/* elice */

수포자를 위한 프로그래밍 수학

데이터 분석 첫 걸음

조웅오 선생님

목차

개요

3장을 배우고 나면!

1. 확률 이론에서의 경우의 수에 대해 알게됩니다.
2. 경우의 수를 구하는 다양한 방법을 알게됩니다.
3. 확률의 의미와 계산법을 알게 됩니다.
4. 독립사건과 종속사건을 구분할 수 있게 됩니다.
5. 데이터를 분석하는 방법 중 회귀 분석에 대해 알게됩니다.

경우의 수

사건이 발생했을 때 일어날 수 있는 경우의 개수

동전 던지기

{앞면, 뒷면} - **2가지**

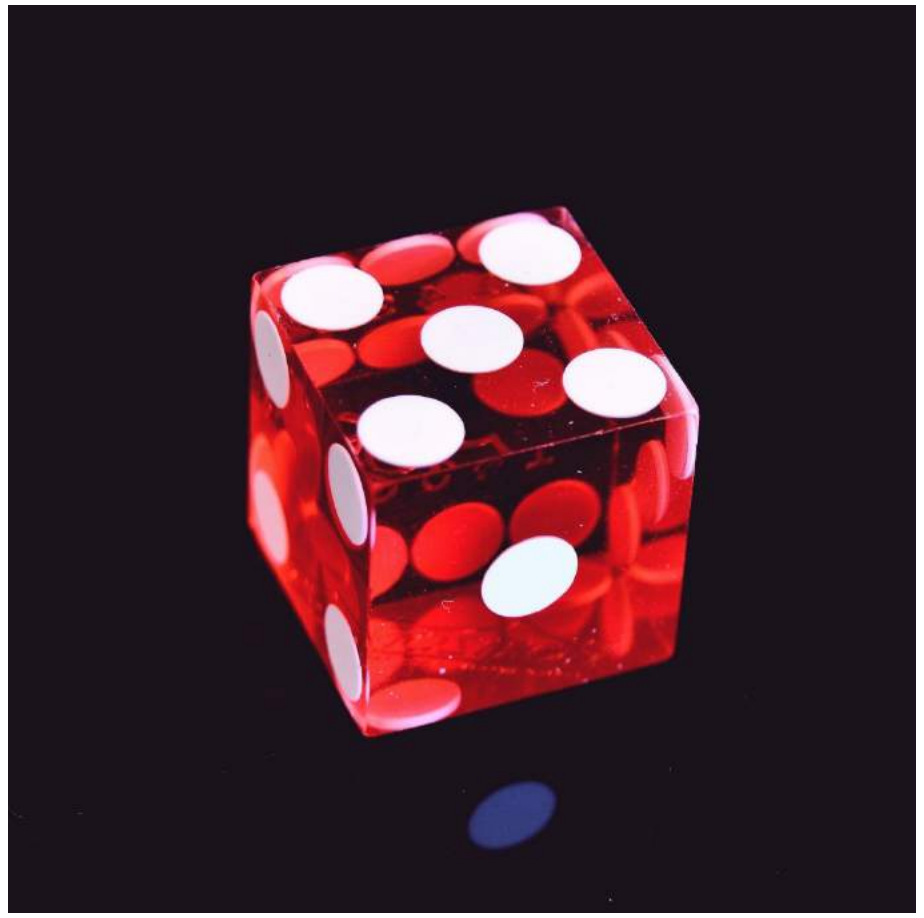
주사위 던지기

{1, 2, 3, 4, 5, 6} - **6가지**

합의 법칙

사건 A 또는 B가 일어나는 경우의 수

= A의 경우의 수 + B의 경우의 수



합의 법칙

- 3이 나오는 경우의 수: {3} (1)
- 4가 나오는 경우의 수: {4} (1)
- 3 또는 4가 나오는 경우의 수: {3, 4} (1+1)

곱의 법칙

사건 A, B가 **동시에** 일어나는 경우의 수

= A의 경우의 수 × B의 경우의 수

곱의 법칙

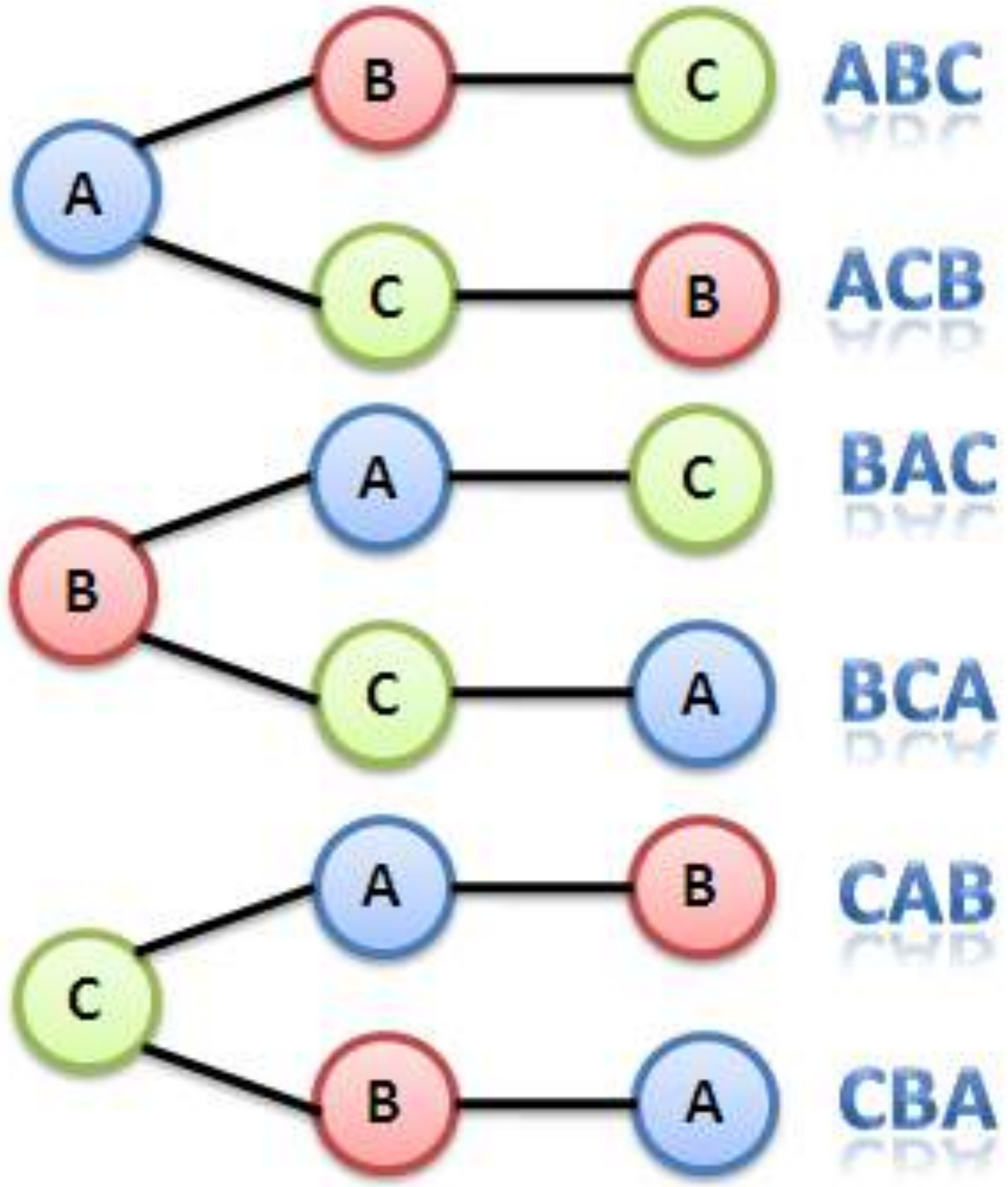
- A: 주사위 1이 3 또는 4가 나오는 경우: {3, 4} (2)
- B: 주사위 2가 5 또는 6이 나오는 경우: {5, 6} (2)
- 사건 A와 B가 동시에 일어나는 경우

  {(3, 5), (3, 6), (4, 5), (4, 6)} (2×2)

순열

n개 중 r개를 뽑아 나열하는 경우의 수

A, B, C를 나열하는 경우의 수는?



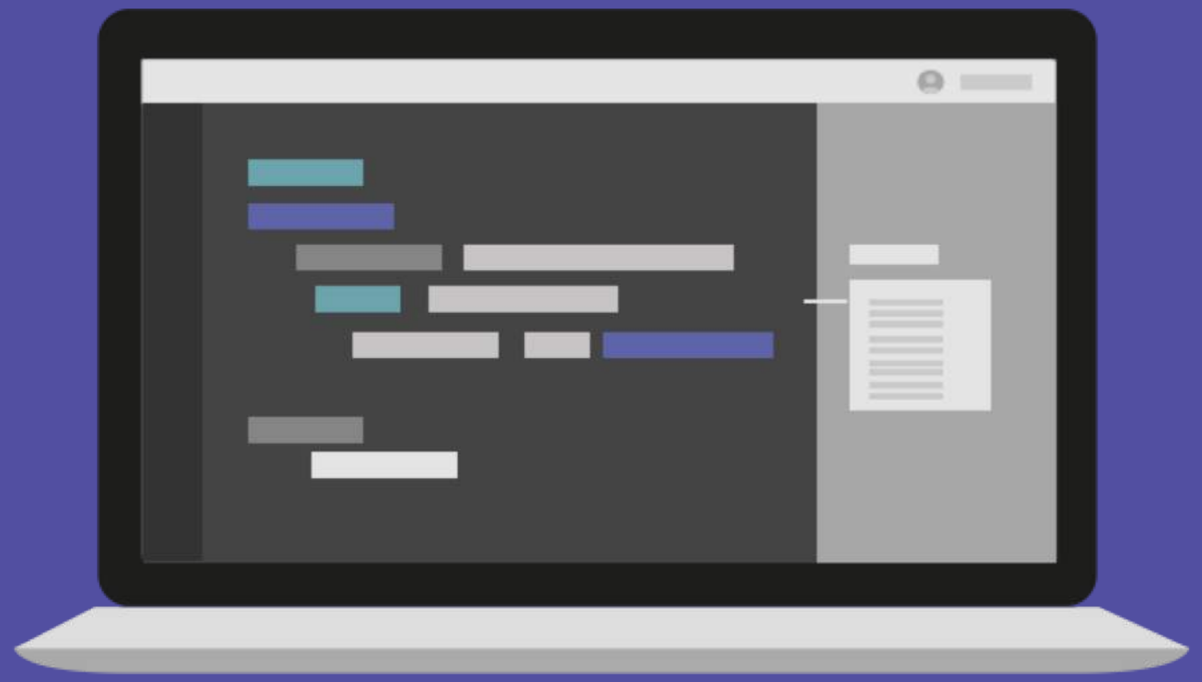
순열

n개 중 r개를 뽑아 나열하는 경우의 수

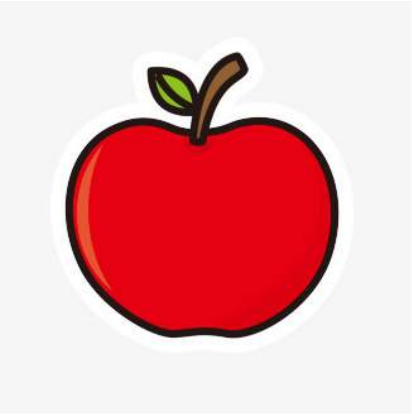
$$_nP_r = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1)$$

$$= \frac{n!}{(n-r)!}$$

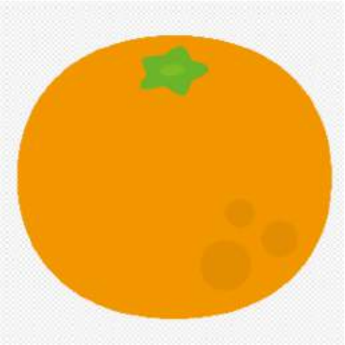
[실습1] 순열



같은 것이 있는 순열



사과 a개와 오렌지 b개를
나열하는 경우의 수는?



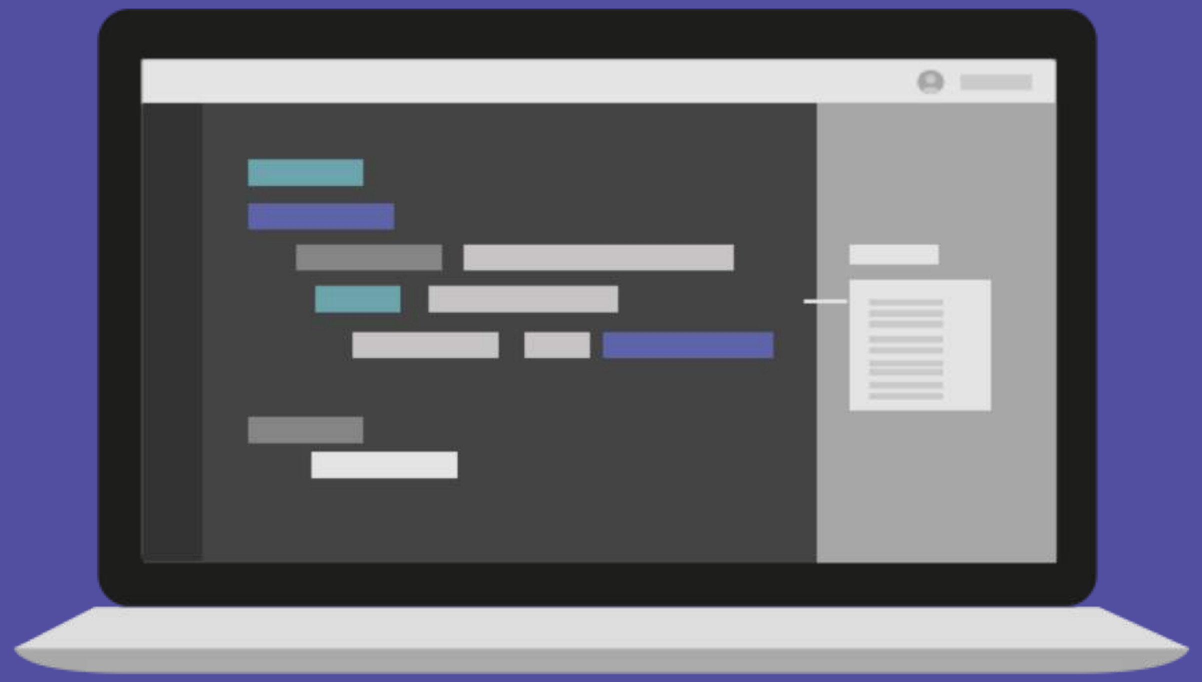
$$\frac{(a+b)!}{a!\times b!}$$

같은 것이 있는 순열

사과 a개와 오렌지 b개 중 r개를 뽑아
나열하는 경우의 수는?

모든 경우를 직접 보는 방법 밖에 없다…

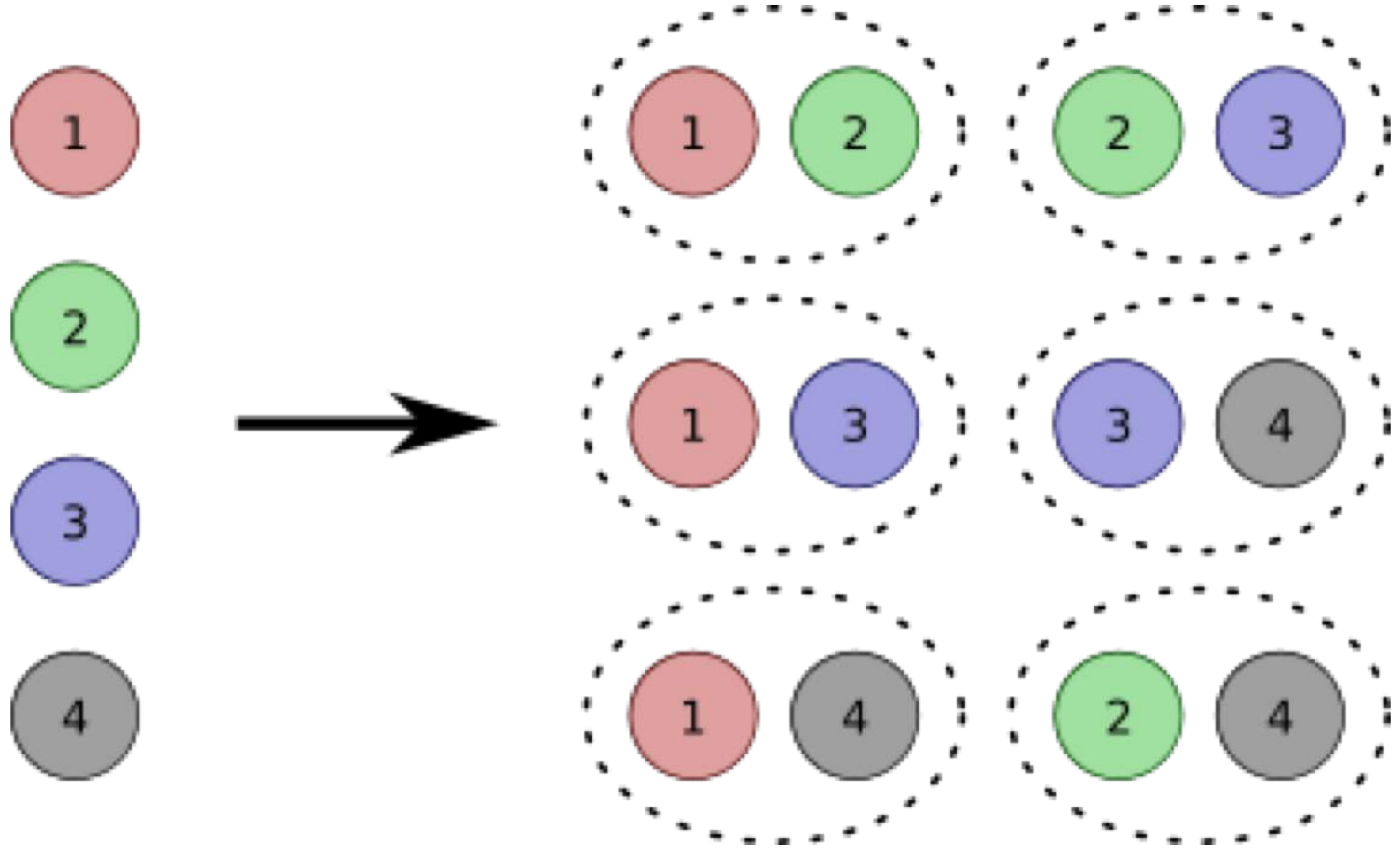
[실습2] 같은 것이 있는 순열



조합

n개 중 r개를 순서에 관계 없이 뽑는 경우의 수

4개 중 2개를 순서에 관계 없이 뽑는 경우의 수는?

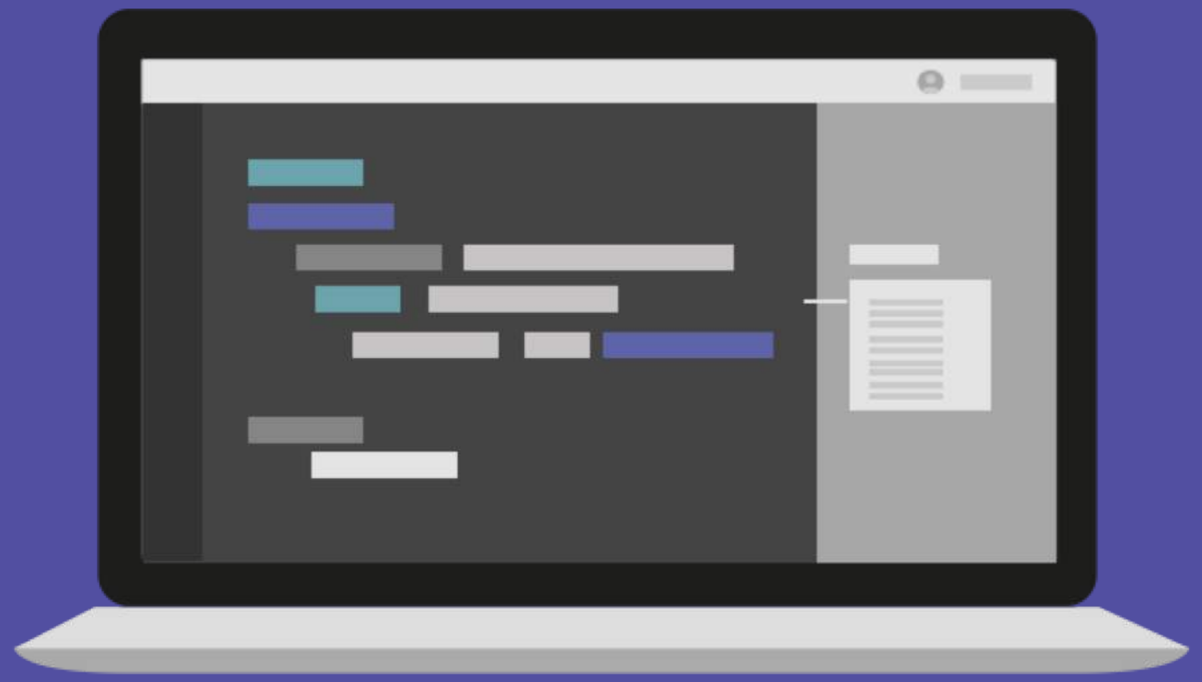


조합

n개 중 r개를 순서에 관계 없이 뽑는 경우의 수

$$_{n}C_{r} = \frac{_{n}P_{r}}{r!} = \frac{n!}{(n-r)!r!}$$

[실습3] 조합



확률의 정의

어떤 사건이 어느 정도의 비율로 일어나는가

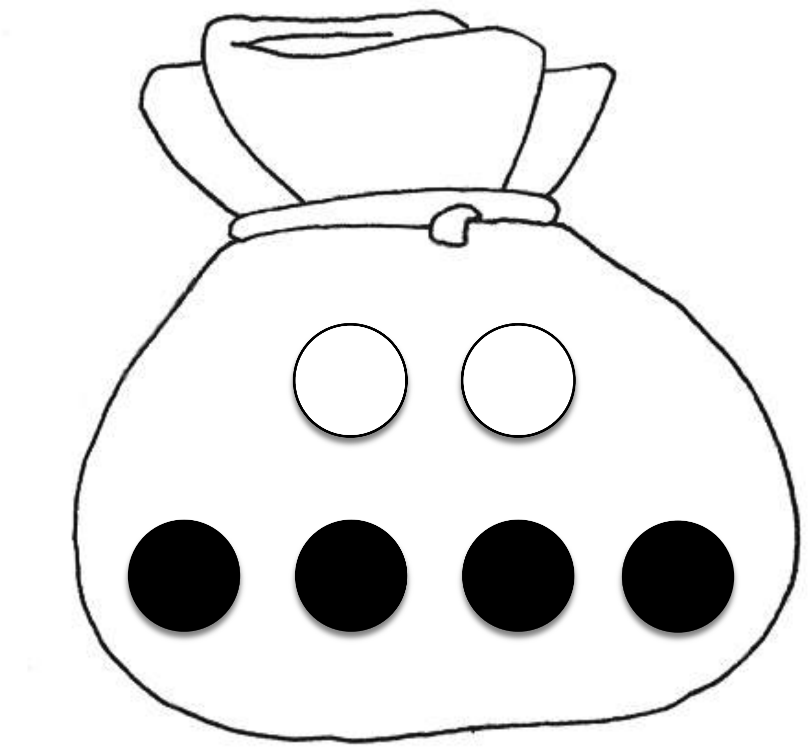
$P(A)$ **= (사건A가 일어날 확률)**

$$(\text{근원사건이 일어날 확률}) = \frac{(\text{근원사건의 원소의 수})}{(\text{전체 원소의 수})}$$

확률의 정의

어떤 사건이 어느 정도의 비율로 일어나는가

흰공 2개와 검은공 4개가 든 주머니에서
공 하나를 뽑는 사건



경우의 수: { ● ○ }

원소의 수: { ● ● ● ● ○ ○ }

통계적 확률

확률을 직접 세어서 구해보자!

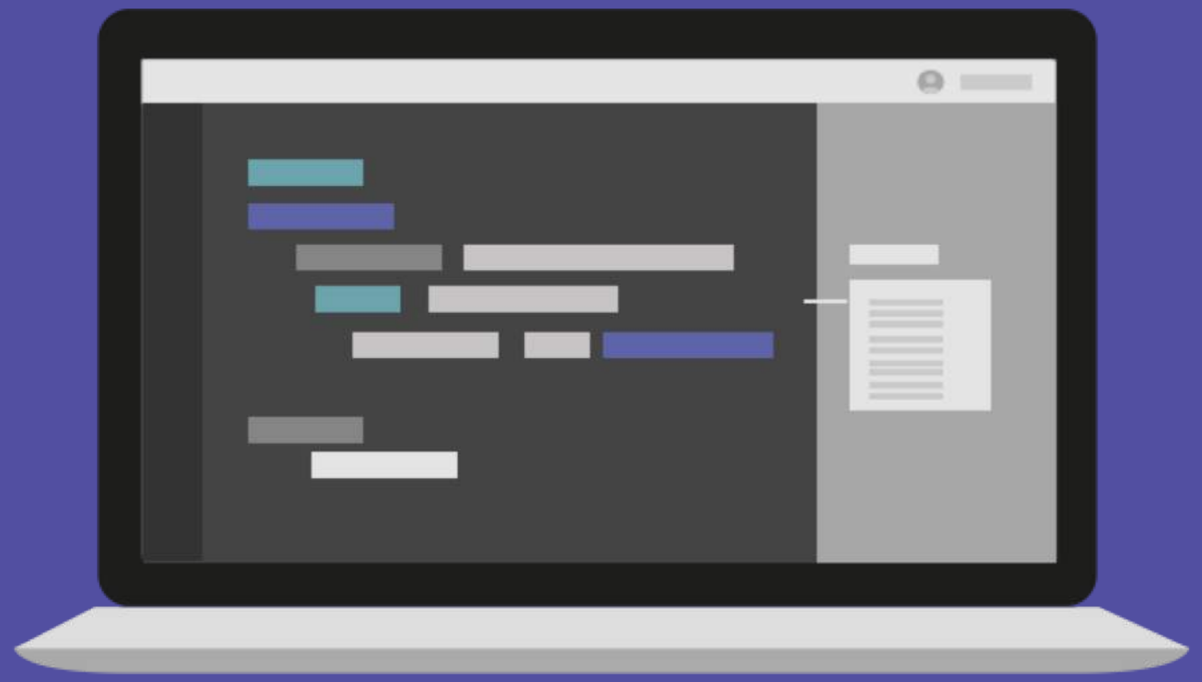
 앞면 5번

 뒷면 4번

앞면이 나올 확률

-> 5/9

[실습4] 통계적 확률



통계적 확률

무수히 많이 시행한다면 수적 확률에 수렴

…

앞면이 나올 통계적 확률 ≒ 1/2

독립사건

다른 사건의 영향을 받지 않는 사건

흰색을 뽑을 확률은 언제나 같다.

종속사건

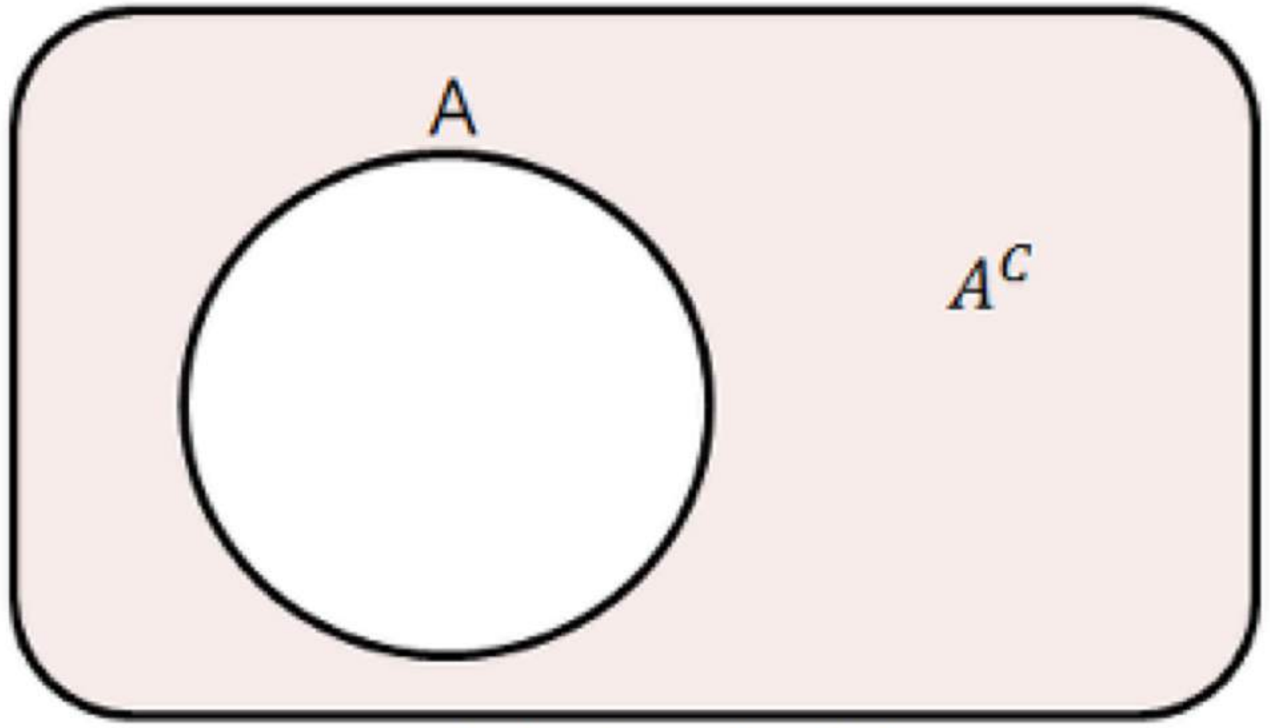
다른 사건의 영향을 받는 사건

흰색을 뽑을 확률은 이전의 결과에 종속

여사건

사건이 일어나지 않는 경우

표본공간 = 전사건 (전체집합 U)



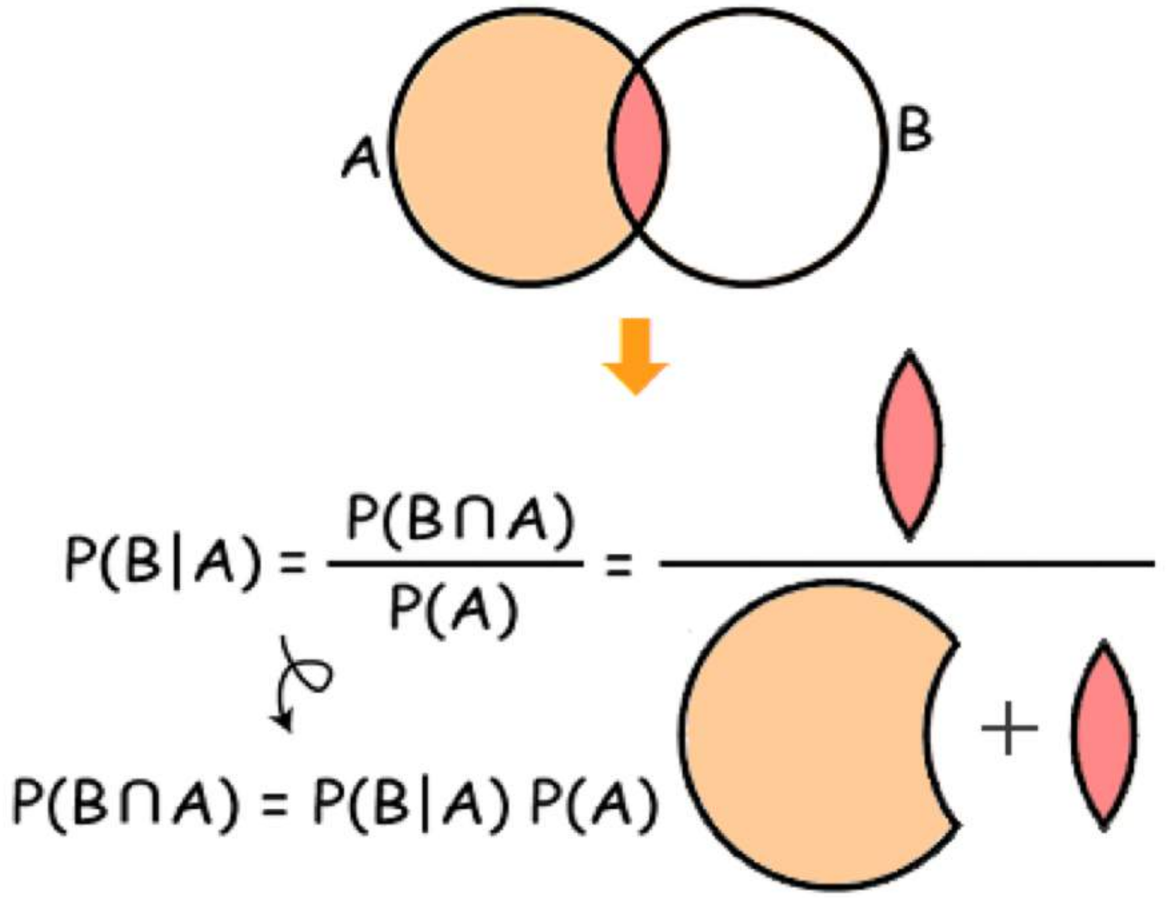
여사건 (여집합)

$$P(A) + P(A^c) = 1$$

조건부 확률

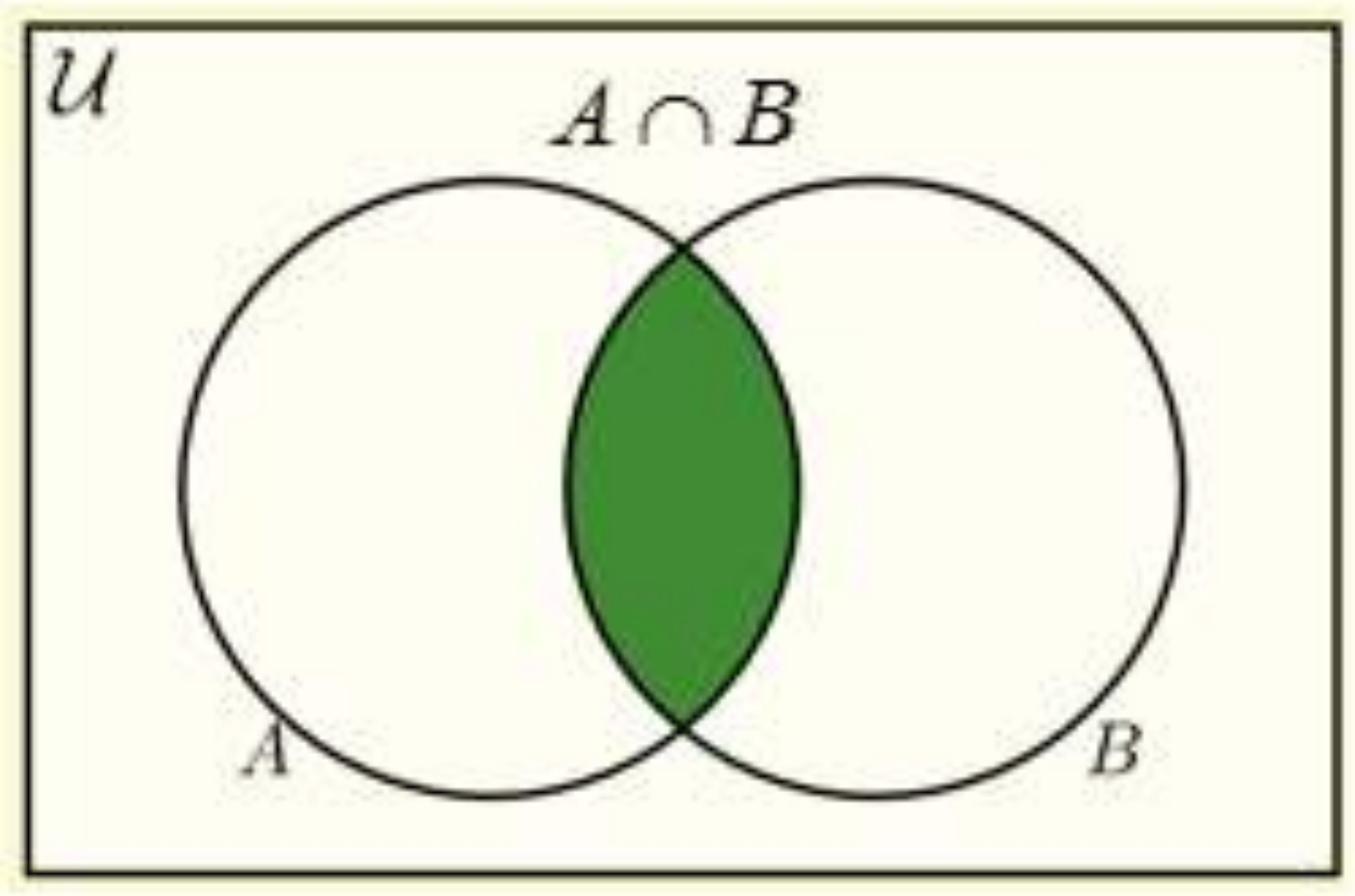
종속사건의 확률 구하기

A가 일어났을 때 종속사건 B의 확률



곱사건

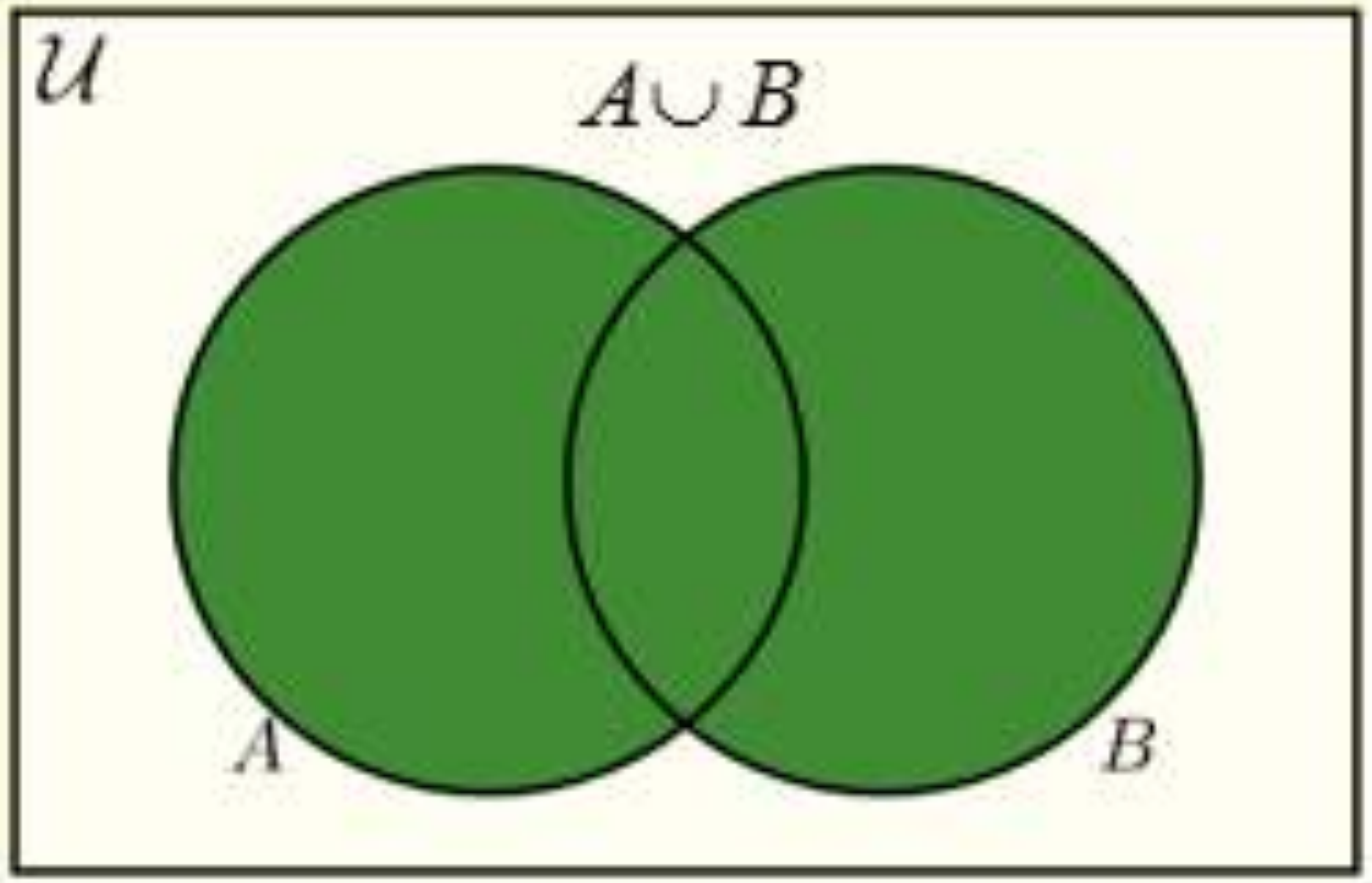
동시에 일어나는 사건의 확률



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

합사건

A 또는 B가 일어나는 사건



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

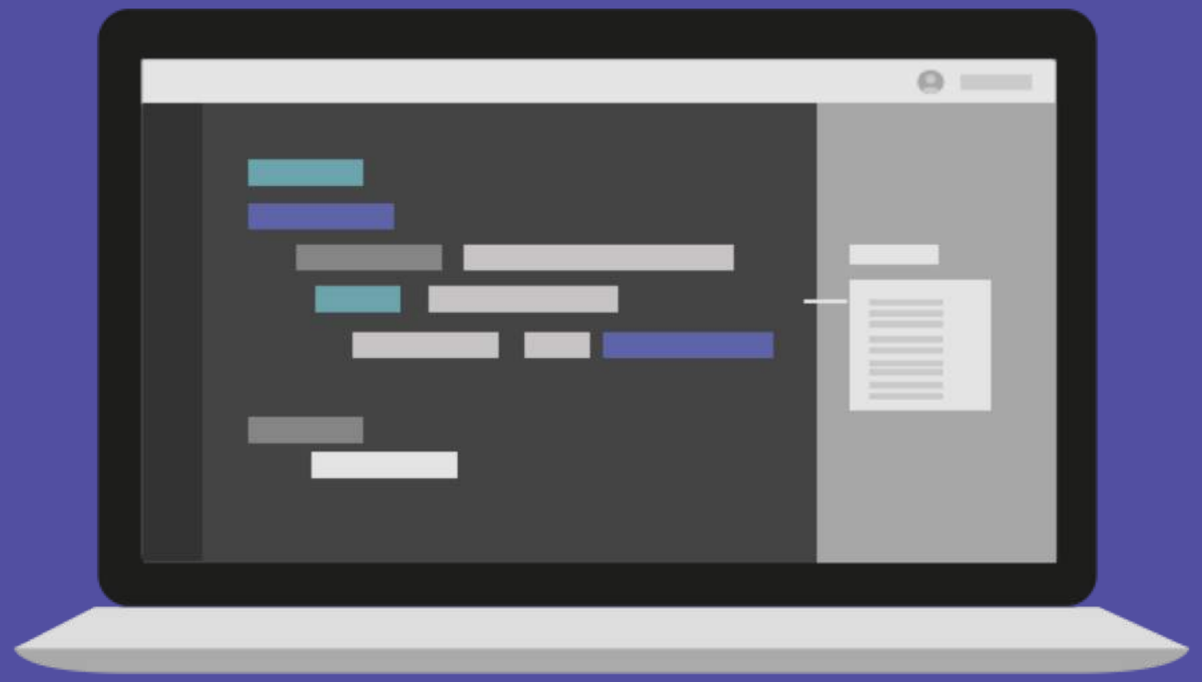
독립시행

n번의 시행 중 사건이 r번 일어날 확률

10번 중 앞면이 3번 나오는 경우의 수는?

(A가 n번 중 r번 일어날 확률) = $P(A)^r P(A^c)^{n-r} \times \frac{n!}{r!(n-r)!}$

[실습5] 독립시행



베이지안 확률

주어진 정보를 바탕으로 미래를 예측하자

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

알고 싶은 정보

알고 있는 정보

베이지안 확률

스팸 메일 분류하기

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

A: 받은 메일이 스팸 메일인 사건

B: 받은 메일에 ‘공짜’ 단어가 포함되어 있는 사건

P(A|B): 받은 메일에 ‘공짜’ 단어가 포함된 메일일 때,
이 메일이 스팸메일일 확률은?

베이지안 확률

스팸 메일 분류하기

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

P(A), P(B), P(B|A)를 알면 계산 가능하다.

-> 과거의 기록을 바탕으로 계산

P(A): 받은 메일이 스팸 메일일 확률

P(B): 받은 메일에 '공짜' 단어가 포함될 확률

P(B|A): 받은 스팸 메일에 '공짜' 단어가 포함될 확률

베이지안 확률

스팸 메일 분류하기

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

P(A): 0.5
P(B): 0.4
P(B|A): 0.6

P(A|B) = 0.5 × 0.6 ÷0.4 = 0.75 (75%)

베이지안 확률

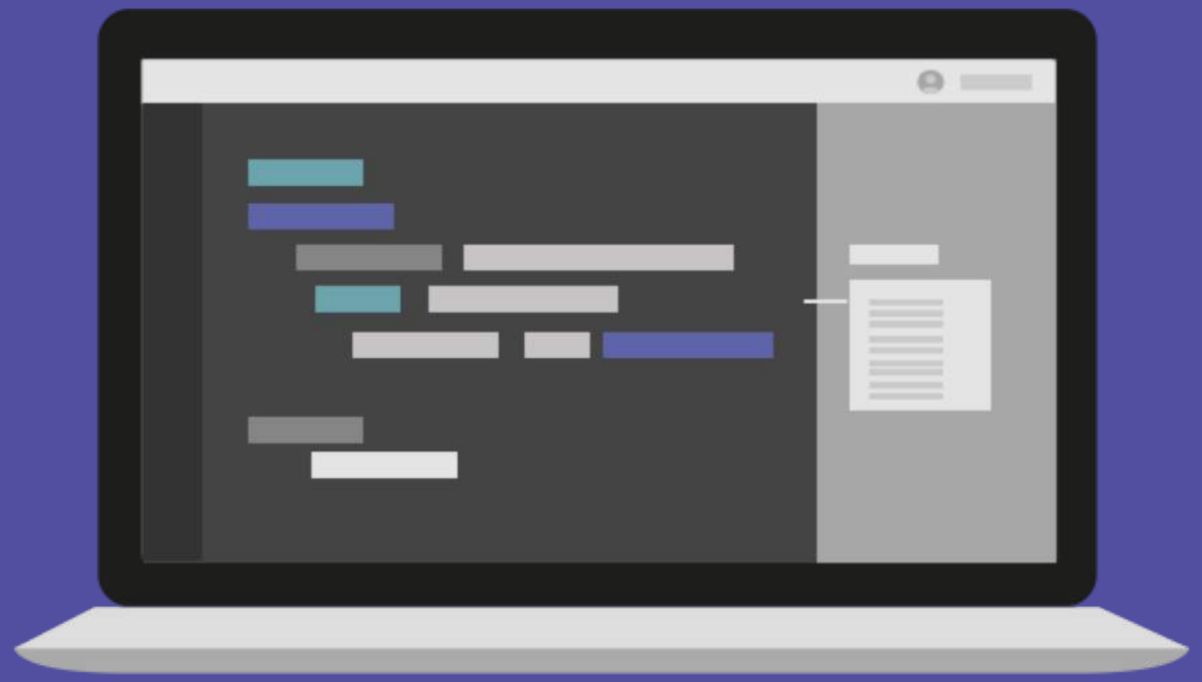
장점:

- 비교적 계산이 간단하다
- 데이터가 많다면 상당히 잘 맞는다

단점:

- 전례가 없으면 계산이 불가능하다
- 데이터가 적으면 신뢰도가 낮다

[실습6] 베이지안 확률



이번 장에서는!

1. 경우의 수의 정의와 합의 법칙,
   곱의 법칙에 대해 배웠습니다.
2. 경우의 수를 계산하는 방법으로
   순열과 조합에 대해 배웠습니다.
3. 확률의 정의와 계산하는 방법을 배웠습니다.
4. 베이지안 확률을 이용한 가장 기초적인
   데이터 분석 법을 배웠습니다.

/* elice */

문의 및 연락처

academy.elice.io
contact@elice.io
facebook.com/elice.io
medium.com/elice