

선수 지식 - 통계

확률 개요

확률 개요 | 딥러닝의 기초가 되는 확률 개념 알아보기

강사 나동빈

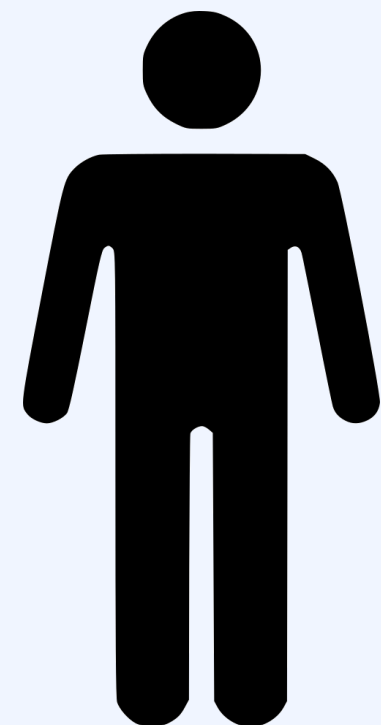
선수 지식 - 통계

확률 개요

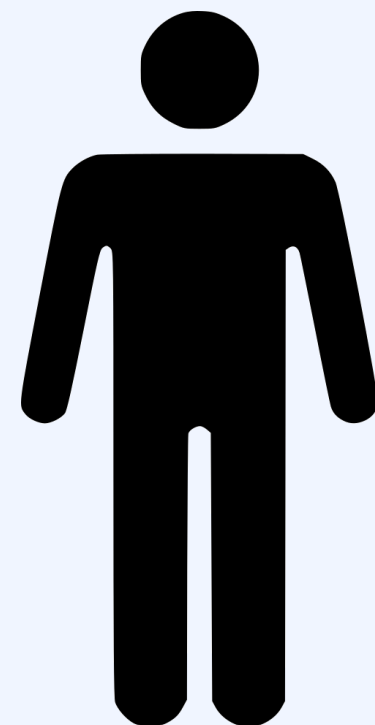
확률(Probability)이란?

- 확률이란, 특정한 사건이 일어날 가능성을 수로 표현한 것을 의미한다.
- 확률은 0부터 1(100%) 사이의 실수로 표현된다.

내가 **내일** 지각할
확률은 얼마일까?



친구 1

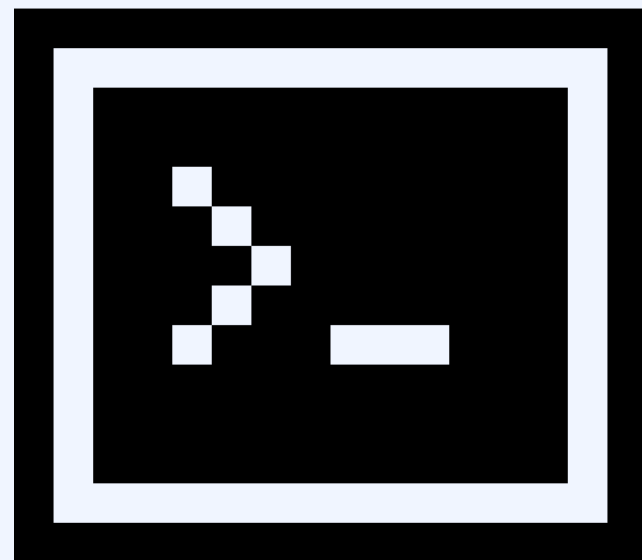


친구 2

지금까지는 매일
지각했으니 **100%** 아닐까?

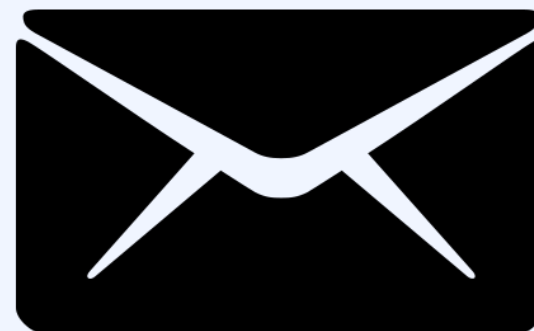
확률이 사용되는 예시

- 스팸 메일 분류 모델을 만들고 싶다고 가정해보자.
- 지금까지 도착한 메일이 총 1,000개이다.
- 이 중에서 700개는 스팸 메일, 300개는 정상 메일이었다.
- 새롭게 하나의 메일이 도착했을 때, 그것이 스팸 메일일 확률은 얼마일까?



스팸 분류 모델

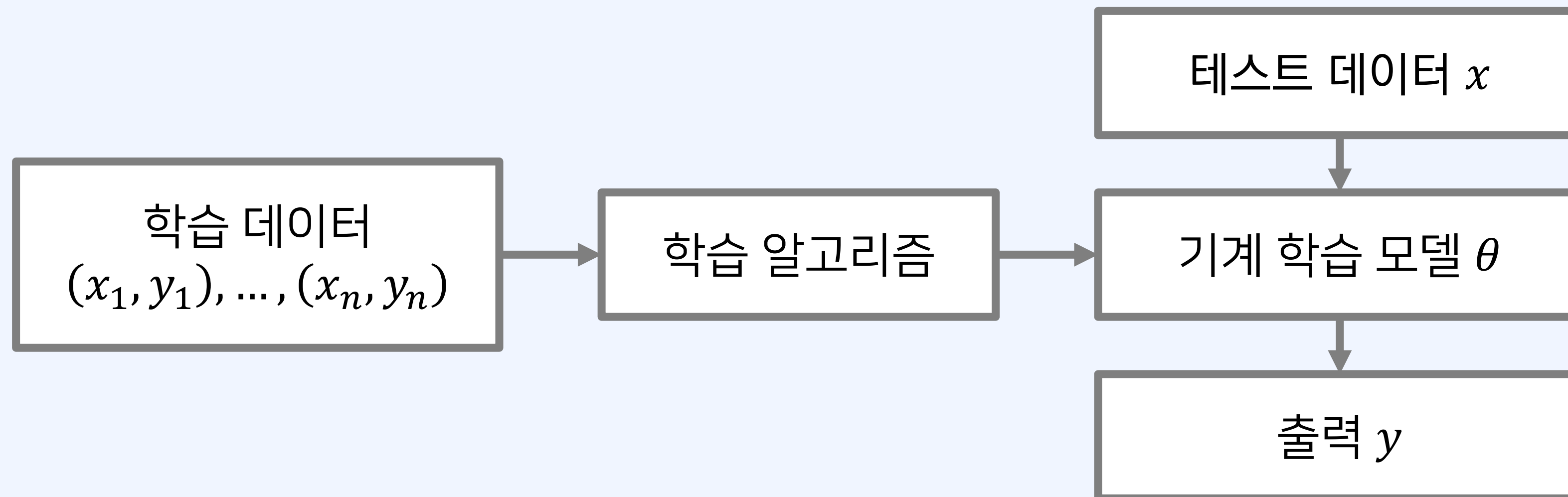
이 메일은 **7/10의 확률**로
스팸 메일이군!



새로운 메일

기계학습 모델을 확률적으로 이해하기

- 우리는 N 개의 학습 데이터로 기계학습 모델을 학습한다.
- 일반적으로 기계학습 모델의 출력은 “확률” 형태를 띈다.
- 예시 1) 이미지 분류 모델 θ 가 이미지 x 에 대해서 75% 확률로 고양이일 것이라고 예측했다.
- 예시 2) 글 쓰기 모델 θ 는 “나는 밥을” 이후에 “먹었다”라는 단어가 나올 확률을 42%로 예측했다.



- 다양한 문제 상황에서의 **확률**을 계산해보자.
- 이를 위해 기본적으로 **경우의 수**를 계산하는 방법을 알아야 한다.
- 경우의 수를 계산하는 방법으로는 **순열(permutation)**과 **조합(combination)**이 있다.

순열(Permutation)

- 서로 다른 n 개에서 r 개를 중복 없이 뽑아 특정한 순서로 나열하는 것을 의미한다.
- 모든 순열의 수(경우의 수)는 다음의 공식으로 계산할 수 있다.

1) 순열의 수 공식($n = r$ 일 때)

$${}_nP_r = n!$$

2) 순열의 수 공식($n \neq r$ 일 때)

- [참고] $0!$ 은 1이다.

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

순열(Permutation) 예시 - 학습에 사용할 이미지의 순서

- 기계 학습 모델 학습을 위해 N 개의 학습 데이터로 학습을 진행하려고 한다.
- 이때, 매번(epoch) 모델에 학습 데이터를 넣을 순서를 섞어서(shuffling) 학습을 진행한다.
- 에포크(epoch): 학습 데이터 세트에 포함된 모든 데이터가 한 번씩 모델을 통과한 횟수를 말한다.



순열(Permutation) 예시 - 학습에 사용할 이미지의 순서

- $N = 3$ 개의 이미지가 있을 때, 3개를 뽑아 임의의 순서로 나열하는 모든 경우의 수는?
- 가지고 있는 학습 이미지를 {1, 2, 3}이라고 했을 때, 가능한 모든 경우는 다음과 같다.
- 모든 경우: [1, 2, 3], [1, 3, 2], [2, 1, 3], [2, 3, 1], [3, 1, 2], [3, 2, 1]
- 경우의 수: $3! = 3 \times 2 \times 1 = 6$

순열(Permutation) 예시 - 학습에 사용할 이미지의 순서

- $N = 4$ 개의 이미지가 있을 때, 2개를 뽑아 임의의 순서로 나열하는 모든 경우의 수는?
- 가지고 있는 학습 이미지를 {1, 2, 3, 4}이라고 했을 때, 가능한 모든 경우는 다음과 같다.
- 모든 경우: [1, 2], [1, 3], [1, 4], [2, 1], [2, 3], [2, 4], [3, 1], [3, 2], [3, 4], [4, 1], [4, 2], [4, 3]
- 경우의 수: $4 \times 3 = 12$

파이썬(Python)을 이용한 순열 계산

- 파이썬을 이용하면 전체 순열을 간단히 계산할 수 있다.

```
from itertools import permutations

arr = ['A', 'B', 'C']
# 원소 중에서 2개를 뽑는 모든 순열 계산
result = list(permutations(arr, 2))
print(result)
```

[실행 결과]

```
[('A', 'B'), ('A', 'C'), ('B', 'A'), ('B', 'C'),  
('C', 'A'), ('C', 'B')]
```

[참고] 아직 파이썬 문법을 잘 모르시는 분께서는

1. 파이썬 문법 강의 파트를 듣고 강의로 돌아와 주시거나
2. 현재는 개괄적으로 이해하고, 추후 다시 복습해주세요.

조합(Combination)

- 서로 다른 n 개에서 r 개를 중복 없이 순서를 고려하지 않고 뽑는 것을 의미한다.
- 모든 조합의 수(경우의 수)는 다음의 공식으로 계산할 수 있다.
- 조합의 수 공식

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

조합(Combination) 예시 - 삼(Siamese) 네트워크

- 딥러닝 아키텍처 중에서 삼 네트워크(Siamese network)가 있다.
- 두 개의 이미지를 받아서, 두 이미지가 유사하다면 1을, 다르다면 0을 출력한다.
- 이때 N 개의 이미지로 구성된 데이터 세트가 있을 때, 매번 2개의 이미지를 뽑아서 2개의 이미지를 네트워크의 입력으로 넣는다. 이 경우, 두 이미지에 대하여 순서는 상관없다.
- 예를 들어 4개의 이미지가 있을 때, 순서를 고려하지 않고 2개의 이미지를 뽑는 경우의 수는?
- 전체 이미지를 {1, 2, 3, 4}라고 했을 때, 가능한 모든 경우는 다음과 같다.
- 모든 경우: [1, 2], [1, 3], [1, 4], [2, 3], [2, 4], [3, 4]
- 경우의 수: $4 \times 3 / 2 = 6$

파이썬(Python)을 이용한 조합 계산

- 파이썬을 이용하면 전체 조합을 간단히 계산할 수 있다.

```
from itertools import combinations

arr = ['A', 'B', 'C']
# 원소 중에서 2개를 뽑는 모든 조합 계산
result = list(combinations(arr, 2))
print(result)
```

[실행 결과]

```
[('A', 'B'), ('A', 'C'), ('B', 'C')]
```

중복 순열(Permutation with Repetition)

- 서로 다른 n 개에서 중복을 포함해 r 개를 뽑아 특정한 순서로 나열하는 것을 의미한다.
- 모든 중복 순열의 수(경우의 수)는 다음의 공식으로 계산할 수 있다.
- 중복 순열의 수 공식

$${}_n\Pi_r = n^r$$

중복 순열(Permutation with Repetition) 예시

- 세 개의 숫자 1, 2, 3을 이용해 만들 수 있는 두 자리 자연수의 개수는?
- (단, 같은 숫자를 여러 번 사용하거나, 사용하지 않은 숫자가 있어도 괜찮다.)
- 모든 경우: 11, 12, 13, 21, 22, 23, 31, 32, 33
- 경우의 수: $3^2 = 9$

파이썬(Python)을 이용한 중복 순열 계산

- 파이썬을 이용하면 전체 중복 순열을 간단히 계산할 수 있다.

```
from itertools import product

arr = ['A', 'B', 'C']
# 원소 중에서 2개를 뽑는 모든 중복 순열 계산
result = list(product(arr, repeat=2))
print(result)
```

[실행 결과]

```
[('A', 'A'), ('A', 'B'), ('A', 'C'), ('B', 'A'),
('B', 'B'), ('B', 'C'), ('C', 'A'), ('C', 'B'),
('C', 'C')]
```

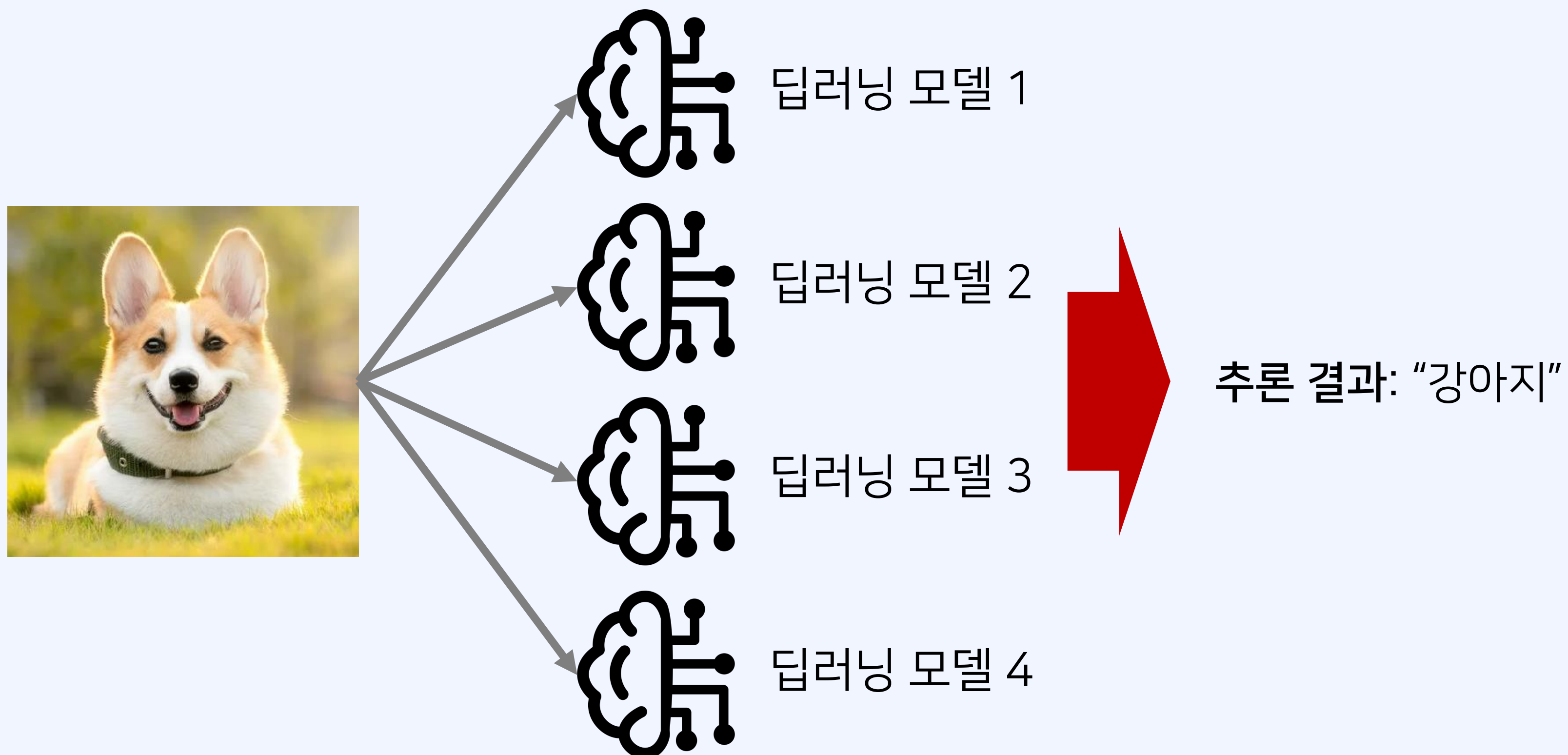
중복 조합(Combination with Repetition)

- 서로 다른 n 개에서 중복을 포함해 순서를 고려하지 않고 r 개를 뽑는 것을 의미한다.
- 모든 중복 조합의 수(경우의 수)는 다음의 공식으로 계산할 수 있다.
- 중복 조합의 수 공식

$${}_nH_r = {}_{n+r-1}C_r$$

중복 조합(Combination with Repetition) 예시

- 딥러닝에서는 학습된 여러 모델의 결과를 활용하여 최종적인 결과를 생성하는 앙상블(ensemble) 방법이 존재한다.



중복 조합(Combination with Repetition) 예시

- 서로 다른 딥러닝 아키텍처 1번, 2번, 3번이 존재한다.
- 중복을 허용하여 4개를 골라서 딥러닝 모델 4개를 학습하고자 한다.

[문제] 이때 가능한 학습의 모든 조합의 수는?

- 같은 아키텍처를 여러 번 사용하거나, 선택하지 않는 아키텍처가 있어도 된다.

딥러닝 아키텍처 1

딥러닝 아키텍처 2

딥러닝 아키텍처 3

- 예시) 학습할 4개의 모델을 [1, 2, 3, 1]로 선택할 수 있다.
- 예시) 1번 아키텍처만 4개 학습하여 [1, 1, 1, 1]도 가능하다.
- 학습 순서는 상관 없기 때문에 [1, 2, 3, 1]과 [1, 2, 1, 3]은 같은 경우에 속한다.

파이썬(Python)을 이용한 중복 조합 계산

- 파이썬을 이용하면 전체 중복 조합을 간단히 계산할 수 있다.

```
from itertools import combinations_with_replacement

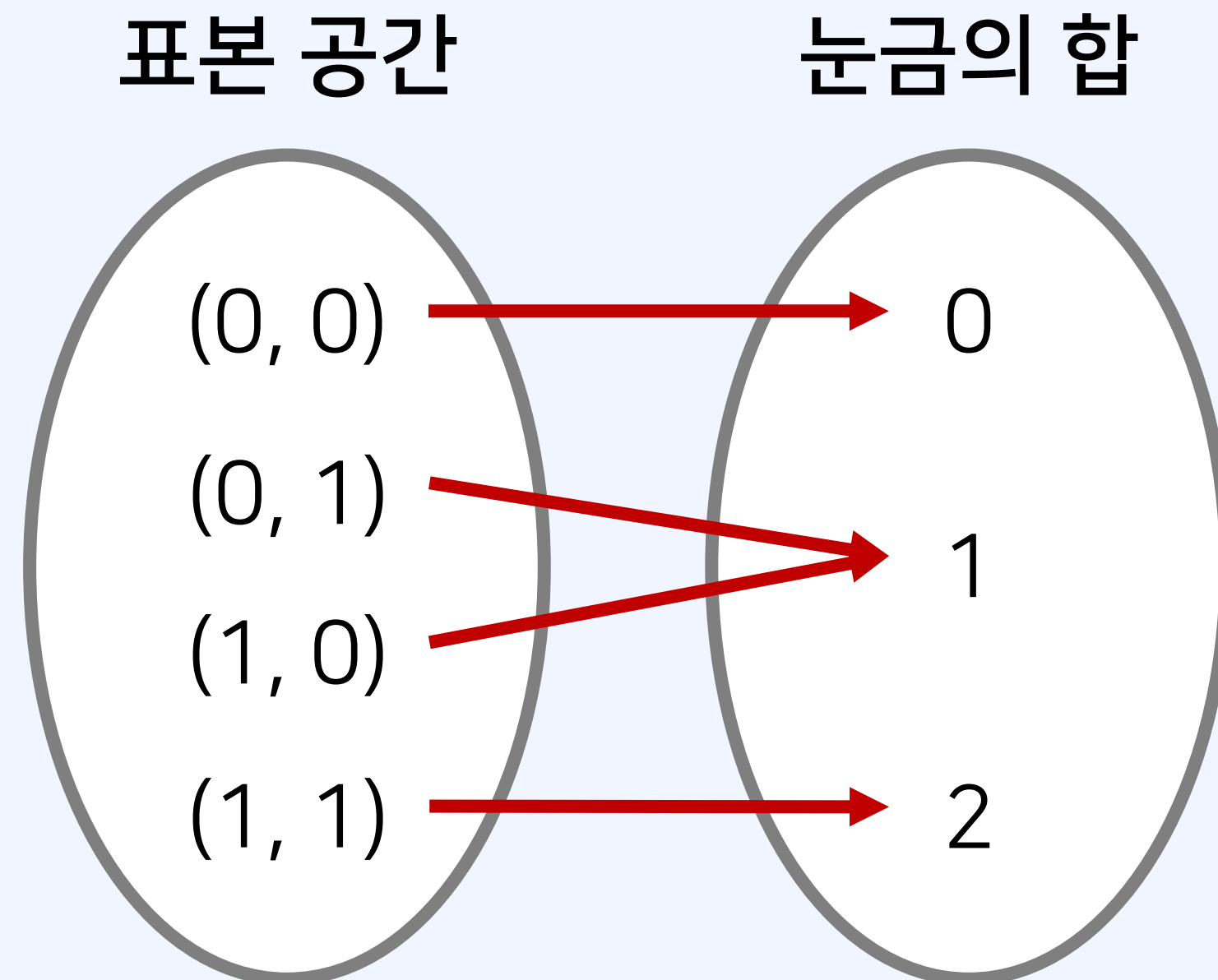
arr = ['A', 'B', 'C']
# 원소 중에서 2개를 뽑는 모든 중복 조합 계산
result = list(combinations_with_replacement(arr, 2))
print(result)
```

[실행 결과]

```
[('A', 'A'), ('A', 'B'), ('A', 'C'), ('B', 'B'), ('B', 'C'), ('C', 'C')]
```

확률(Probability)

- S 를 전체 사건(event)의 집합(표본 공간 = sample space)라고 하자.
- 사건 X 가 일어날 확률(probability) $P(X)$ 는 다음과 같다.
- $P(X) = \text{사건 } X \text{가 일어나는 경우의 수} / \text{전체 경우의 수} = n(X)/n(S)$
- 앞면에 1, 뒷면에 0이 쓰여 있는 2개의 동전을 2번 던졌을 때, 눈금의 합이 1일 확률은?



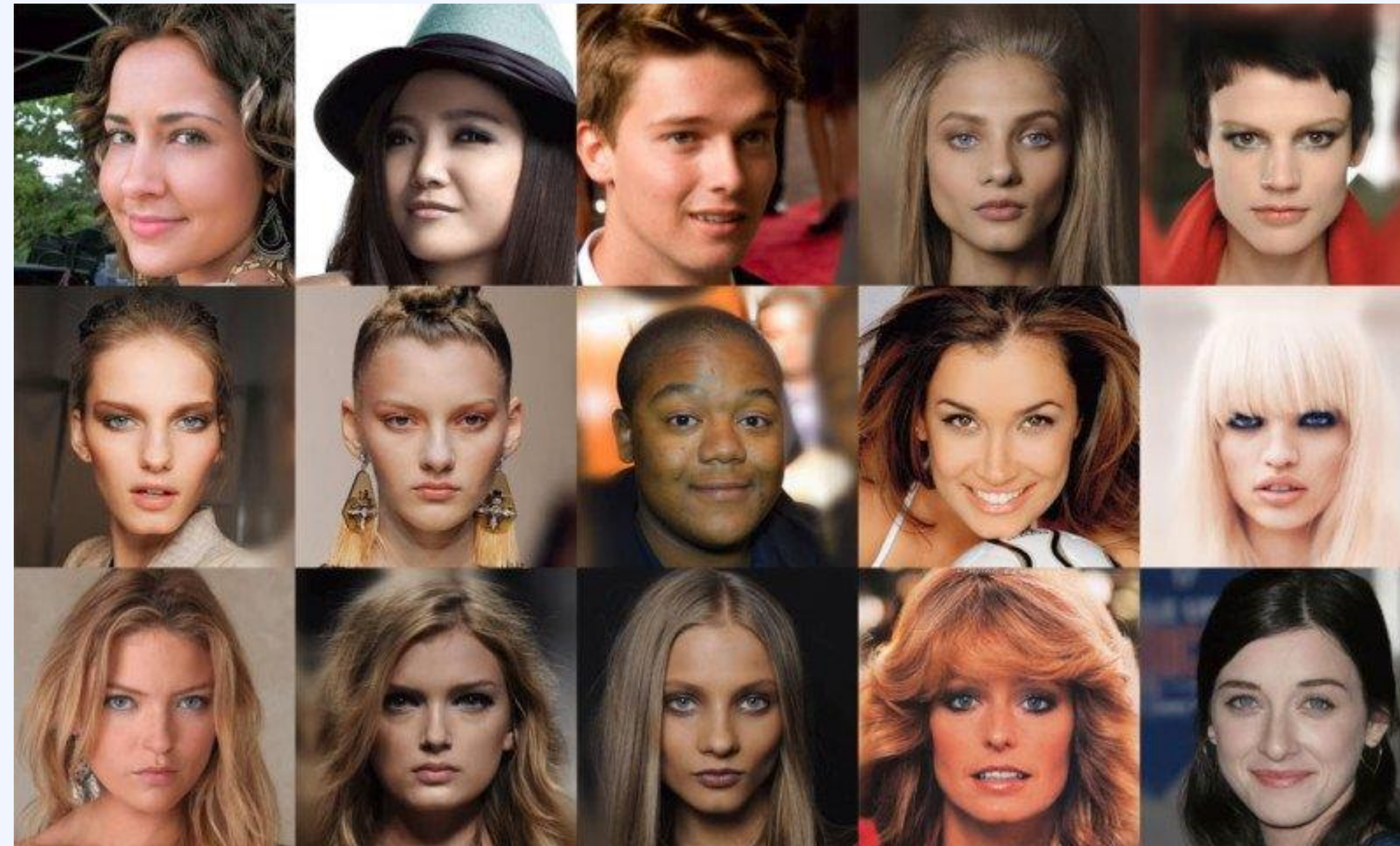
- 주사위를 던졌을 때 눈금이 5가 나올 확률은 $1/6$ 다.
- 그렇다면, 6번 던졌을 때, 반드시 눈금이 5인 경우가 나올까?
- 시행 횟수가 적기 때문에 장담할 수 없다.
- 하지만 6,000번 던졌다면 1,000번에 가까운 횟수로 등장할 것이다.

- 동일한 시행을 N 번 반복해 사건 X 가 발생한 횟수를 R 이라고 하자.
- 시행 횟수 N 을 무한히 크게 만들었을 때, R/N 이 수렴하는 값을 사건 X 의 통계적 확률이라 한다.
- 프로그램을 이용해 많은 시행을 수행하여, 확률을 계산해 볼 수 있다.
- 목표 확률: 0.167

시행 횟수	6	60	600	6000	60000
눈금이 5인 횟수	2	12	104	1025	10074
확률	0.333	0.200	0.173	0.171	0.1679

실제 사례 예시 - 생성 모델

- 생성 모델은 실제로 존재하지 않지만, 그럴싸한 이미지를 만들 수 있다.



* Progressive Growing of GANs for Improved Quality, Stability, and Variation (2018)