- 선형대수학: 어떻게 조사 대상을 형식화할 것인가?
  - 학습 모델의 매개변수집합, 데이터, 선형연산의 결합 등을 벡터 또는 행렬로 간결하게 표현 가능
  - 데이터를 분석하여 유용한 정보를 알아내거나 특징 공간을 변환 수행
- 확률과 통계: 데이터의 특징을 어떻게 알 수 있을까?
  - 데이터에 포함된 불확실성을 표현하고 처리
  - 데이터의 특징들을 추출하고 많은 데이터의 중요한 속성만 간추리는 역할
  - 모델 설계에 활용
- 최적화 이론: 알고리즘을 어떻게 이해하고, 훈련 과정을 최적화 할 수 있을까?
  - 목적함수를 최소화하는 최적해를 찾는데 활용 (주로 미분을 활용)

선형대수

- 벡터와 행렬
- 놈과 유사도
- 퍼셉트론의 해석
- 선형결합과 벡터공간
- 역행렬
- 행렬 분해

선형대수 | 벡터와 행렬

- 벡터
  - 샘플을 특징 벡터로feature vector 표현
  - 예) Iris 데이터에서 꽃받침의 길이, 꽃받침의 너비, 꽃잎의 길이, 꽃잎의 너비라는 4개의 특징이 각각 5.1, 3.5, 1.4, 0.2인 샘플

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

  - 여러 개의 특징 벡터를 첨자로 구분

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

선형대수 | 벡터와 행렬

- 행렬
  - 여러 개의 벡터를 담음
  - 훈련집합을 담은 행렬을 설계행렬이라 부름
  - 예) Iris 데이터에 있는 150개의 샘플을 설계 행렬 **X**로 표현

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

행 row

열 column

선형대수 | 벡터와 행렬

- 행렬 $\mathbf{A}$의 전치행렬 $\mathbf{A}^{\mathrm{T}}$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

예를 들어, $\mathbf{A} = \begin{pmatrix} 3 & 4 & 1 \\ 0 & 5 & 2 \end{pmatrix}$라면 $\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 3 & 0 \\ 4 & 5 \\ 1 & 2 \end{pmatrix}$

- Iris의 설계 행렬을 전치행렬 표기에 따라 표현하면,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_{150}^{\mathrm{T}} \end{pmatrix}$$

선형대수 | 벡터와 행렬

- 행렬을 이용하면 수학을 간결하게 표현할 수 있음
  - 예) 다항식의 행렬 표현

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_1, x_2, x_3) \\
&= 2x_1x_1 - 4x_1x_2 + 3x_1x_3 + x_2x_1 + 2x_2x_2 + 6x_2x_3 - 2x_3x_1 + 3x_3x_2 + 2x_3x_3 + 2x_1 + 3x_2 - 4x_3 + 5 \\
&= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 2 & -4 & 3 \\ 1 & 2 & 6 \\ -2 & 3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 2 & 3 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 5 \\
&= \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + c
\end{aligned}
$$

- 특수한 행렬들

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

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

선형대수 | 벡터와 행렬

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

    $2*3$ 행렬 $\mathbf{A} = \begin{pmatrix} 3 & 4 & 1 \\ 0 & 5 & 2 \end{pmatrix}$와 $3*3$행렬 $\mathbf{B} = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 0 & 5 \\ 4 & 5 & 1 \end{pmatrix}$을 곱하면 $2*3$ 행렬 $\mathbf{C} = \mathbf{AB} = \begin{pmatrix} 14 & 5 & 24 \\ 13 & 10 & 27 \end{pmatrix}$

    - 교환법칙 성립하지 않음: $\mathbf{AB} \neq \mathbf{BA}$
    - 분배법칙과 결합법칙 성립: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$이고 $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

  - 벡터의 내적

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

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

선형대수 | 벡터와 행렬

- 텐서
  - 3차원 이상의 구조를 가진 숫자 배열
  - 예) 3차원 구조의 RGB 컬러 영상

$$
\mathbf{A} = \begin{pmatrix} 3 & 0 & 1 & 2 & 6 & 7 \\ 3 & 1 & 2 & 3 & 5 & 6 \\ 1 & 2 & 2 & 2 & 2 & 3 \\ 3 & 0 & 0 & 1 & 1 & 0 \\ 5 & 4 & 1 & 3 & 3 & 3 \\ 2 & 2 & 1 & 2 & 2 & 1 \end{pmatrix}
$$

선형대수 | 놈과 유사도

- 벡터와 행렬의 크기를 놈으로 측정
  - 벡터의 p차 놈

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

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

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

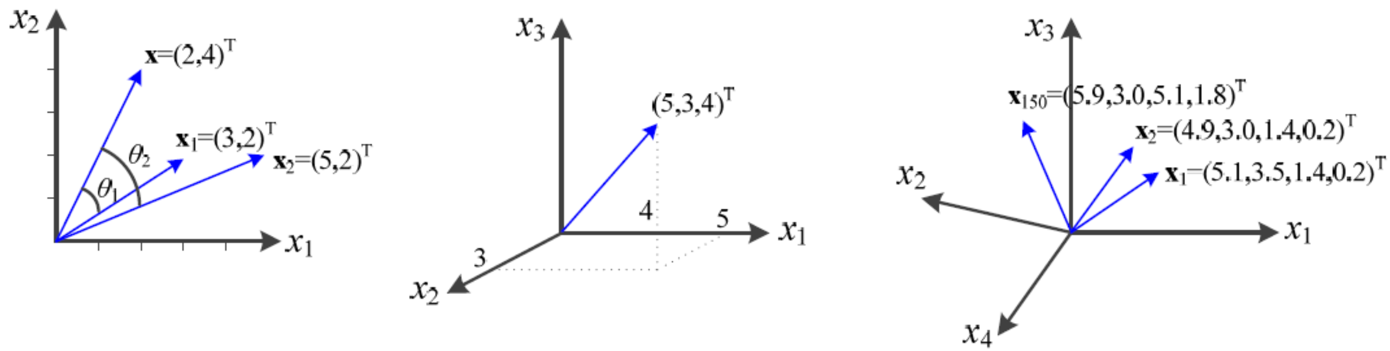
  - 행렬의 프로베니우스 놈

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

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

선형대수 | 놈과 유사도

- 유사도와 거리
  - 벡터를 기하학적으로 해석



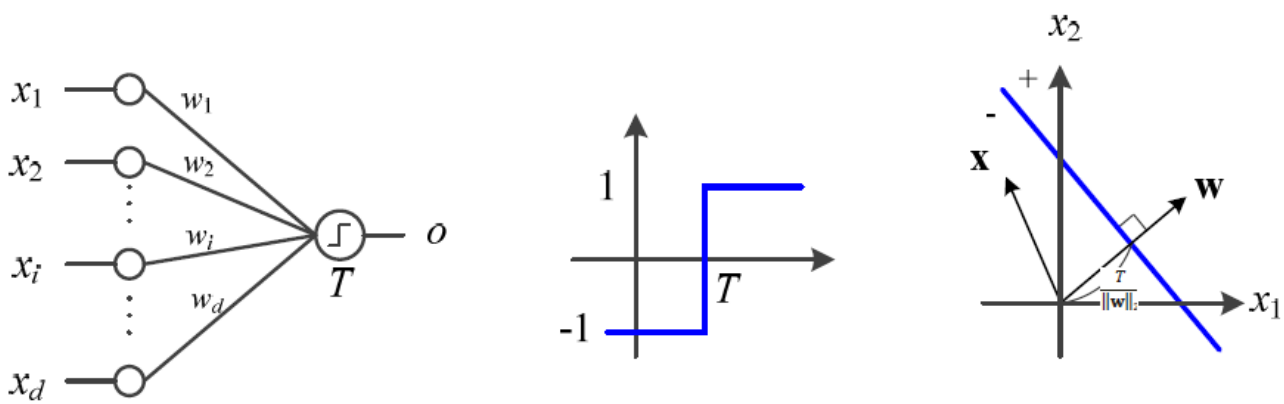
(a) 2차원 벡터 (b) 3차원 벡터 (c) 4차원 벡터(Iris 데이터)

  - 코사인 유사도

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

선형대수 | 퍼셉트론의 해석

- 퍼셉트론
  - 1958년 로젠블렛이 고안한 분류기 모델



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

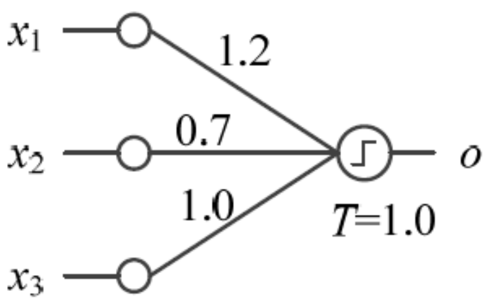
  - 퍼셉트론의 동작을 수식으로 표현하면,

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

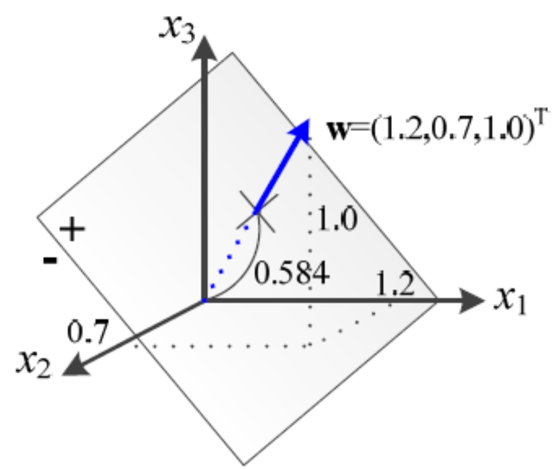
    - 활성 함수 τ로는 계단함수 사용

선형대수 | 퍼셉트론의 해석

- 퍼셉트론
  - 파란 직선은 두 개의 부분공간을 나누는 결정직선 decision line
    - $\mathbf{w}$에 수직이고 원점으로부터 $\frac{T}{\|\mathbf{w}\|_2}$만큼 떨어져 있음
  - 3차원 특징공간은 결정평면 decision plane, 4차원 이상은 결정 초평면 decision hyperplane
  - 예) 3차원 특징공간을 위한 퍼셉트론



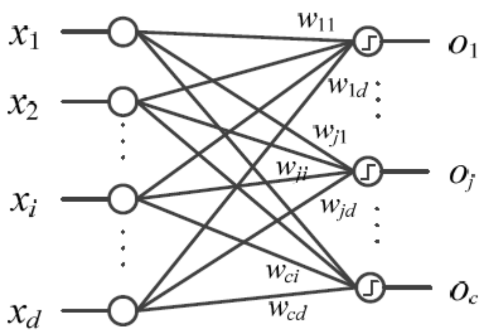
(a) 퍼셉트론



(b) 공간 분할(2부류 분류)

선형대수 | 퍼셉트론의 해석

- 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,

행렬로 간결하게 쓰면 $\mathbf{o} = \boldsymbol{\tau}(\mathbf{W}\mathbf{x})$

$$\mathbf{o} = \boldsymbol{\tau}\begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{x} \\ \mathbf{w}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{w}_c \cdot \mathbf{x} \end{pmatrix} \longrightarrow \quad \text{이때 } \mathbf{W} = \begin{pmatrix} \mathbf{w}_1^{\mathrm{T}} \\ \mathbf{w}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{w}_c^{\mathrm{T}} \end{pmatrix}$$

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

선형대수 | 퍼셉트론의 해석

- 학습의 정의
  - 아래 식은 학습을 마친 프로그램을 현장에 설치했을 때 일어나는 과정

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

  - 학습 과정: 학습은 훈련집합의 샘플에 대해 아래 식을 가장 잘 만족하는 **W**를 찾아내는 작업

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

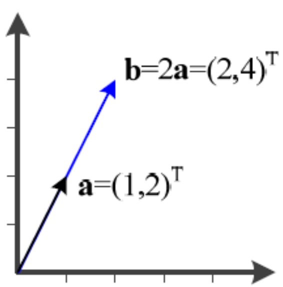
- 현대 기계 학습에서 퍼셉트론의 중요성
  - 딥러닝은 퍼셉트론을 여러 층으로 확장하여 만듦

선형대수 | 선형결합과 벡터공간

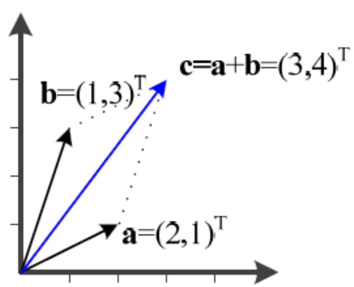
- 벡터
  - 공간상의 한 점으로 화살표 끝이 벡터의 좌표에 해당
- 선형결합이 만드는 벡터공간
  - 기저벡터 **a**와 **b**의 선형결합

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

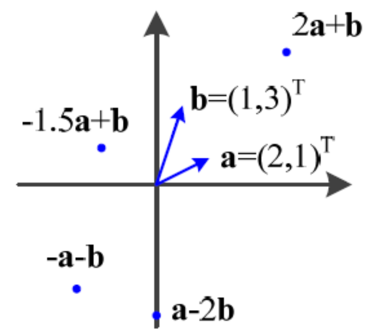
  - 선형결합으로 만들어지는 공간을 벡터공간이라 부름



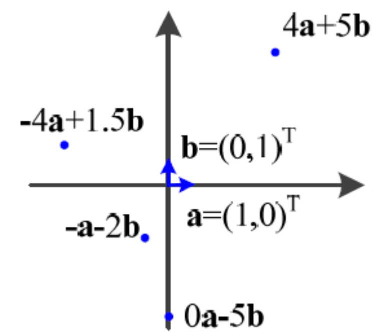
(a) 벡터에 스칼라 곱



(b) 두 벡터의 덧셈



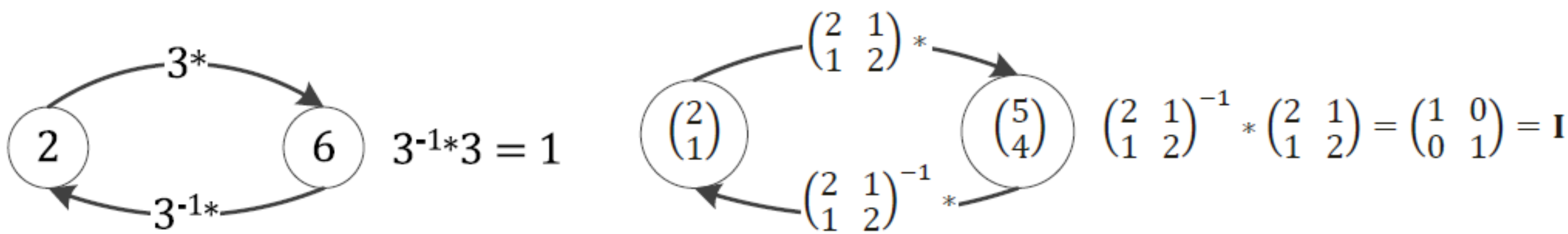
(a) 기저 벡터와 벡터공간



(b) 정규직교 기저 벡터

선형대수 | 역행렬

- 역행렬의 원리



(a) 역수의 원리 (b) 역행렬의 원리

  - 정사각행렬 **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 $\begin{pmatrix} 2 & -0.5 \\ -3 & 1 \end{pmatrix}$

선형대수 | 역행렬

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

선형대수 | 역행렬

- 역행렬 공식

$$A^{-1} = \frac{\operatorname{adj} A}{\det A}$$

- 2 x 2 행렬에 대한 역행렬

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

- 3 x 3 행렬에 대한 역행렬

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}^{-1} = \frac{1}{K} \begin{bmatrix} A_{22}A_{33} - A_{23}A_{32} & A_{13}A_{32} - A_{12}A_{33} & A_{12}A_{23} - A_{13}A_{22} \\ A_{23}A_{31} - A_{21}A_{33} & A_{11}A_{33} - A_{13}A_{31} & A_{13}A_{21} - A_{11}A_{23} \\ A_{21}A_{32} - A_{22}A_{31} & A_{12}A_{31} - A_{11}A_{32} & A_{11}A_{22} - A_{12}A_{21} \end{bmatrix}$$

$$K = A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}$$

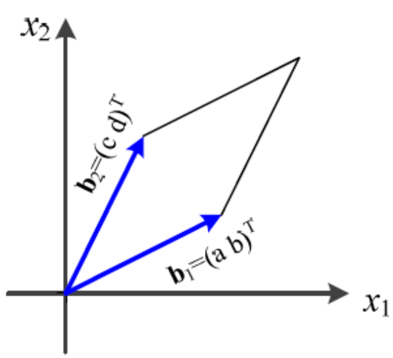
선형대수 | 역행렬

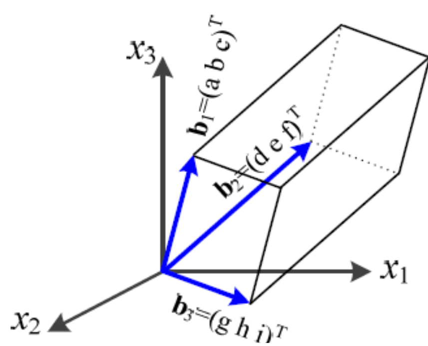
- 행렬 $\mathbf{A}$의 행렬식 $det(\mathbf{A})$

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

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

- 기하학적 의미
  - 2차원에서는 2개의 행 벡터가 이루는 평행사변형의 넓이
  - 3차원에서는 3개의 행 벡터가 이루는 평행사각기둥의 부피





선형대수 | 행렬 분해

- 분해란?
  - 정수 3717은 특성이 보이지 않지만, 3*3*7*59로 소인수 분해를 하면 특성이 보이듯이, 행렬도 분해하면 여러모로 유용함

- 고윳값과 고유 벡터
  - 고유 벡터 $\mathbf{v}$와 고윳값 λ

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

  - 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 $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

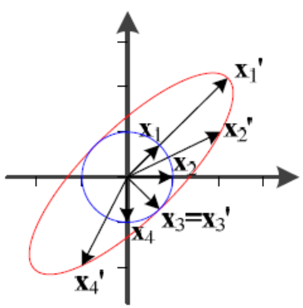
선형대수 | 행렬 분해

- 고윳값과 고유 벡터의 기하학적 해석

[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 3배 만큼, $\mathbf{x}_3$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.



선형대수 | 행렬 분해

- 고윳값 분해eigen value decomposition

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

  - $\mathbf{Q}$는 $\mathbf{A}$의 고유 벡터를 열에 배치한 행렬이고 $\mathbf{\Lambda}$는 고윳값을 대각선에 배치한 대각행렬
  - 예를 들어, $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$
  - 고윳값 분해는 정사각행렬에만 적용 가능한데, 기계 학습에서는 정사각행렬이 아닌 경우의 분해도 필요하므로 고윳값 분해는 한계를 가짐

선형대수 | 행렬 분해

- $n*m$ 행렬 **A**의 특잇값 분해SVD(singular value decomposition)

$$\mathbf{A} = \mathbf{U\Sigma V^T} \tag{2.22}$$

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

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

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

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

Thank you