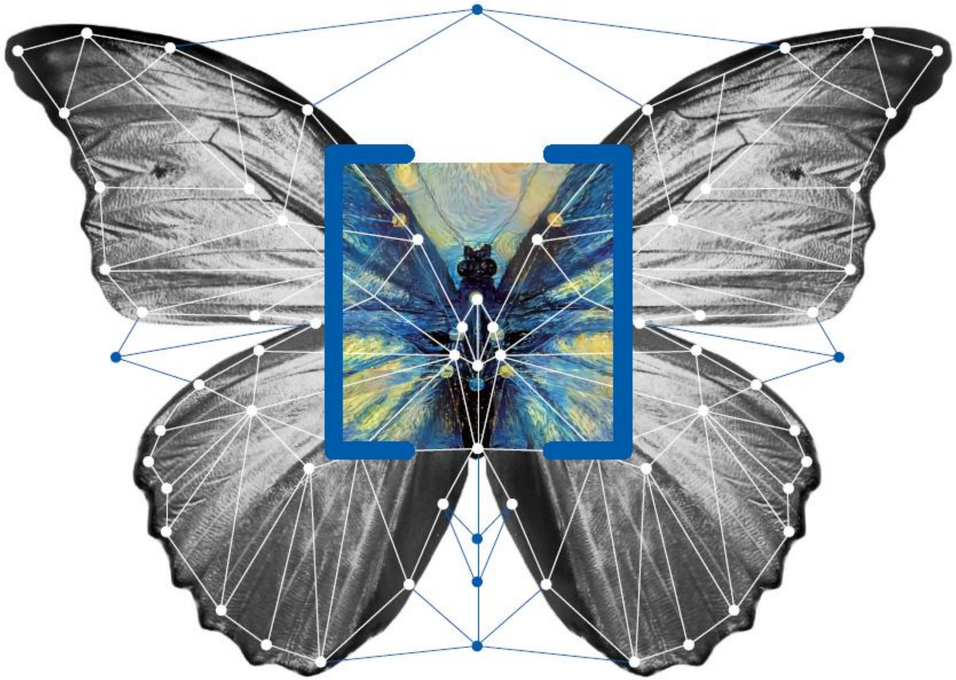
한빛아카데미 Hanbit Academy, Inc.



MACHINE LEARNING 기계 학습

오일석 지음

2장. 기계 학습과 수학

벡터의 표현

- 벡터 : 크기와 방향을 가지는 임의의 물리량

- 패턴 인식에서는 인식 대상이 되는 객체가 특징으로 표현되고, 특징은 차원을 가진 벡터로 표현된다. 이러한 벡터를 **특징 벡터(feature vector)**라고 한다.

- 특징 벡터에 대한 **대수학적 계산을 위해서 특징 벡터를 행렬로 표현**하여 N차원 공간상의 한 점의 데이터로 특징을 다루게

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \cdots \\ \mathbf{a}_N \end{bmatrix}$$

열벡터 (Column Vector) 사용

2.1.1 벡터와 행렬

벡터

- 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

꽃 150개의 특징벡터들

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

2.1.1 벡터와 행렬

행렬

- 여러 개의 벡터를 담음
- 훈련집합을 담은 행렬을 설계행렬이라 부름
- 예) Iris 데이터에 있는 150개의 샘플을 설계 행렬 $\mathbf{X}$로 표현

행과 열번호를 유의깊게 관찰

$$
\mathbf{X} = \begin{pmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 4.7 & 3.2 & 1.3 & 0.2 \\ 4.6 & 3.1 & 1.5 & 0.2 \\ \vdots & \vdots & \vdots & \vdots \\ 6.2 & 3.4 & 5.4 & 2.3 \\ 5.9 & 3.0 & 5.1 & 1.8 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{149,1} & x_{149,2} & x_{149,3} & x_{149,4} \\ x_{150,1} & x_{150,2} & x_{150,3} & x_{150,4} \end{pmatrix} & \begin{array}{l} \leftarrow \text{행 row} \\ \\ \\ \\ \\ \\ \\ \end{array} \\ \begin{array}{l} \uparrow \\ \text{열 column} \end{array} & \end{array}
$$

150 X 4 행렬
행번호: 1 – 150
열번호: 1 – 4

2.1.1 벡터와 행렬

행렬 A의 전치행렬 $\mathbf{A}^\mathrm{T}$

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

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,

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

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

2.1.1 벡터와 행렬

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

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

열벡터 기준으로 표기

- **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 $\begin{pmatrix} 1 & 2 & 11 \\ 2 & 21 & 5 \\ 11 & 5 & 1 \end{pmatrix}$

2.1.1 벡터와 행렬

대각 행렬

- 행렬의 대각 성분을 제외하고는 모두 0인 행렬
- 스칼라 행렬 : 대각 성분이 모두 같고, 비대각 성분이 모두 0인 정방행렬

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

항등 행렬 혹은 단위 행렬

- 대각 성분이 모두 1이고 그밖의 성분이 모두 0인 정방행렬

대칭 행렬

- 대각선을 축으로 모든 성분이 대칭되는 행렬

행렬 연산

- 행렬 곱셈 $\mathbf{C} = \mathbf{AB}$, 이때 $c_{ij} = \sum_{k=1,s} a_{ik} b_{kj}$ (2.1)

2*3 행렬 $\mathbf{A} = \begin{pmatrix} 3 & 4 & 1 \\ 0 & 5 & 2 \end{pmatrix}$와 3*3행렬 $\mathbf{B} = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 0 & 5 \\ 4 & 5 & 1 \end{pmatrix}$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 $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^{\mathrm{T}}\mathbf{b} = \sum_{k=1,d} a_k b_k$ (2.2)

$\mathbf{x}_1 = \begin{pmatrix} 5.1 \\ 3.5 \\ 1.4 \\ 0.2 \end{pmatrix}$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

2.1.4 선형결합과 벡터공간

벡터

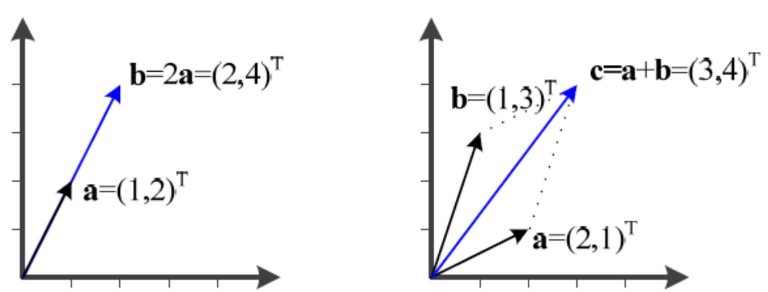
- 공간상의 한 점으로 화살표 끝이 벡터의 좌표에 해당

선형결합이 만드는 벡터공간

- 기저벡터 a와 b의 선형결합

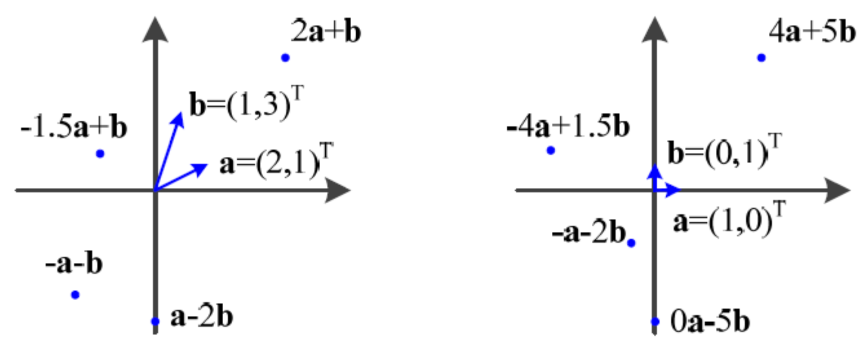
$$\mathbf{c} = \alpha_1 \mathbf{a} + \alpha_2 \mathbf{b} \quad (2.12)$$

- 선형결합으로 만들어지는 공간을 벡터공간이라 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

선형결합의 2차 형식

- 3개의 변수 x,y,z의 2차 형식(quadratic form)은 다음과 같은 동차 함수식을 말함
- 2차 함수이기 때문에 2차 형식이라고 한다.

$$F = ax^2 + by^2 + cx^2 + 2fxy + 2gyz + 2hzx$$

- 행렬을 이용하여 표시하면

$$F(x, y, z) = (x, y, z)\begin{pmatrix} a & f & h \\ f & b & g \\ h & g & c \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

- 일반화하면

$$F(x) = a_{11}x_1{}^2 + a_{22}x_2^2 + \ldots + a_{nn}x_n{}^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \ldots + 2a_{n-1,n}x_{n-1}x_n$$

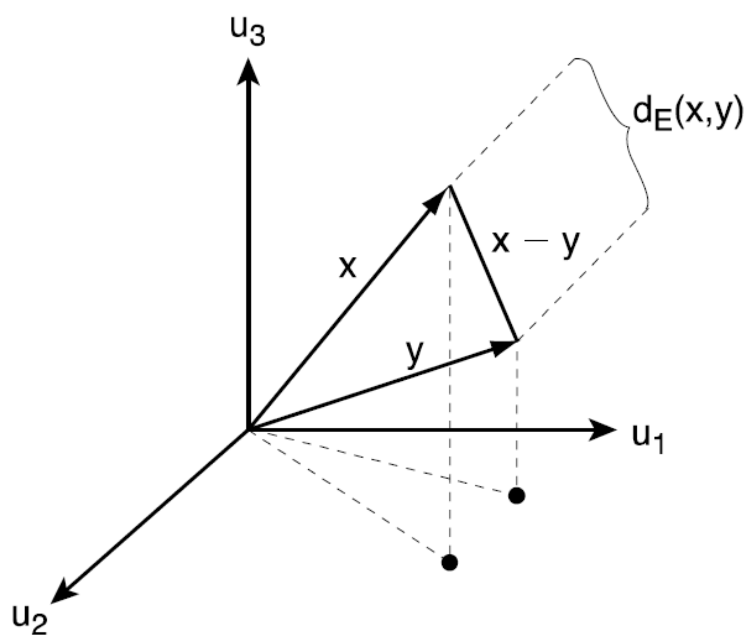
$$= (x_1, x_2, \ldots, x_n)\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}$$

※ 이때 $a_{ij} = a_{ji}$

유사도

- **유클리디안 거리**
  - ➢ 벡터 공간상에서 두 점 간의 거리는 점 사이 벡터 차의 크기로 정의

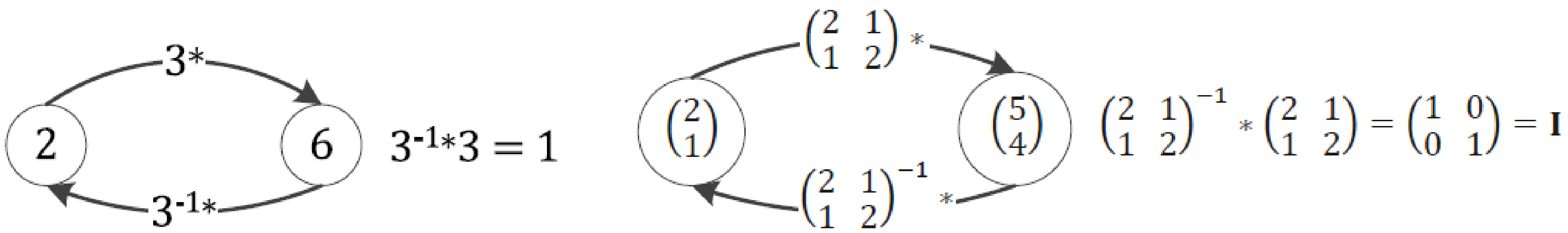
$$d_E(\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}| = \left[\sum_{k=1}^{N}(\mathbf{x}_k - \mathbf{y}_k)^2\right]^{1/2}$$



[그림 2-3] 유클리디안 거리

2.1.5 역행렬

■ 역행렬의 원리



(a) 역수의 원리 (b) 역행렬의 원리

그림 2-9 역행렬

- **정사각행렬 A의 역행렬 A^{-1}**

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

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$(\mathbf{A}^{-1})^{T} = (\mathbf{A}^{T})^{-1}$$

행렬 A의 행렬식

- 행렬식은 행렬을 어떠한 하나의 실수 값으로 표현한 것을 말함
- d×d 정방 행렬 A에 대해 행렬식은 |A| 혹은 det A 으로 표현하며 다음과 같은 성질을 가짐

  » 2x2 행렬의 행렬식

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

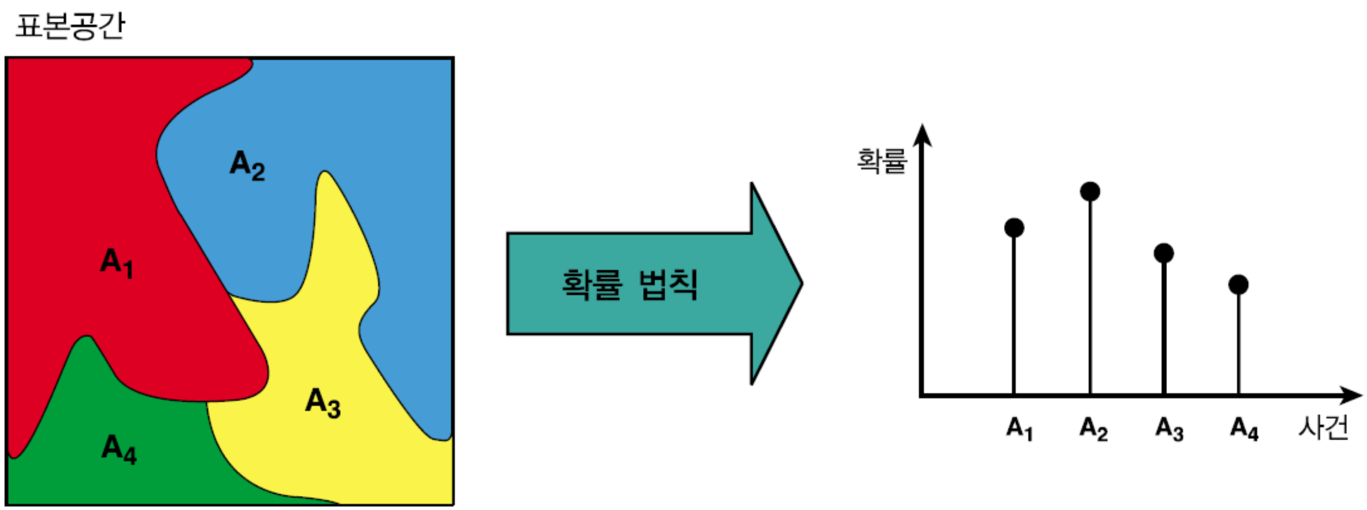
  » 3x3 행렬의 행렬식

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

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

2.2.1 확률 기초

- **확률(probability)**
  - **통계적 현상의 확실함의 정도를 나타내는 척도.**
  - **랜덤 시행에서 어떠한 사건이 일어날 정도를 나타내는 사건에 할당된 수들을 말함**
  - **확률 법칙**
    - 랜덤 시행에서 사건에 확률을 할당하는 규칙을 말함
    - 랜덤 시행의 표본 공간 S가 모든 가능한 출력 집합이 된다.



2.2.1 확률 기초

▪ 확률에 관한 정리

정리 1: $0 \le P[A_i]$

정리 2: $P[S] = 1$

정리 3: 만약 $A_i \cap A_j = \varnothing$이면, $P[A_i \cup A_j] = P[A_i] + P[A_j]$

▪ 확률에 관한 성질

성질 1: $P[A^c] = 1 - P[A]$

성질 2: $P[A] \le 1$

성질 3: $P[\varnothing] = 0$

성질 4: $\{A_1, A_2, \ldots A_N\}$이 주어질 때, 만약 $\{A_i \cap A_j = \varnothing \;\; \forall i, j\}$이면, $P[\bigcup_{k=1}^{N} A_k] = \sum_{k=1}^{N} P[A_k]$

성질 5: $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$

성질 6: $P[\bigcup_{k=1}^{N} A_k] = \sum_{k=1}^{N} P[A_k] - \sum_{j<k}^{N} P[A_k \cap A_j] + \ldots + (-1)^{N+1} P[A_1 \cap A_2 \cap \ldots \cap A_N]$

성질 7: 만약 $A_1 \subset A_2$이면, $P[A_1] \le P[A_2]$

2.2.1 확률 기초

■ **확률변수random variable**

- 예) 윷

x : 뒤집힌 것의 개수 관찰

그림 2-13 윷을 던졌을 때 나올 수 있는 다섯 가지 경우(왼쪽부터 도, 개, 걸, 윷, 모)

- 수식으로 표현하려면?? 다섯 가지 경우 중 한 값을 갖는 확률변수 x 값이라는 숫자 활용
- x의 정의역은 {도, 개, 걸, 윷, 모}

{1, 2, 3, 4, 0}

2.2.1 확률 기초

확률분포

$$P(x=\text{도})=\frac{4}{16}, P(x=\text{개})=\frac{6}{16}, P(x=\text{걸})=\frac{4}{16}, P(x=\text{윷})=\frac{1}{16}, P(x=\text{모})=\frac{1}{16}$$

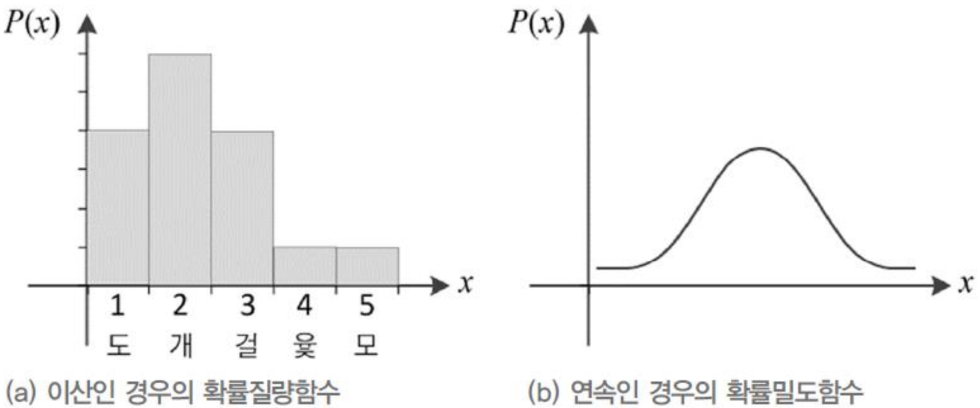


그림 2-14 확률분포

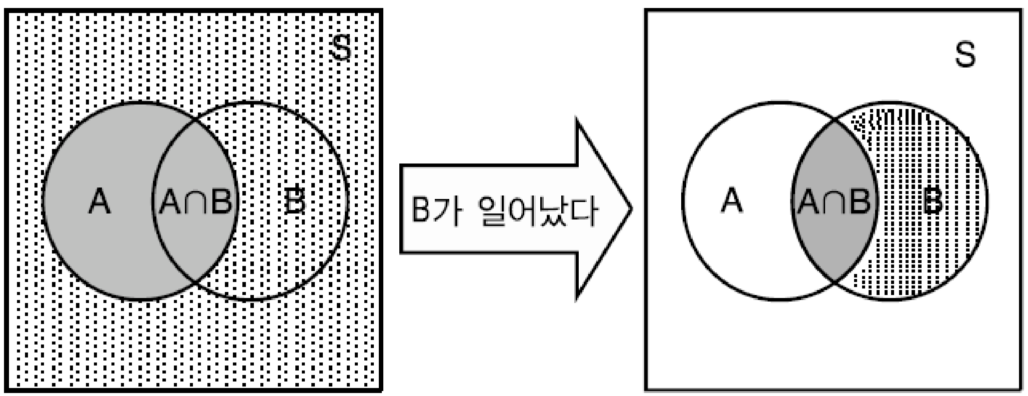
확률벡터random vector

- 예) Iris에서 확률벡터 $\mathbf{x}$는 4차원 $\mathbf{x}=(x_1,x_2,x_3,x_4)^\mathrm{T}=$ (꽃받침 길이, 꽃받침 너비$_1$, 꽃잎 길이, 꽃잎 너비$)^\mathrm{T}$

조건부 확률

- **조건부 확률**
  - **A와 B 두 개의 사건이 있을 경우, 사건 B가 일어날 확률이 이미 알려져 있을 경우에 사건 A가 일어날 확률**

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} \qquad ※\ P[B] > 0 \text{일 경우}$$

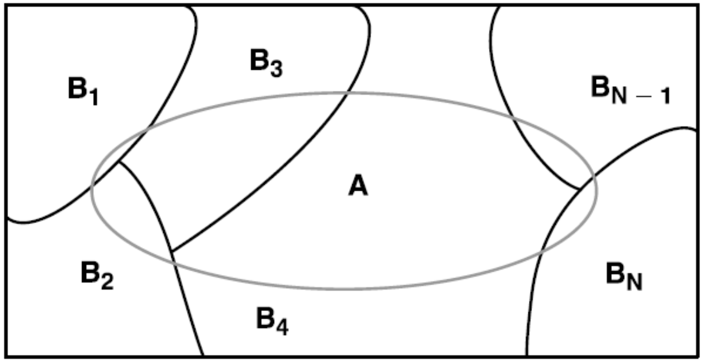


주어진 B에 대한 A의 확률

전체확률

- 전체 확률 법칙 (law of total probability)
  - $B_1, B_2, \ldots, B_n$의 합집합이 표본 공간이고, 서로 상호 배타적인 사건이라고 하자. 표본 공간 S의 분할 영역으로 이들 집합을 나타낼 수 있다. 이 때, 사건 A은 아래 식과 같이 표현된다.

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \ldots \cup B_N) = (A \cap B_1) \cup (A \cap B_2) \cup \ldots (A \cap B_N)$$



A라는 사건이 발생했는데 그것이 B1에서 발생했을 확률은?
이런 질문에 답할 수 있는 시스템을 만드는 과정에 기계학습 필요!

  - $B_1, B_2, \ldots, B_n$은 상호 배타적이므로

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_N]$$

  - 그러므로 아래식이 성립하며 이를 사건 A의 전체 확률이라고 한다.

$$P[A] = P[A \mid B_1]P[B_1] + \cdots + P[A \mid B_N]P[B_N] = \sum_{k=1}^{N} P[A \mid B_k]P[B_k]$$

2.2.1 확률 기초

간단한 확률실험 장치

- 주머니에서 번호를 뽑은 다음, 번호에 따라 해당 병에서 공을 뽑고 색을 관찰함
- 번호를 y, 공의 색을 x라는 확률변수로 표현하면 정의역은 $y \in \{①, ②, ③\}$, $x \in \{$파랑, 하양$\}$

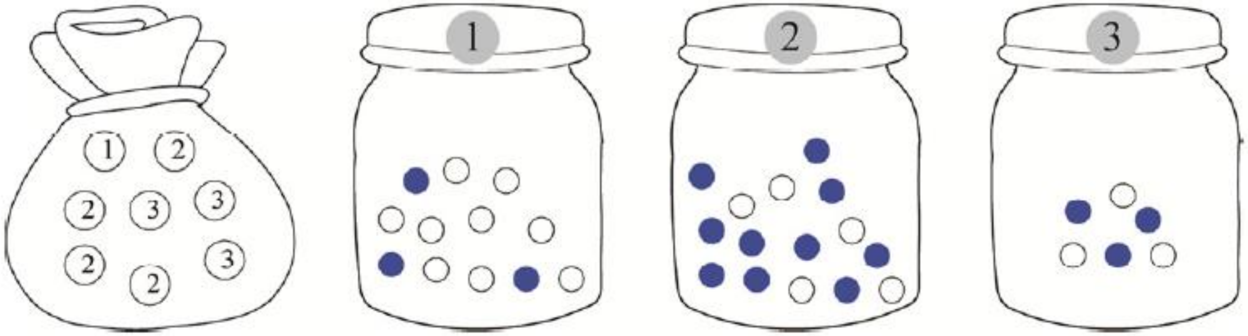


그림 2-15 확률 실험

2.2.1 확률 기초

곱 규칙과 합 규칙

- ①번 카드를 뽑을 확률은 $P(y=①)=P(①)=1/8$
- 카드는 ①번, 공은 하얀일 확률은 $P(y=①,x=\text{하얀})=P(①,\text{하얀})$ ← 결합확률

$$P(y=①, x=\text{하얀}) = P(x=\text{하얀}|y=①)P(y=①) = \frac{9}{12}\frac{1}{8} = \frac{3}{32}$$

- **곱 규칙**

$$\text{곱 규칙: } P(y,x) = P(x|y)P(y)$$

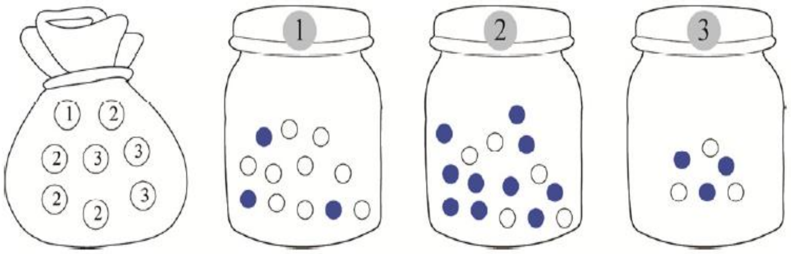


그림 2-15 확률 실험

- **하얀 공이 뽑힐 확률**

$$P(\text{하얀}) = P(\text{하얀}|①)P(①) + P(\text{하얀}|②)P(②) + P(\text{하얀}|③)P(③)$$
$$= \frac{9}{12}\frac{1}{8} + \frac{5}{15}\frac{4}{8} + \frac{3}{6}\frac{3}{8} = \frac{43}{96}$$

- **합 규칙**

$$\text{합 규칙: } P(x) = \sum_y P(y,x) = \sum_y P(x|y)P(y) \tag{2.24}$$

2.2.2 베이즈 정리와 기계 학습

베이즈 정리 (식 (2.26))

$$P(y,x) = P(x|y)P(y) = P(x,y) = P(y|x)P(x)$$

$$\longrightarrow \quad P(y|x) = \frac{P(x|y)P(y)}{P(x)} \qquad (2.26)$$

- **다음 질문을 식 (2.27)로 쓸 수 있음**

"하얀 공이 나왔다는 사실만 알고 어느 병에서 나왔는지 모르는데, 어느 병인지 추정하라."

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, P(y|x) \qquad (2.27)$$

확률이 가장 높게 나오는 병으로 추정하는 것이 합리적

2.2.2 베이즈 정리와 기계 학습

베이즈 정리 (식 (2.26))

- 베이즈 정리를 적용하면 $\hat{y} = \underset{y}{\mathrm{argmax}}\, P(y|x = \text{하양}) = \underset{y}{\mathrm{argmax}} \frac{P(x = \text{하양}|y)P(y)}{P(x = \text{하양})}$

- 세 가지 경우에 대해 확률을 계산하면,

$$P(①|\text{하양}) = \frac{P(\text{하양}|①)P(①)}{P(\text{하양})} = \frac{\frac{9}{12}\frac{1}{8}}{\frac{43}{96}} = \frac{9}{43}$$

$$P(②|\text{하양}) = \frac{P(\text{하양}|②)P(②)}{P(\text{하양})} = \frac{\frac{5}{15}\frac{4}{8}}{\frac{43}{96}} = \frac{16}{43}$$

→ ③번 병일 확률이 가장 높음

$$P(③|\text{하양}) = \frac{P(\text{하양}|③)P(③)}{P(\text{하양})} = \frac{\frac{3}{6}\frac{3}{8}}{\frac{43}{96}} = \frac{18}{43}$$

베이즈 정리의 해석

$$\overbrace{P(y|x)}^{\text{사후확률}} = \frac{\overbrace{P(x|y)}^{\text{우도}}\ \overbrace{P(y)}^{\text{사전확률}}}{P(x)}$$

2.2.2 베이즈 정리와 기계 학습

기계 학습에 적용

- 예) Iris 데이터 분류 문제
  - 특징 벡터 **x**, 부류 $y \in$ {setosa, versicolor, virginica}
  - 분류 문제를 argmax로 표현하면 식 (2.29)



$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y|\mathbf{x}) \tag{2.29}$$

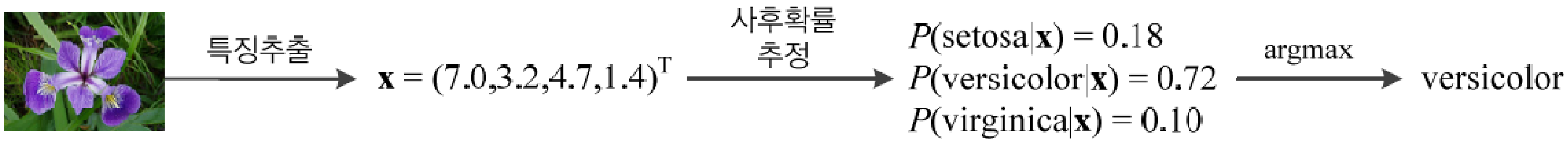


그림 2-16 붓꽃의 부류 예측 과정

- 사후확률 $P(y|\mathbf{x})$를 직접 추정하는 일은 아주 단순한 경우를 빼고 불가능
- 따라서 베이즈 정리를 이용하여 추정함

▪ 베이즈의 정리

$$P[B_j \mid A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A \mid B_j] \cdot P[B_j]}{\sum_{k=1}^{N} P[A \mid B_k] \cdot P[B_k]}$$

$$P[\omega_j \mid \mathrm{x}] = \frac{P[\mathrm{x} \mid \omega_j] \cdot P[\omega_j]}{\sum_{k=1}^{N} P[\mathrm{x} \mid \omega_k] \cdot P[\omega_k]} = \frac{P[\mathrm{x} \mid \omega_j] \cdot P[\omega_j]}{P[\mathrm{x}]}$$

※ ω_j : j번째 클래스

x : 특징 벡터

- $P[\omega_j]$: 클래스 ω_j 의 사전 확률(prior probability)
- $P[\omega_j \mid \mathrm{x}]$: 관측 x가 주어질 경우 클래스 ω_j 에 대한 사후 확률(posterior probability)
- $P[\mathrm{x} \mid \omega_j]$: 우도(likelihood: 클래스 ω_j 가 주어질 경우 관측 x가 일어날 조건부 확률)
- $P[\mathrm{x}]$: x 가 일어날 확률로 결정에 영향을 미치지 않은 정규화 상수