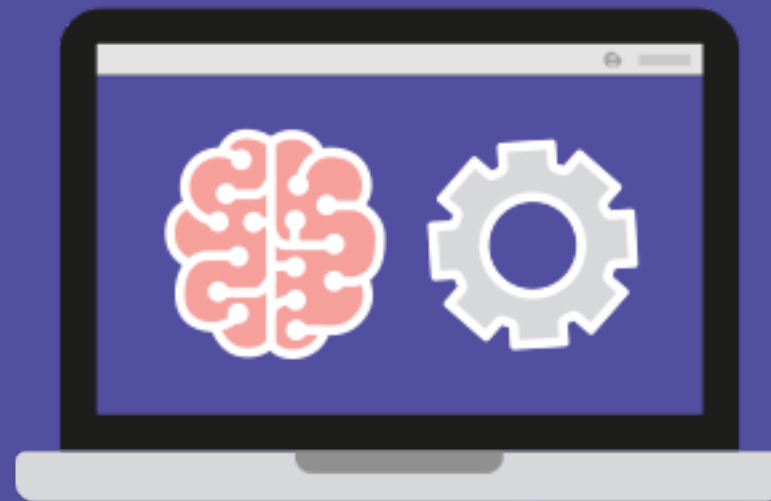


/\* elice \*/

# 인공지능/머신러닝 맛보기 - Python

Module 5: 확률 기초 및 베이지안 확률론

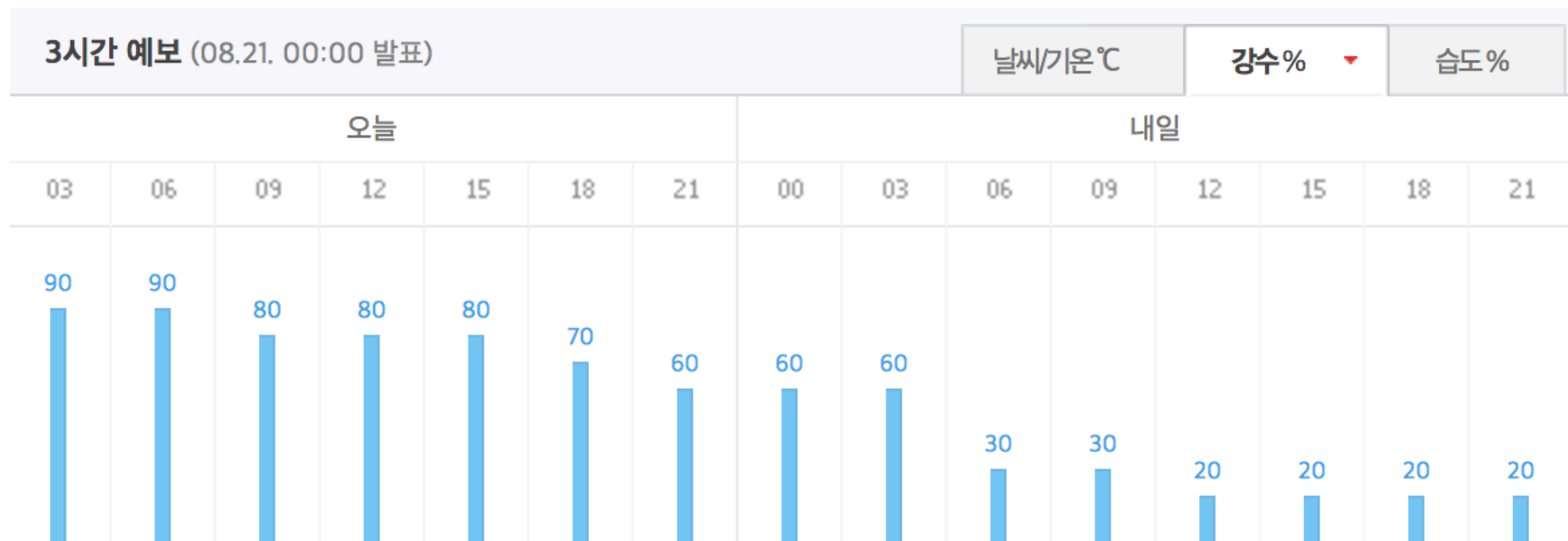


# 확률 기초

## Introduction to Probability

# 확률이란

어떤 사건이 일어날 것인지 혹은  
일어났는지에 대한 **지식** 혹은 **믿음**을 표현하는 방법



# Early Probability Theory

A와 B가 주사위 게임을 해서  
먼저 6번을 이긴 사람이 80만원을  
차지하기로 했다.

A가 5번, B가 3번 이긴 후  
게임이 중단되었다면  
판돈을 어떻게 나눠야 하는가?



Luca Pacioli

# Early Probability Theory

“게임이 중단되기 전까지의 성적에 따라 나누자”

지금까지 **A는 5번, B는 3번** 이겼으므로  
판돈도  $\frac{5}{8}$  과  $\frac{3}{8}$  의 비율로 나누면 된다!

A는 50만원, B는 30만원 — 합리적인가?

# Early Probability Theory

현재 상황: **6번** 이기면 판돈을 모두 가져감  
**A는 5번, B는 3번 승리**

	9번째 판	10번째 판	11번째 판
A가 승리	O		
B가 승리	X		

$$\frac{1}{2}$$



Blaise Pascal

# Early Probability Theory

현재 상황: **6번** 이기면 판돈을 모두 가져감  
**A는 5번, B는 3번 승리**

	9번째 판	10번째 판	11번째 판
A가 승리	X	O	
B가 승리	O	X	

$$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$



Blaise Pascal

# Early Probability Theory

현재 상황: **6번** 이기면 판돈을 모두 가져감  
**A는 5번, B는 3번 승리**

	9번째 판	10번째 판	11번째 판
A가 승리	X	X	O
B가 승리	O	O	X

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$



Blaise Pascal



# Early Probability Theory

현재 상황: **6번** 이기면 판돈을 모두 가져감  
**A는 5번, B는 3번 승리**

	9번째 판	10번째 판	11번째 판
A가 승리	X	X	X
B가 승리	O	O	O

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$



Blaise Pascal

# Early Probability Theory

현재 상황: **6번** 이기면 판돈을 모두 가져감  
**A는 5번, B는 3번 승리**

**A가 승리할 확률**  $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$

**B가 승리할 확률**  $\frac{1}{8}$

**7:1로 나누면 공평하다.**



Blaise Pascal

# 확률 예제

엘리스를 수강한 학생 1,000명에게 질문

“강아지를 기르시나요?": 예/아니오

사건 Yes: 강아지를 기른다 -> 88명

사건 No: 강아지를 기르지 않는다 -> 912명

$$P(\text{Yes}) = 88/1000 = 0.088$$

$$P(\text{No}) = 912/1000 = 0.912$$

# 확률 예제

완벽한 주사위를 한 번 굴린다.

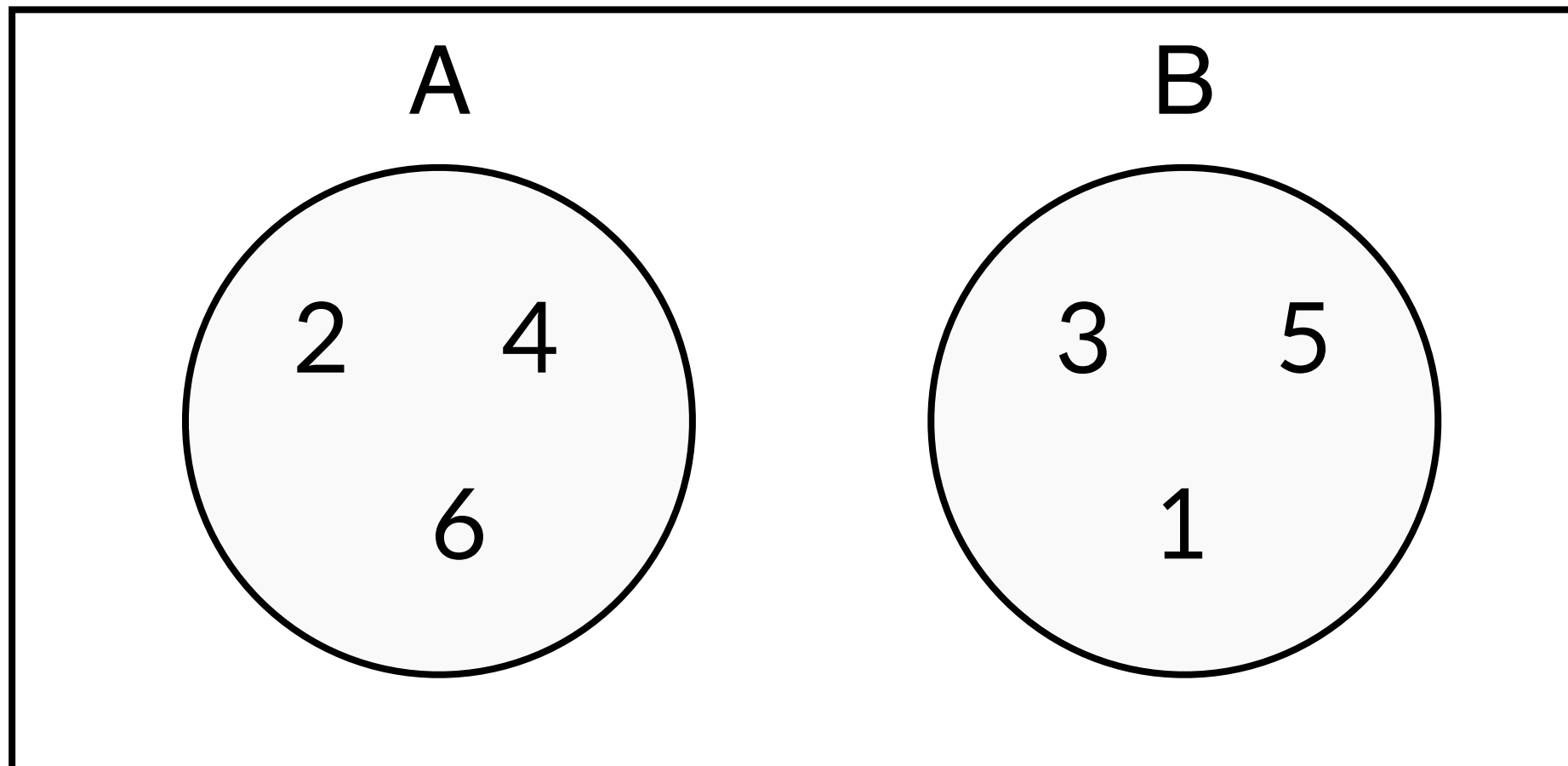
짝수가 나오는 사건  $A: \{2, 4, 6\}$

홀수가 나오는 사건  $B: \{1, 3, 5\}$

$$P(A) = 0.5$$

$$P(B) = 0.5$$

# 확률 기초



$$P(A) = 0.5$$

$$P(B) = 0.5$$

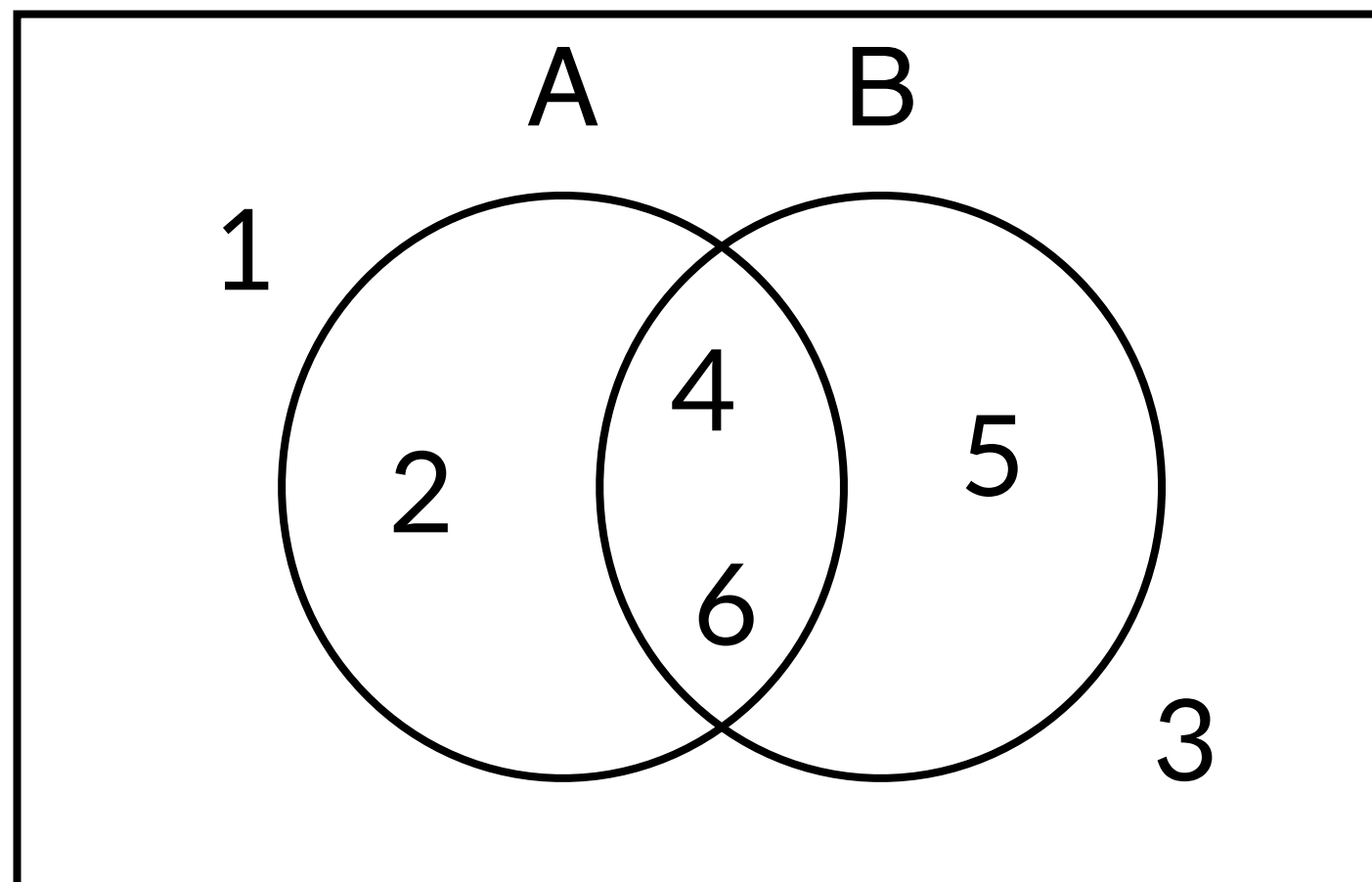
$$\text{합집합 } P(A \cup B) = 1$$

$$\text{교집합 } P(A \cap B) = 0$$

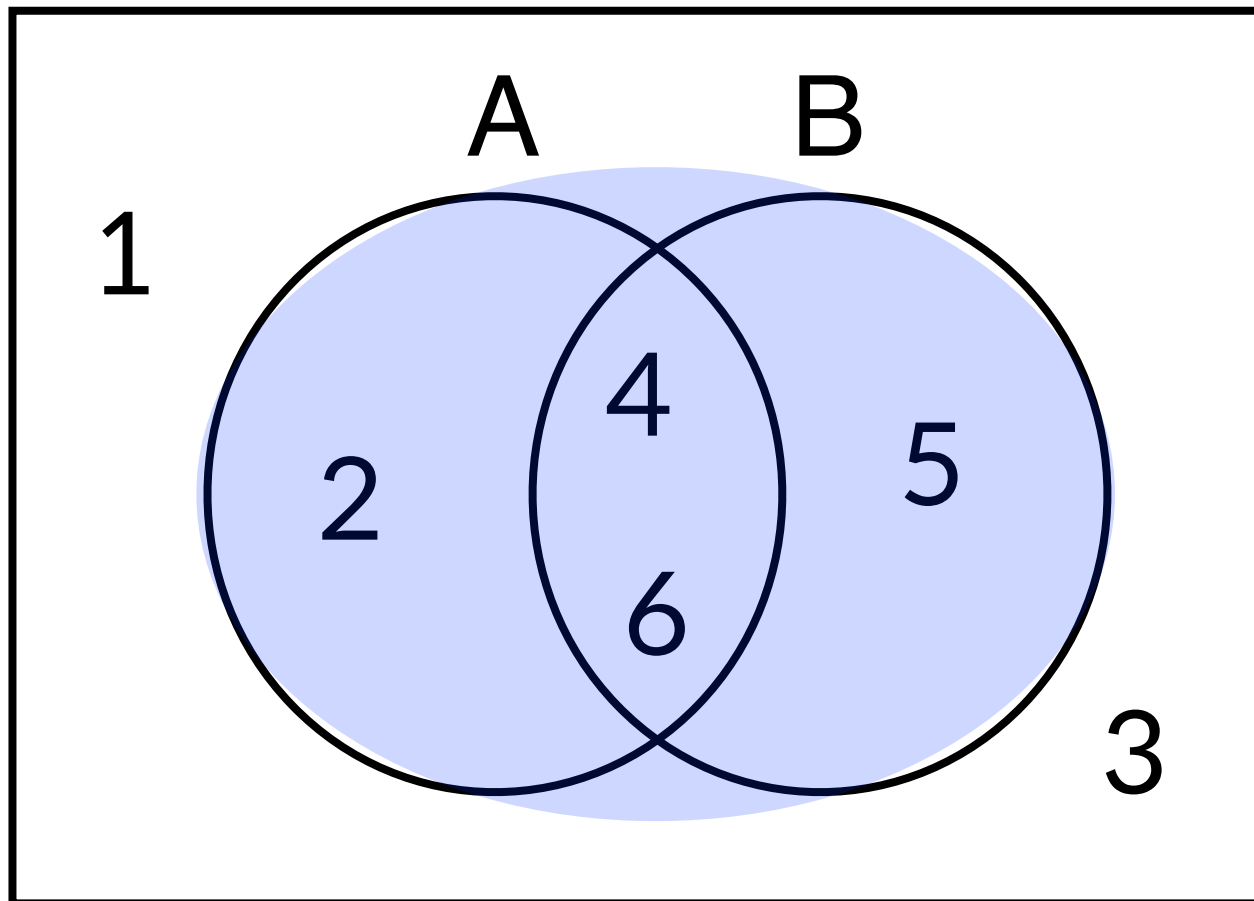
# 확률 기초

짝수가 나오는 사건  $A: \{2, 4, 6\}$

3보다 큰 수가 나오는 사건  $B: \{4, 5, 6\}$



# 확률 기초



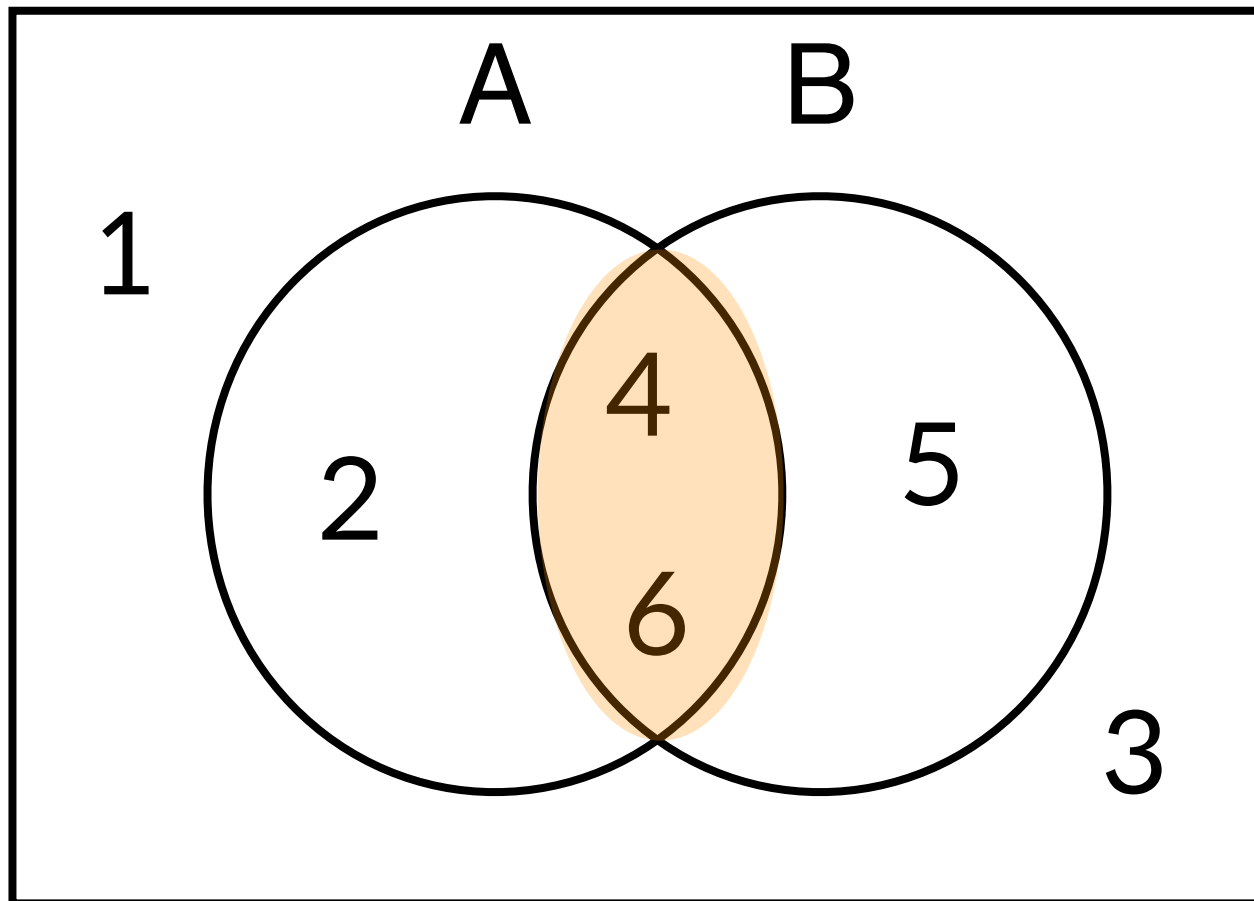
$$P(A) = 3/6 = 1/2$$

$$P(B) = 3/6 = 1/2$$

합집합  $P(A \cup B) = 4/6$

교집합  $P(A \cap B) = 2/6$

# 확률 기초



$$P(A) = 3/6 = 1/2$$

$$P(B) = 3/6 = 1/2$$

$$\text{합집합 } P(A \cup B) = 4/6$$

$$\text{교집합 } P(A \cap B) = 2/6$$



# 확률 기초

짝수가 나오는 사건  $A: \{2, 4, 6\}$

3보다 큰 수가 나오는 사건  $B: \{4, 5, 6\}$

**조건부확률** : 사건 B가 일어났을 때 A가 일어날 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(A|B)$  = 3보다 큰 수가 나왔을 때 **그 수**가 짝수일 확률

$$= \frac{P(\{4, 6\})}{P(\{4, 5, 6\})} = 2/3$$

# 확률 기초

짝수가 나오는 사건  $A: \{2, 4, 6\}$

3보다 큰 수가 나오는 사건  $B: \{4, 5, 6\}$

**조건부확률** : 사건 B가 일어났을 때 A가 일어날 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(B|A)$  = 짝수가 나왔을 때 그 수가 3보다 클 확률

$$\frac{P(\{4, 6\})}{P(\{2, 4, 6\})} = \frac{2}{3}$$

# 실습: 확률로 $\pi$ 구해보기

`/* elice */`

# 베이즈 법칙

Bayes' Rule

# 빈도주의자 vs 베이지주의자

“동전 하나를 던졌을 때 앞면이 나올 확률은 50%이다.”

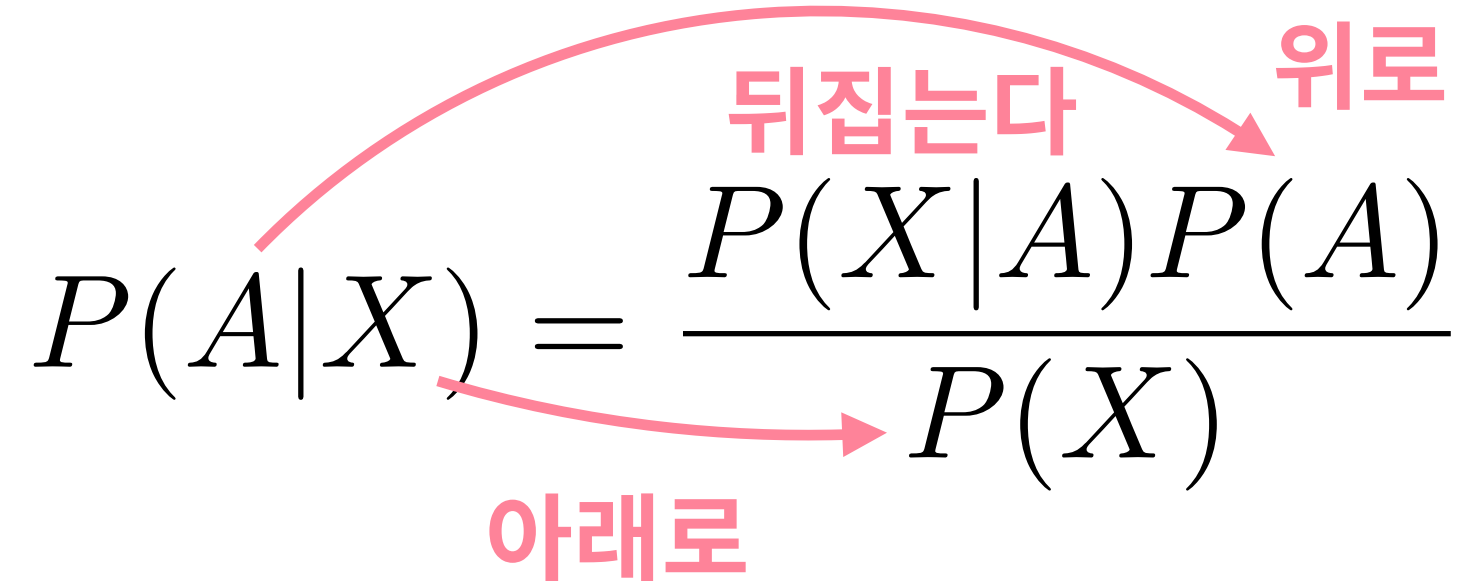
**빈도주의자:** 이 동전을 수천, 수만 번 던졌을 때  
그 중 앞면이 50%, 뒷면이 50% 나온다.

**베이지주의자:** 동전 던지기의 결과가 앞면이 나올  
것이라는 **확신** (혹은 믿음) 이 50%이다.

# 베이즈 법칙

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

# 베이즈 법칙



The diagram shows the formula for Bayes' Theorem:  $P(A|X) = \frac{P(X|A)P(A)}{P(X)}$ . A red curved arrow starts from the term  $P(A|X)$  and points to the term  $P(X|A)$  in the numerator, with the Korean text "뒤집는다" (flip) written above it. Another red curved arrow starts from the term  $P(A|X)$  and points to the term  $P(X)$  in the denominator, with the Korean text "아래로" (downward) written below it. A third red curved arrow starts from the term  $P(X|A)$  and points to the term  $P(A)$  in the numerator, with the Korean text "위로" (upward) written above it.

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

# 베이즈 법칙: 유도하기

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

$$P(A|X) = \frac{P(A \cap X)}{P(X)}$$



# 베이즈 법칙: 유도하기

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

$$P(A|X) = \frac{P(A \cap X)}{P(X)}$$

$$P(A|X)P(X) = P(A \cap X)$$

$$P(X \cap A) = P(X|A)P(A)$$

# 베이즈 법칙: 유도하기

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

$$P(A|X) = \frac{P(A \cap X)}{P(X)}$$

$$P(A|X)P(X) = P(A \cap X)$$

$$P(X \cap A) = P(X|A)P(A)$$

$$P(A|X)P(X) = P(X|A)P(A)$$

# 베이즈 법칙: 유도하기

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

$$P(A|X) = \frac{P(A \cap X)}{P(X)}$$

$$P(A|X)P(X) = P(A \cap X)$$

$$P(X \cap A) = P(X|A)P(A)$$

$$P(A|X)P(X) = P(X|A)P(A)$$

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

# 예제: 암 검사 키트

암 A에 대한 테스트 키트가 있다.

임의의 사람이 이 암에 걸릴 확률은 1%이다.

즉, 전체 인구 중 암에 걸린 사람은 1%이다.

이 암을 진단할 수 있는 키트가 있는데,

암에 걸린 사람은 99%의 확률로 양성 반응이 나오고,  
걸리지 않은 사람은 1%의 확률로 양성 반응이 나온다.

키트 검사 결과 양성 반응이 나왔다면, 암에 걸렸을 확률은?

# 예제: 암 검사 키트

암에 걸린 사건:  $A$

키트에서 양성반응이 나온 사건:  $X$

암에 걸렸을 때  
키트에서 양성반응이 나올 확률

암에 걸렸을 확률

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

키트에서 양성반응이 나왔을 때  
암에 실제로 걸렸을 확률

키트에서 양성반응이  
나올 확률

# 예제: 암 검사 키트

암에 걸린 사건: **A**

키트에서 양성반응이 나온 사건: **X**

$$P(X|A) = 0.99$$

이 암을 진단할 수 있는 키트가 있는데,

암에 걸린 사람은 99%의 확률로 양성 반응이 나오고,  
걸리지 않은 사람은 1%의 확률로 양성 반응이 나온다.

# 예제: 암 검사 키트

암에 걸린 사건: **A**

키트에서 양성반응이 나온 사건: **X**

$$P(A) = 0.01$$

임의의 사람이 이 암에 걸릴 확률은 1%이다.

# 예제: 암 검사 키트

암에 걸린 사건: **A**

키트에서 양성반응이 나온 사건: **X** not

$$P(X) = P(X \cap A) + P(X \cap \neg A)$$

=

$$P(X \cap A) = P(X|A)P(A)$$



# 예제: 암 검사 키트

암에 걸린 사건: **A**

키트에서 양성반응이 나온 사건: **X** not

$$\begin{aligned} P(X) &= P(X \cap A) + P(X \cap \neg A) \\ &= P(X|A)P(A) + P(X|\neg A)P(\neg A) \end{aligned}$$

$$P(X \cap A) = P(X|A)P(A)$$

# 예제: 암 검사 키트

암에 걸린 사건: **A**

키트에서 양성반응이 나온 사건: **X**

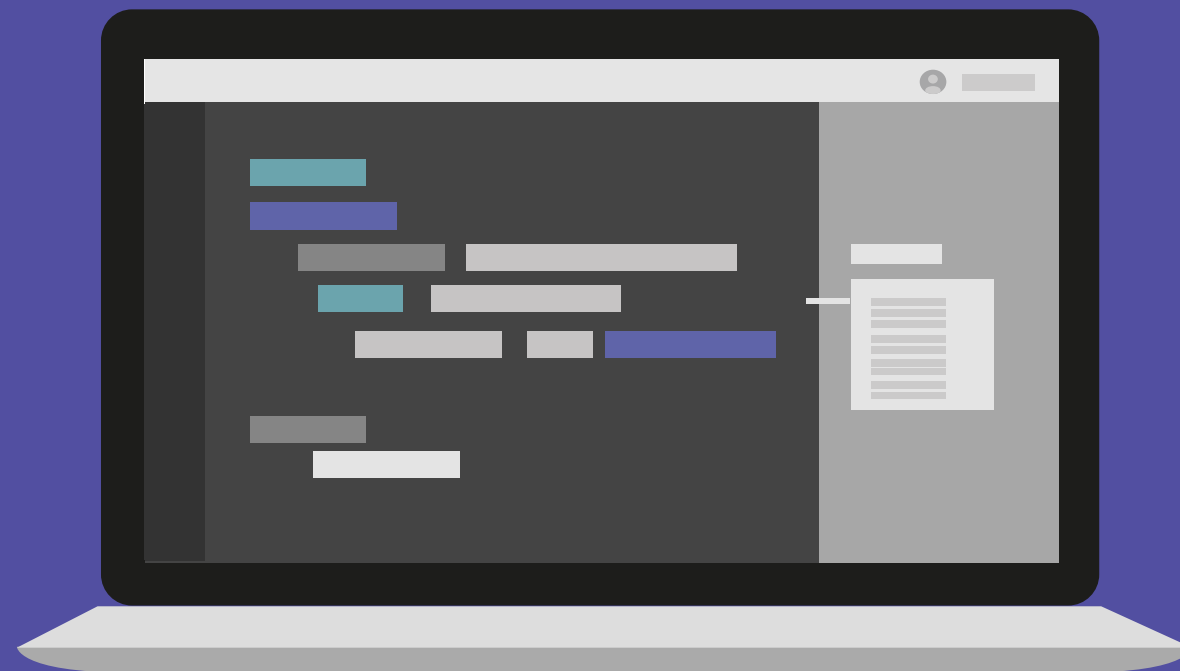
$$\begin{aligned} P(X) &= P(X \cap A) + P(X \cap \neg A) \\ &= P(X|A)P(A) + P(X|\neg A)P(\neg A) \\ &\quad \quad \quad \mathbf{0.99} \quad \mathbf{0.01} \quad \quad \quad \mathbf{0.01} \quad \mathbf{0.99} \\ &= 0.0198 \end{aligned}$$

# 예제: 암 검사 키트

$$\begin{aligned} P(A|X) &= \frac{P(X|A)P(A)}{P(X)} \\ &= \frac{0.99 \times 0.01}{0.0198} \\ &= 0.5 \end{aligned}$$

키트에서 양성 반응이 나왔는데  
암에 걸렸을 확률이 50%밖에 안 된다고?

# 실습: 유방암 검사 키트

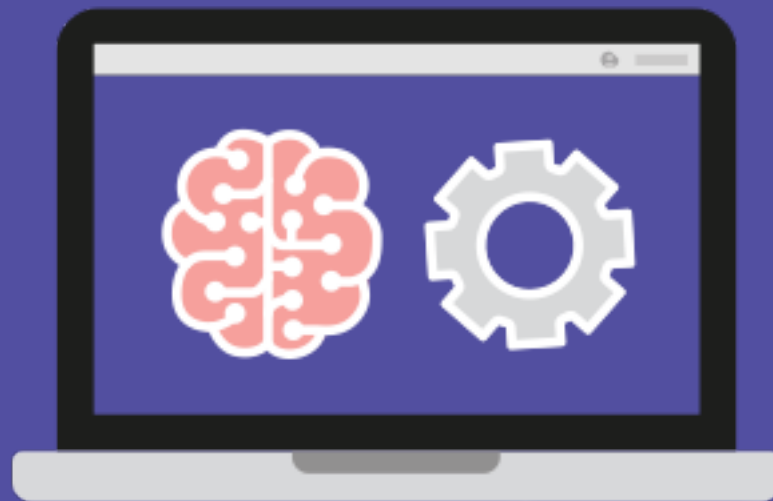


`/* elice */`

/\* elice \*/

# 인공지능/머신러닝 맛보기 - Python

## Module 6: Naive Bayes 분류기

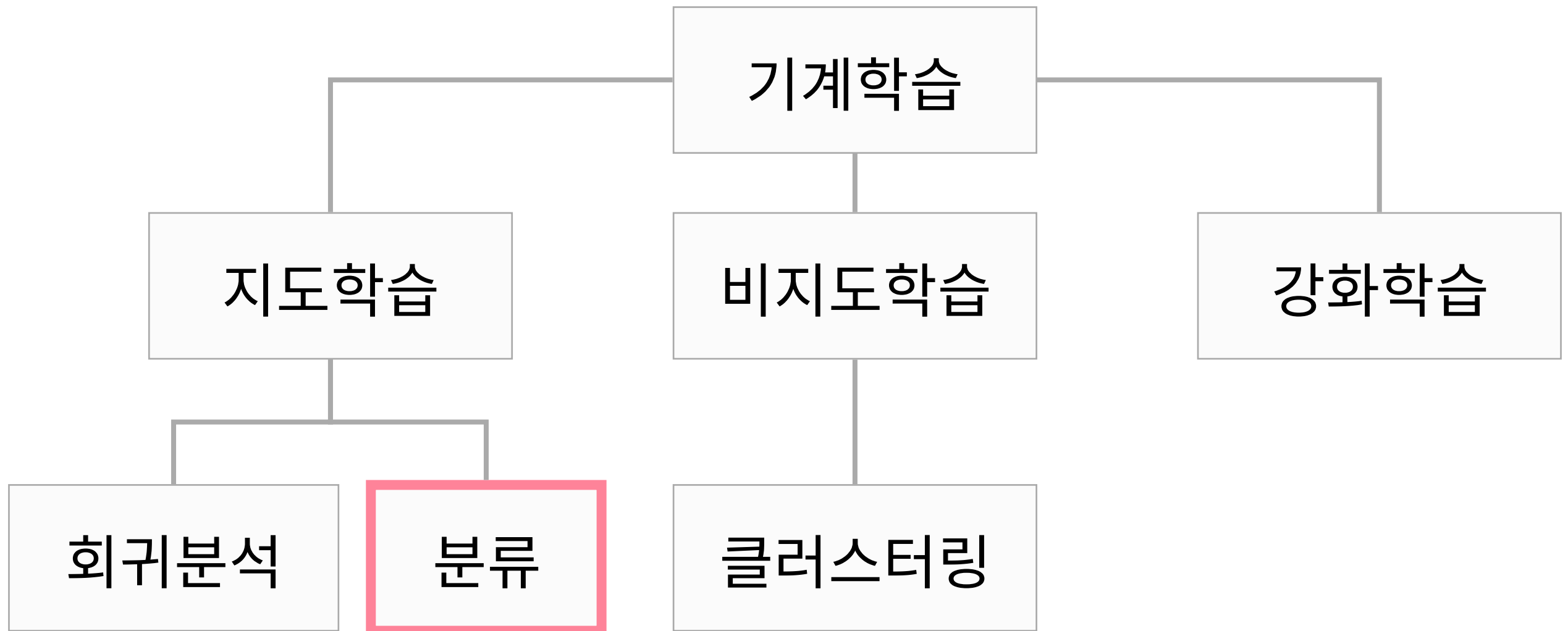


# 나이프 베이지스 분류기

`/* elice */`

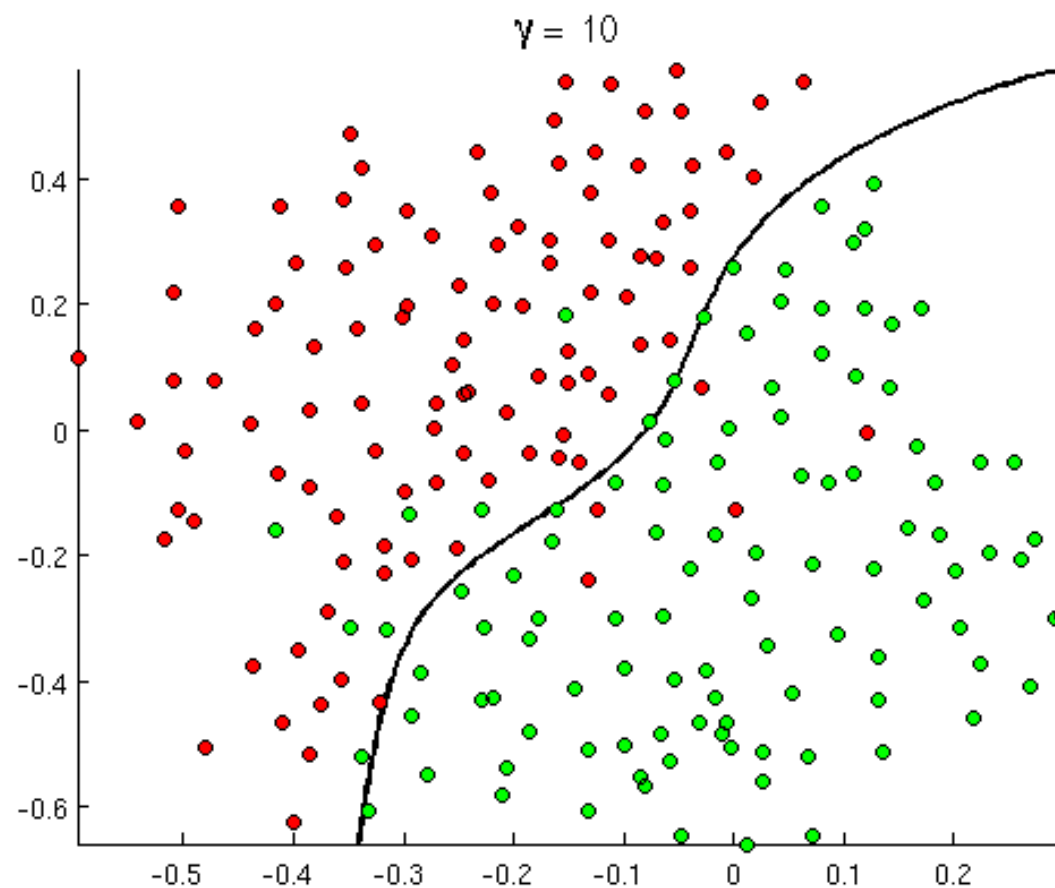
# 분류 작업

주어진 데이터가 어떤 클래스에 속하는지 알아내는 작업



# 분류기

주어진 데이터가 어떤 클래스에 속하는지  
알아내는 방법을 자동으로 학습하는 알고리즘



`/* elice */`



# 사탕 기계

사탕 기계 A, B가 있다.

이 둘은 **같은 종류의 사탕**을 내놓지만,  
들어 있는 사탕의 **비율**이 다르다.



# 사탕 기계

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1



# 사탕 기계

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

문제: 사탕 10개를 뽑아서  
빨강색 4개, 노랑색 5개, 초록색 1개를 뽑았다면  
이 사탕은 어느 기계에서 뽑은 것일까?

# 사탕 기계: 수학적 정의

**X:** 사탕 10개를 뽑아 그 결과를 관측한 사건

**A:** 사탕 기계 A에서 사탕을 뽑은 사건

**B:** 사탕 기계 B에서 사탕을 뽑은 사건

**$P(A|X)$  와  $P(B|X)$  를 비교하면 어떤 것이 더 클까?**

# 사탕 기계: 수학적 정의

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

# 사탕 기계: 수학적 정의

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

A 기계에서 사탕을 뽑았을 때  
사탕을 빨4, 노5, 초1로 뽑을 확률

A 기계에서  
사탕을 뽑을 확률

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

사탕을 빨4, 노5, 초1로 뽑았을 때  
그게 기계 A에서 뽑았을 확률

어느 기계에서 뽑든  
사탕을 빨4, 노5, 초1로 뽑을 확률

# 사탕 기계: 수학적 정의

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

B 기계에서 사탕을 뽑았을 때  
사탕을 빨4, 노5, 초1 로 뽑을 확률

B 기계에서  
사탕을 뽑을 확률

$$P(B|X) = \frac{P(X|B)P(B)}{P(X)}$$

사탕을 빨4, 노5, 초1로 뽑았을 때  
그게 기계 B에서 뽑았을 확률

어느 기계에서 뽑든  
사탕을 빨4, 노5, 초1로 뽑을 확률

# 사탕 기계: 수학적 정의

$P(A|X)$  와  $P(B|X)$  를 비교하면 어떤 것이 더 클까?

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

$$P(B|X) = \frac{P(X|B)P(B)}{P(X)}$$



# 사탕 기계: 수학적 정의

$P(A|X)$  와  $P(B|X)$  를 비교하면 어떤 것이 더 클까?

$$P(A|X) : P(B|X) = \frac{P(X|A)P(A)}{P(X)} : \frac{P(X|B)P(B)}{P(X)}$$

계산할  
필요 없음

$$= P(X|A)P(A) : P(X|B)P(B)$$

# 먼저 사전 확률부터

$P(A)$ ?  $P(B)$ ?

A 기계보다 B 기계가 조금 더 좋은 자리에 있어서  
일반적으로 더 많이 팔린다.

$$P(A) = 0.4, P(B) = 0.6$$

# Likelihood (우도)

$P(X|A)$  부터 계산해 보자

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

가정: 사탕 기계가 매우 커서, 그 안에 있는 사탕 수의  
비율은 몇 개를 꺼내도 일정하게 유지된다

# Likelihood

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

A에서 **빨강** 사탕을 꺼낼 확률 =  $\frac{2}{5}$

A에서 **노랑** 사탕을 꺼낼 확률 =  $\frac{2}{5}$

A에서 **초록** 사탕을 꺼낼 확률 =  $\frac{1}{5}$

# Likelihood

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

A에서 **빨강** 사탕 4개를 꺼낼 확률 =  $(\frac{2}{5})^4$

A에서 **노랑** 사탕 5개를 꺼낼 확률 =  $(\frac{2}{5})^5$

A에서 **초록** 사탕 1개를 꺼낼 확률 =  $(\frac{1}{5})^1$

# Likelihood

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

A에서 **빨강** 사탕 4개,

**노랑** 사탕 5개,

**초록** 사탕 1개를 꺼낼 확률 =  $\left(\frac{2}{5}\right)^4 \times \left(\frac{2}{5}\right)^5 \times \left(\frac{1}{5}\right)^1$

# Likelihood

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

$$P(X|A) = \left(\frac{2}{5}\right)^4 \times \left(\frac{2}{5}\right)^5 \times \left(\frac{1}{5}\right)^1$$

# Likelihood

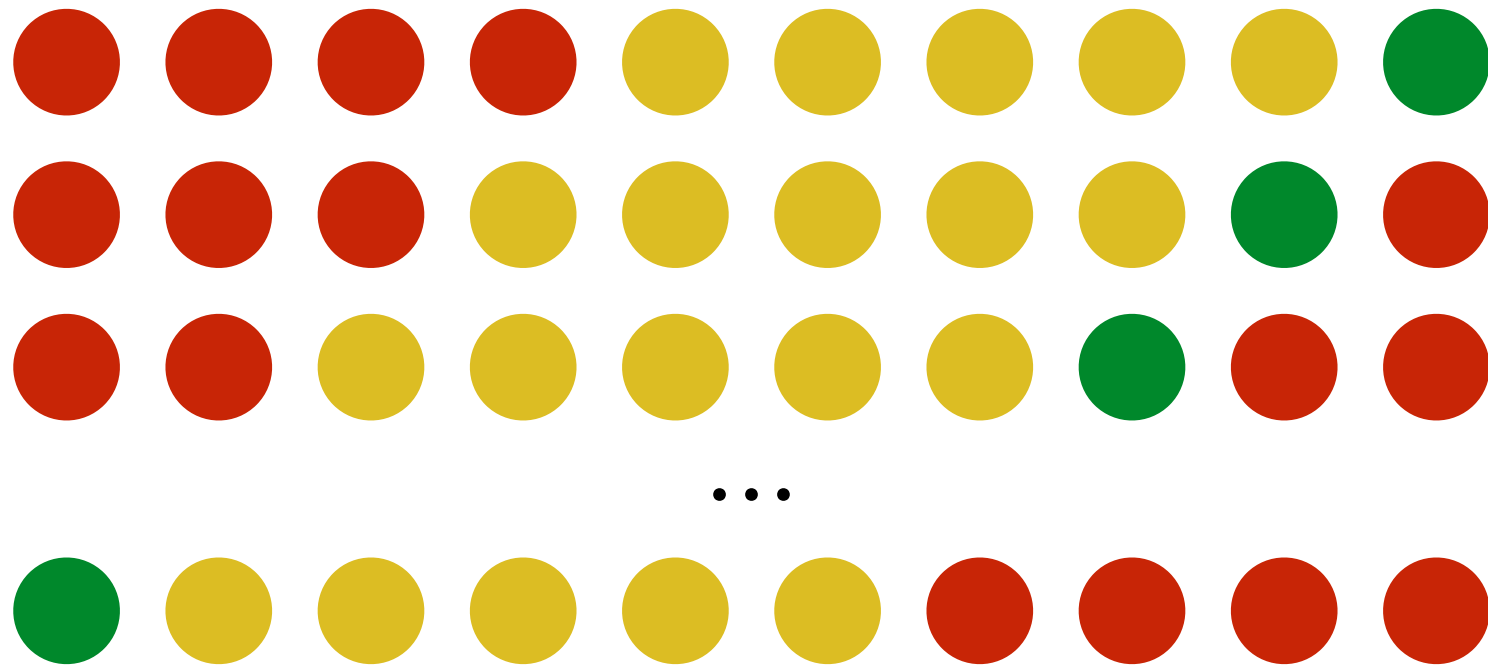
비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

$$P(X|A) = \left(\frac{2}{5}\right)^4 \times \left(\frac{2}{5}\right)^5 \times \left(\frac{1}{5}\right)^1$$

$$P(X|B) = \left(\frac{1}{3}\right)^4 \times \left(\frac{1}{3}\right)^5 \times \left(\frac{1}{3}\right)^1$$



# Likelihood



$$P(X|A) = \left(\frac{2}{5}\right)^4 \times \left(\frac{2}{5}\right)^5 \times \left(\frac{1}{5}\right)^1 \times C$$

$$P(X|B) = \left(\frac{1}{3}\right)^4 \times \left(\frac{1}{3}\right)^5 \times \left(\frac{1}{3}\right)^1 \times C$$

# Likelihood

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

$$P(X|A) = \left(\frac{2}{5}\right)^4 \times \left(\frac{2}{5}\right)^5 \times \left(\frac{1}{5}\right)^1 = 5.243 * 10^{-5}$$

$$P(X|B) = \left(\frac{1}{3}\right)^4 \times \left(\frac{1}{3}\right)^5 \times \left(\frac{1}{3}\right)^1 = 1.694 * 10^{-5}$$

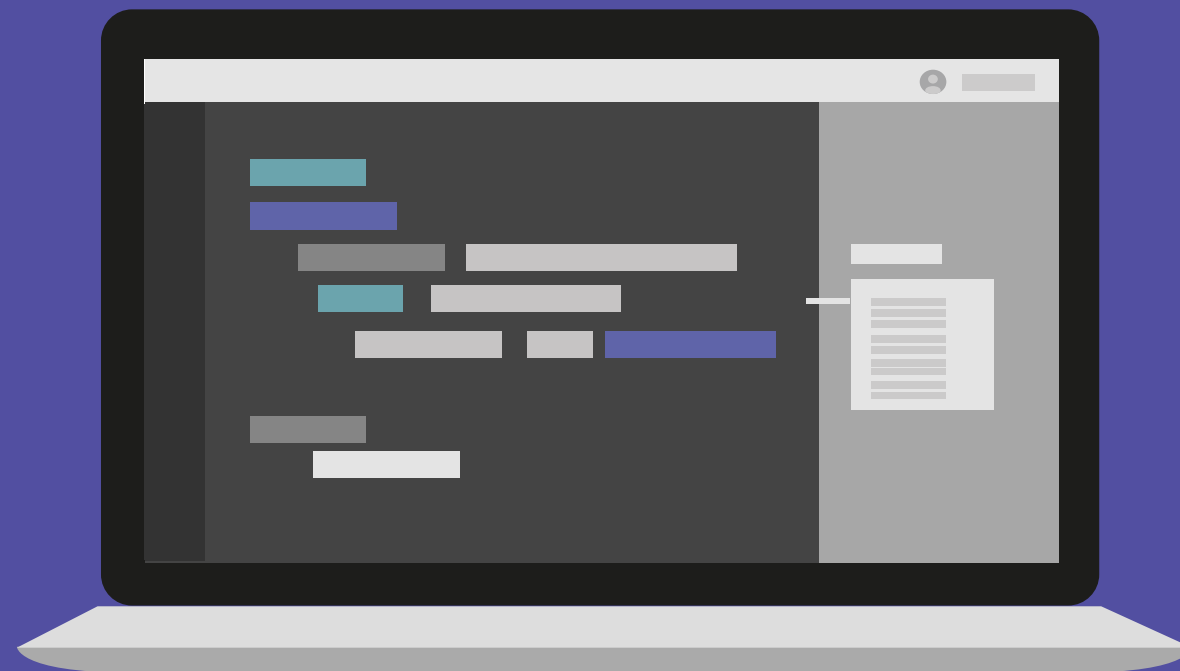
# Naive Bayes

$$\begin{aligned}P(A|X) : P(B|X) &= \frac{P(X|A)P(A)}{P(X)} : \frac{P(X|B)P(B)}{P(X)} \\&= P(X|A)P(A) : P(X|B)P(B) \\&= 5.243 * 10^{-5} \times 0.4 : 1.694 * 10^{-5} \times 0.6 \\&= 0.674 : 0.326\end{aligned}$$

꺼낸 10개의 사탕은 A에서 나왔을 확률이  
B에서 나왔을 확률보다 두 배 더 높다.

실습:

# Naive Bayes Classifier

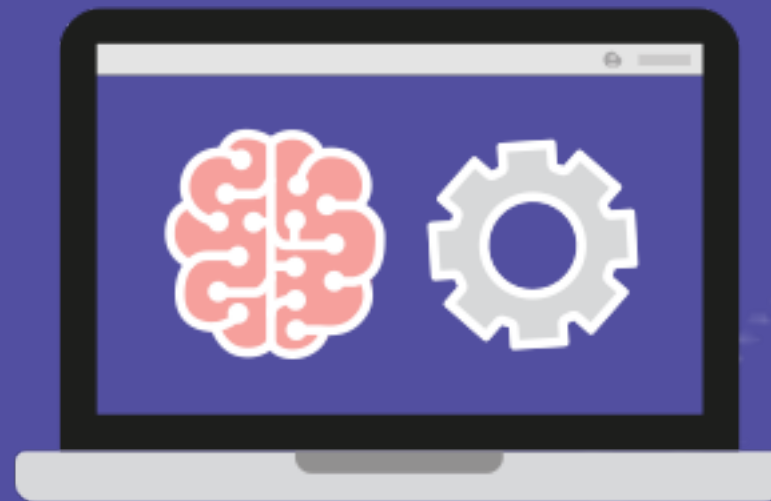


```
/* elice */
```

/\* elice \*/

# 인공지능/머신러닝 맛보기 - Python

## Module 7: Bag of Words 와 감정분석



# Bag of Words

`/* elice */`

# 텍스트 데이터의 분류

오늘의 목표:

감정분석 알고리즘을 만들어 본다



`/* elice */`

# 텍스트 데이터의 분류

비율	좋아	최고	싫어	별로
긍정	3	5	1	1
부정	1	1	2	4

긍정  
사탕



부정  
사탕





# Bag of Words

오늘 나는 밥을 먹었다. 어제 나는 햄버거를 먹었다.

↓ 특수 문자 제거

오늘 나는 밥을 먹었다 어제 나는 햄버거를 먹었다

↓ Tokenize

오늘 / 나는 / 밥을 / 먹었다 /  
어제 / 나는 / 햄버거를 / 먹었다

# Bag of Words

오늘 / 나는 / 밥을 / 먹었다 /  
어제 / 나는 / 햄버거를 / 먹었다



/\* elice \*/

# Bag of Words



BoW 모델에서

순서는

중요하지 않다

**Python Dictionary로 표현하면 쉽다.**

# Bag of Words



BoW 모델에서

순서는

중요하지 않다

{"오늘": 1, "나는": 2, "먹었다": 2,  
"햄버거를": 1, "밥을": 1, "어제": 1}

# Bag of Words

오늘 나는 밥을 먹었다. 어제 나는 햄버거를 먹었다.



{"오늘": 1, "나는": 2, "먹었다": 2,  
"햄버거를": 1, "밥을": 1, "어제": 1}

# 사탕 기계와 Bag of Words

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1



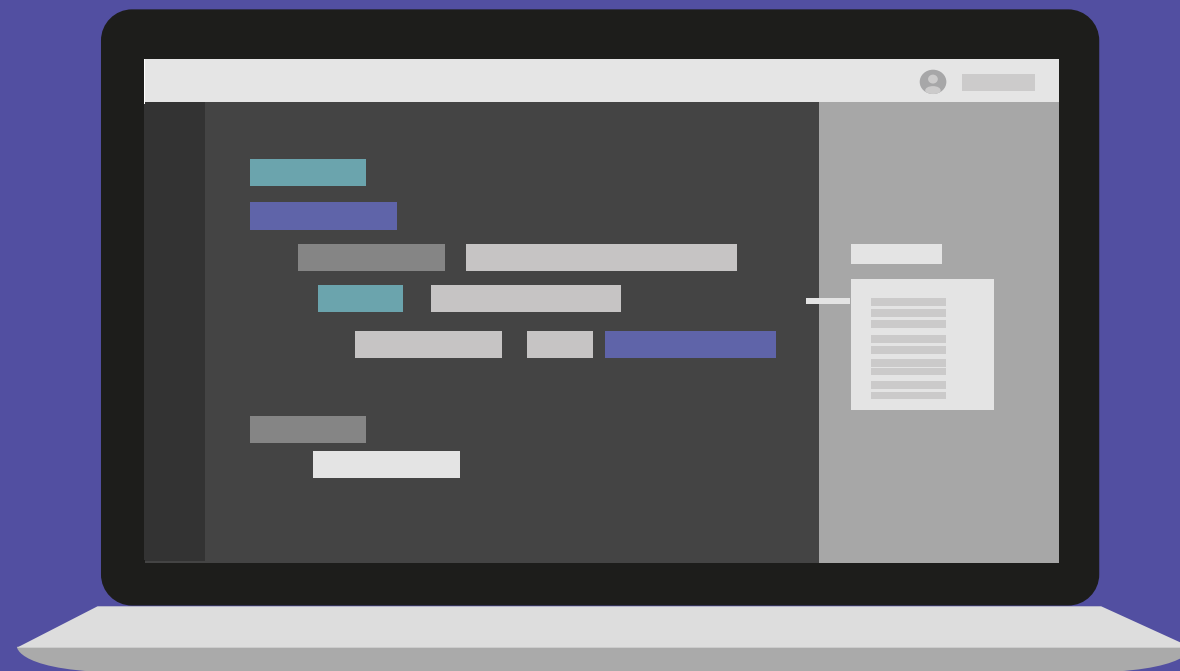
# 사탕 기계와 Bag of Words

비율	빨강색	노랑색	초록색
A	2	2	1
B	1	1	1

A {"초록색": 1, "노랑색": 2, "빨강색": 2}

B {"빨강색": 1, "초록색": 1, "노랑색": 1}

# 실습: Bag of Words



`/* elice */`



# 감정 분류기

## Sentiment Classifier

# 사탕 기계와 Bag of Words

A {"초록색" : 1, "노랑색" : 2, "빨강색" : 2}

B {"빨강색" : 1, "초록색" : 1, "노랑색" : 1}

A 기계에서 노랑색 사탕이 나올 확률은?  $\frac{2}{5}$

$$\frac{\text{노랑색 사탕의 개수}}{\text{초록} + \text{노랑} + \text{빨강 사탕의 개수}} = \frac{\text{노랑색 사탕의 개수}}{\text{전체 사탕의 개수}}$$

# 사탕 기계와 Bag of Words

A {"초록색" : 1, "노랑색" : 2, "빨강색" : 2}

B {"빨강색" : 1, "초록색" : 1, "노랑색" : 1}

B 기계에서 초록색 사탕이 나올 확률은?  $\frac{1}{3}$

$$\frac{\text{초록색 사탕의 개수}}{\text{초록} + \text{노랑} + \text{빨강 사탕의 개수}} = \frac{\text{초록색 사탕의 개수}}{\text{전체 사탕의 개수}}$$

# Likelihood 다시 보기

A {"초록색" : 1, "노랑색" : 2, "빨강색" : 2}

B {"빨강색" : 1, "초록색" : 1, "노랑색" : 1}

A에서 빨강 사탕 4개,

노랑 사탕 5개,

초록 사탕 1개를 꺼낼 확률 =  $\left(\frac{2}{5}\right)^4 \times \left(\frac{2}{5}\right)^5 \times \left(\frac{1}{5}\right)^1$

# Naive Bayes

$$\begin{aligned}P(A|X) : P(B|X) &= \frac{P(X|A)P(A)}{P(X)} : \frac{P(X|B)P(B)}{P(X)} \\&= P(X|A)P(A) : P(X|B)P(B) \\&= 5.243 * 10^{-5} \times 0.4 : 1.694 * 10^{-5} \times 0.6 \\&= 0.674 : 0.326\end{aligned}$$

이것과 **똑같이**

# Sentiment Classifier

긍정 {"좋아": 3, "최고": 5, "싫어": 1, "별로": 1}

부정 {"좋아": 1, "최고": 1, "싫어": 2, "별로": 4}

“긍정” 기계에서 “최고” 단어가 나올 확률은?  $\frac{5}{10}$

$$\frac{\text{“최고” 단어의 개수}}{\text{전체 단어의 개수}}$$

# Sentiment Classifier

긍정 {"좋아": 3, "최고": 5, "싫어": 1, "별로": 1}

부정 {"좋아": 1, "최고": 1, "싫어": 2, "별로": 4}

“부정” 기계에서 “싫어” 단어가 나올 확률은?  $\frac{2}{8}$

$$\frac{\text{“싫어” 단어의 개수}}{\text{전체 단어의 개수}}$$

# 학습: Training

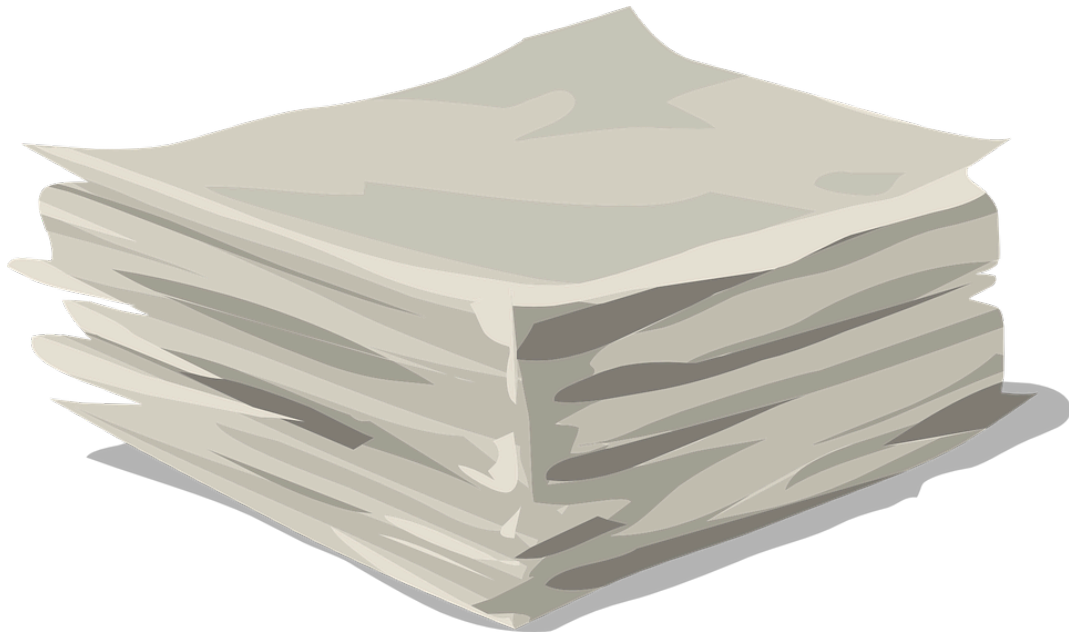
긍정/부정 BoW는 어떻게 알아내야 하나?

손으로? 단어 사전? ...

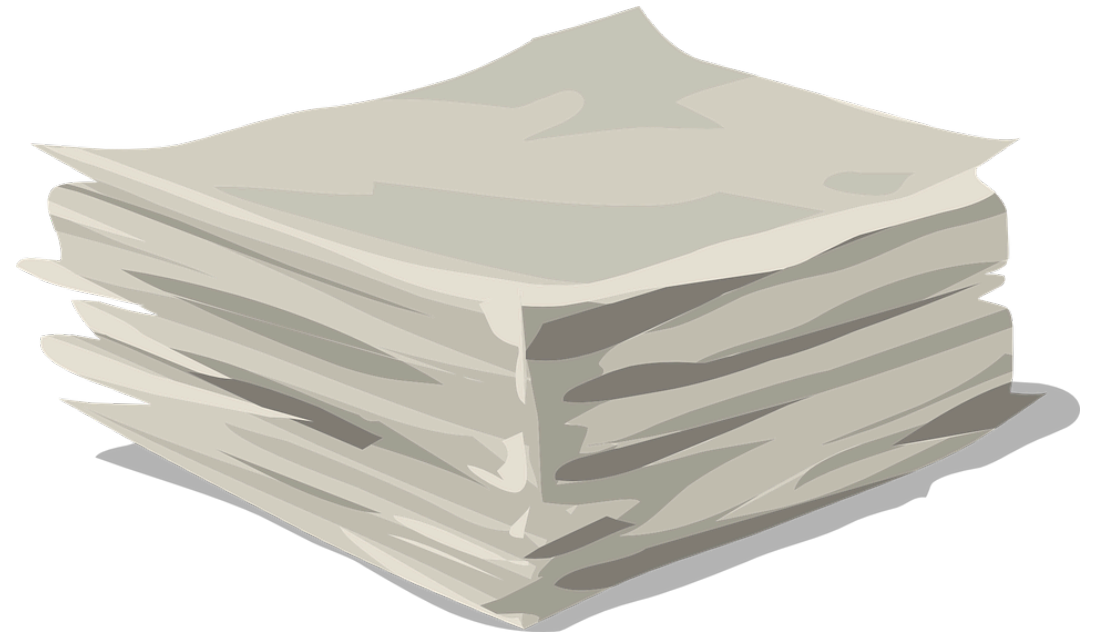
기계학습 알고리즘은  
자동으로 학습되게 한다.



# 학습: Training



긍정적인 문서 2000개



부정적인 문서 2000개

각각의 문서 셋들에서 나오는 단어의 빈도 수를 측정

# 트레이닝 결과

각각의 문서 셋들에서 나오는 단어의 빈도 수를 측정

부정

```
[('영화', 11452),  
 ('너무', 5417),  
 ('진짜', 4698),  
 ('정말', 3836),  
 ('그냥', 3537),  
 ('이', 3279),  
 ('왜', 3108),  
 ('이런', 2659),  
 ('이건', 2253),  
 ('좀', 2203),  
 ('다', 2132),  
 ('더', 2036),  
 ('영화는', 2013),  
 ('쓰레기', 1991),  
 ...]
```

긍정

```
[('영화', 13930),  
 ('정말', 7544),  
 ('너무', 6046),  
 ('진짜', 4325),  
 ('이', 3988),  
 ('최고의', 2776),  
 ('수', 2695),  
 ('잘', 2630),  
 ('더', 2436),  
 ('보고', 2178),  
 ('최고', 2035),  
 ('영화를', 2005),  
 ('ㅋㅋ', 1998),  
 ('그', 1960),  
 ...]
```

# 트레이닝 결과

부정  
전체 단어 수  
796,466

[ ('영화', 11452),  
('너무', 5417),  
('진짜', 4698),  
('정말', 3836),  
('그냥', 3537),  
('이', 3279),  
('왜', 3108),  
('이런', 2659),  
('이건', 2253),  
('좀', 2203),  
('다', 2132),  
('더', 2036),  
('영화는', 2013),  
('쓰레기', 1991),  
...

$$P(\text{“쓰레기”}|부정) = 0.00249$$

긍정  
전체 단어 수  
773,502

[ ('영화', 13930),  
('정말', 7544),  
('너무', 6046),  
('진짜', 4325),  
('이', 3988),  
('최고의', 2776),  
('수', 2695),  
('잘', 2630),  
('더', 2436),  
('보고', 2178),  
('최고', 2035),  
('영화를', 2005),  
('ㅋㅋ', 1998),  
('그', 1960),  
...

$$P(\text{“쓰레기”}|긍정) = 0.000111$$

# 트레이닝 결과

## 마음이 따뜻해지는 최고의 영화

$$P(\text{“마음이”}|\text{부정}) = 0.000498$$

$$P(\text{“마음이”}|\text{긍정}) = 0.000228$$

$$P(\text{“따뜻해지는”}|\text{부정}) = 0.0000142$$

$$P(\text{“따뜻해지는”}|\text{긍정}) = 0.00135$$

$$P(\text{“최고의”}|\text{부정}) = 0.000111$$

$$P(\text{“최고의”}|\text{긍정}) = 0.00249$$

$$P(\text{“영화”}|\text{부정}) = 0.0180$$

$$P(\text{“영화”}|\text{긍정}) = 0.0143$$

$$1.420 * 10^{-14}$$

:

$$1.114 * 10^{-11}$$

$$0.001$$

:

$$0.999$$

# 단어가 없으면 어떡할까?

이거 볼 시간에 엘리스 에서 공부하자.

$$P(\text{“시간에”}|\text{부정}) = 0.0000313 \quad P(\text{“시간에”}|\text{긍정}) = 0.0000459$$

$$P(\text{“엘리스”}|\text{부정}) = 0 \quad P(\text{“엘리스”}|\text{긍정}) = 0$$

$$P(\text{“에서”}|\text{부정}) = 0.00184 \quad P(\text{“에서”}|\text{긍정}) = 0.000838$$

$$P(\text{“공부하자”}|\text{부정}) = 0.00484 \quad P(\text{“공부하자”}|\text{긍정}) = 0.00112$$

0 : 0

0 : 0

# 단어가 없으면 어떡할까?

이거 볼 시간에 엘리스 에서 공부하자.

$$P(\text{“시간에”}|부정) = 0.0000313 \quad P(\text{“시간에”}|긍정) = 0.0000459$$

$$P(\text{“엘리스”}|부정) = 0.000000001 \quad P(\text{“엘리스”}|긍정) = 0.000000001$$

$$P(\text{“에서”}|부정) = 0.00184 \quad P(\text{“에서”}|긍정) = 0.000838$$

$$P(\text{“공부하자”}|부정) = 0.00484 \quad P(\text{“공부하자”}|긍정) = 0.00112$$

$$2.787 * 10^{-18} \quad : \quad 4.308 * 10^{-19}$$

$$0.866 \quad : \quad 0.133$$

`/* elice */`

**문의 및 연락처**

[academy.elice.io](http://academy.elice.io)

[contact@elice.io](mailto:contact@elice.io)

[facebook.com/elice.io](https://facebook.com/elice.io)

[blog.naver.com/elicer](https://blog.naver.com/elicer)