Expected Value? 그게 뭔~ 가요? (기댓값??)

소프트웨어 꼰대 강의

노기섭 교수

(kafa46@cju.ac.kr)

이게 도대체 뭐래요? 헐!!!

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathbb{E}_{x \sim P(x)} \left[\mathbb{E}_{y \sim P(y|x)} [\ln P(y|x; \theta)] \right]$$

머신러닝(특히 딥러닝)에서 많이 보게 됩니다.

웬... 무슨... 이런 만행을!!
외계어냐?

$\mathbb{E}$ 는 기댓값을 의미합니다.

오늘 강의를 들으면 무슨 뜻인지 감잡을 수 있어요^^

우리가 사는 세상.... 어떤 현상을 표현하려면?

내가 아는 어떤 인물

⬇

이 사진을
표현할 수 있는 방법?

⤷
- 이름
- 주민번호
- 군번
- 사원번호
- ⋮

자료출처: https://hangyo.com/mobile/article.html?no=85073

초등학교 교실의 학생들

⬇

이 사진을
표현할 수 있는 방법?

⬇

"학년 - 반" 활용?

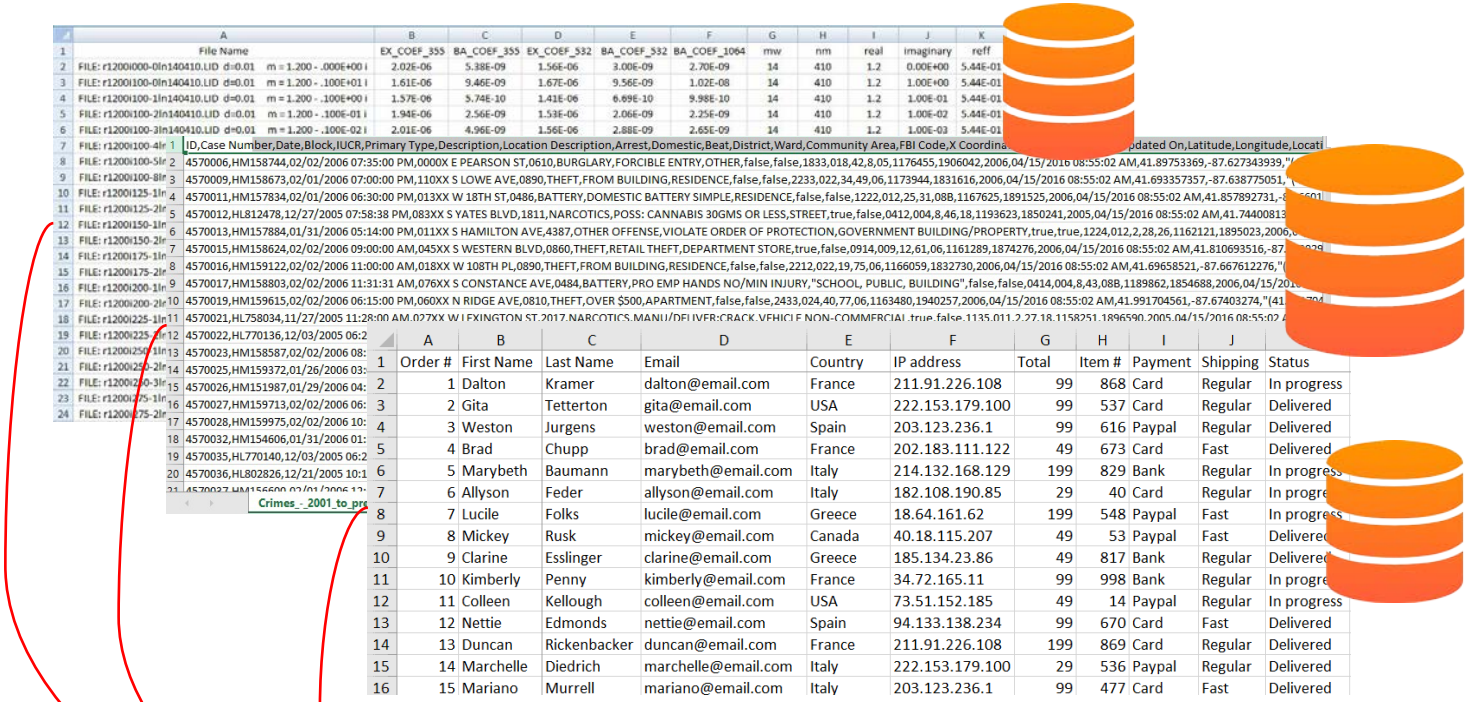
5학년 3반

5-3

0503

⋮

데이터를 표현하는 방법은?



Dataset...

데이터를
표현할 수
있는 방법?

- 갯수
- 최대/최솟값
- 평균
- 표준편차
- ⋮

이런 질문을 자주 받습니다. ^^

- **사실은... 저도 늘 헷갈렸던 부분입니다...**
  - 음... 뭐냐면....

기댓값을 이해하려면
확률변수 개념을 먼저 이해하셔야 해요^^

평균과 기댓값은 다릅니다.
하지만 같은 값을 가질수도 있습니다.

결론부터 이야기 하면,

이미지 출처:
https://pngtree.com/freepng/education--thinking--question-mark_5923515.html

교수님....

평균과 기댓값은

같은 건가요?

ㅠㅠ

이미지 출처:
https://www.cleanpng.com/png-professor-teacher-clip-art-mad-professor-5552425/

평균 vs. 기댓값?

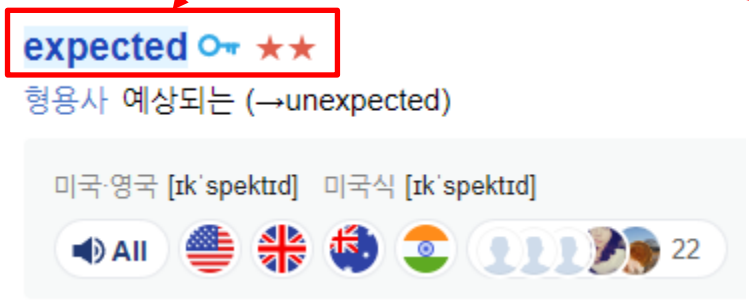
평균 (Mean)

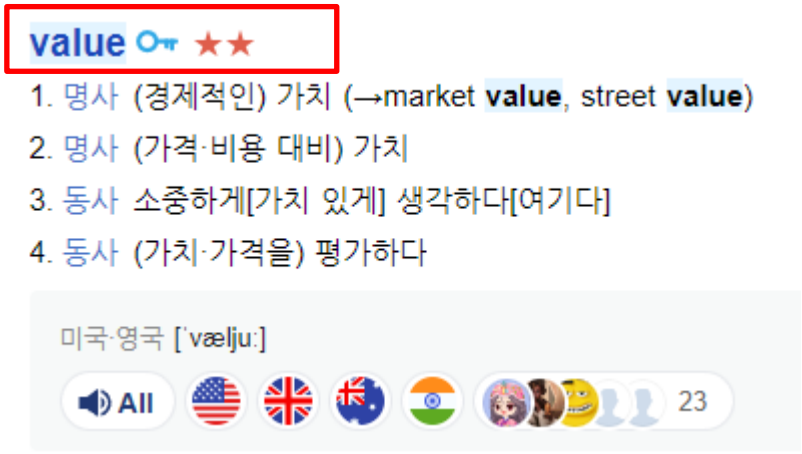
- 어떤 집합의 모든 원소 값을 더한 후 원소의 갯수로 나눈 값

$$Mean\ (\mu) = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Note:
평균의 종류는 많습니다^^
여기서 '평균'은 산술평균을 의미합니다.

기댓값 (Expected Value)





Expected value: 기대하는 값

정확히 값을 안다면 '기대하는' 표현 X

'기대하는' → 확률적으로 계산해야 할 것임

기댓값 표기

Definition of expected value (https://en.wikipedia.org/wiki/Expected_value)

- In probability theory, the expected value (also called expectation, expectancy, expectation operator, mathematical expectation, mean, average, or first moment) is a generalization of the weighted average.
- Informally, the expected value is the arithmetic mean of a large number of independently selected outcomes of **a random variable**.

요거를 수학 기호로 표시하면, 이렇게 됩니다.

$$E[X] = x_1\ p_1 + x_2\ p_2 + \cdots + x_n\ p_n = \sum_{i=1}^{n} x_i p_i$$

- X: $a\ random\ variable$
- x_i: $possible\ outcome\ of\ X$
- p_i: $probability\ of\ x_i$

기댓값은 확률 변수로부터!
(확률변수는 확률과 값을 갖는다)

굳이 기댓값을 사용하는 이유

- **뭐하러 복잡하게 기댓값을 만들어서 쓰나요?**
  - 그냥 평균, 중앙값 이런 거 쓰면 되잖아요?
- **물론 그래도 되죠 ㅎㅎ**
- **어떤 일을 결정할 때 (확률적으로) 예상되는 값을 알게 된다면…**
  - 더 좋지 않을까요?

(예) 도박을 하는 경우

- 이길 확률: 60% → 이기면 1만원 받기
- 질 확률 : 100% - 60% = 40% → 지면 2만원 주기
- 현재까지 데이터

게임 데이터	결과	수익
Game1	Win	+1만원
Game2	Win	+1만원
Game3	Lose	-2만원
Game4	Win	+1만원

평균수익

$$Mean\ (\mu) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{1+1-2+1}{4} = \frac{1}{4} = +0.25 \text{ 만원}$$

평균 수익이 플러스네… ㅋ
계속 해서 부자 되야지^^

기댓값

$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i) = 0.6 \times 1 + 0.4 \times (-2)$$

$$= 0.6 - 0.8 = -0.2 \text{ 만원}$$

이런 게임 해야 할까요?

평균을 기댓값으로 해석한다면?

■ 수학 성적

$$E[X] = \sum_{i=1}^{n} x_i p(x_i) = 100 \cdot \frac{1}{6} + 90 \cdot \frac{1}{6} + 80 \cdot \frac{1}{6} + 70 \cdot \frac{1}{6} + 60 \cdot \frac{1}{6} + 50 \cdot \frac{1}{6}$$

$$= \frac{100 + 90 + 80 + 70 + 60 + 50}{6} = \frac{450}{6} = 75$$

X

Ω	E	확률
100	100	1/6
90	90	1/6
80	80	1/6
70	70	1/6
60	60	1/6

가능한 점수

가능한 점수가 나올 확률을

동일하게 지정하면,

기댓값과 평균값은 같다.

평균은 기댓값의 Special Case

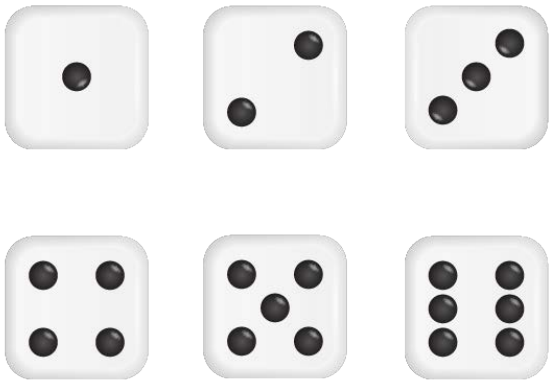
(동일한 가중치!)

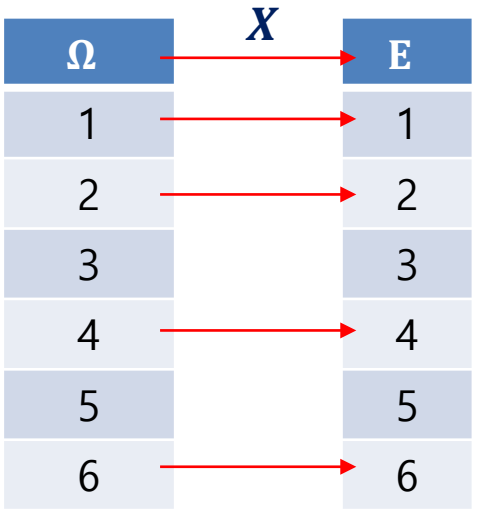
$$Mean\ (\mu) = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{100 + 90 + 80 + 70 + 60 + 50}{6} = \frac{450}{6} = 75$$

기댓값 구하는 간단한 예제

■ **실험: 주사위 1개를 던지는 실험**

- 먼저, Random variable X (맵핑 함수) 정하기
- X: 주사위 던졌을 때 나온 값이라고 하자...

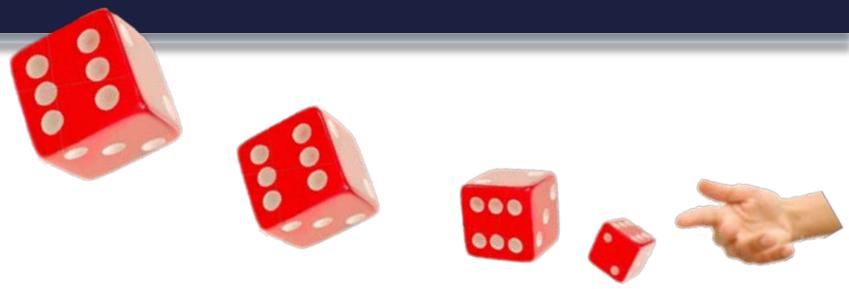




Random variable X의 기댓값

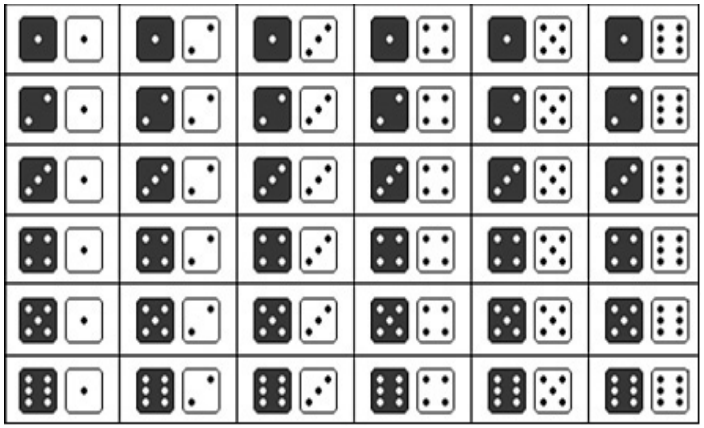
$$E[X] = \sum_{i=1}^{n} x_i p_i = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

기댓값과 평균 예시 - 조금만 더 복잡한 걸루...(1/2)



주사위를 던지는 실험

- 먼저, Random variable X (맵핑 함수) 정하기
- X: 주사위 2개의 합이라고 하자...



Ω	X	E
1, 1	→	2
1, 2	→	3
⋮		⋮
2, 1	→	3
⋮		⋮
6, 6	→	12

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- 주사위 2개를 던졌을 때, 합이 7이 될 확률은? $P(X=7)=\frac{6}{36}=\frac{1}{6}$
- 주사위 2개를 던졌을 때, 합이 9가 될 확률은? $P(X=9)=\frac{4}{36}=\frac{1}{9}$

평균과 비교하면?
(다음 슬라이드)

- 기댓값은?

$$E[X]=\sum_{i=1}^{n} x_i p_i = 2\cdot\frac{1}{36}+3\cdot\frac{2}{36}+4\cdot\frac{3}{36}+\cdots+12\cdot\frac{1}{36}=7$$

기댓값과 평균 예시 - 조금만 더 복잡한걸루... (2/2)

- **주사위 2개를 던진 실험을 10회 dataset 확보했다.**

  - Dataset의 평균은?

- **Dataset #1**

평균과 기댓값이 같은 경우

실험	1	2	3	4	5	6	7	8	9	10
결과값	3, 4	1, 5	5, 1	3, 3	1, 2	6, 3	6, 6	1, 6	2, 4	4, 4
RV X	7	6	6	6	3	9	12	7	6	8

$$\boldsymbol{Mean\ (\mu) = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{7+6+6+6+3+9+12+7+6+8}{10} = \frac{70}{10} = 7}$$

- **Dataset #2**

평균과 기댓값이 다른 경우

실험	1	2	3	4	5	6	7	8	9	10
결과값	3, 4	3, 6	6, 6	5, 6	1, 2	2, 6	5, 1	1, 6	2, 4	4, 4
RV X	7	9	12	11	7	8	6	7	6	8

$$\boldsymbol{Mean\ (\mu) = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{7+9+12+11+7+8+6+7+6+8}{10} = \frac{81}{10} = 8.1}$$

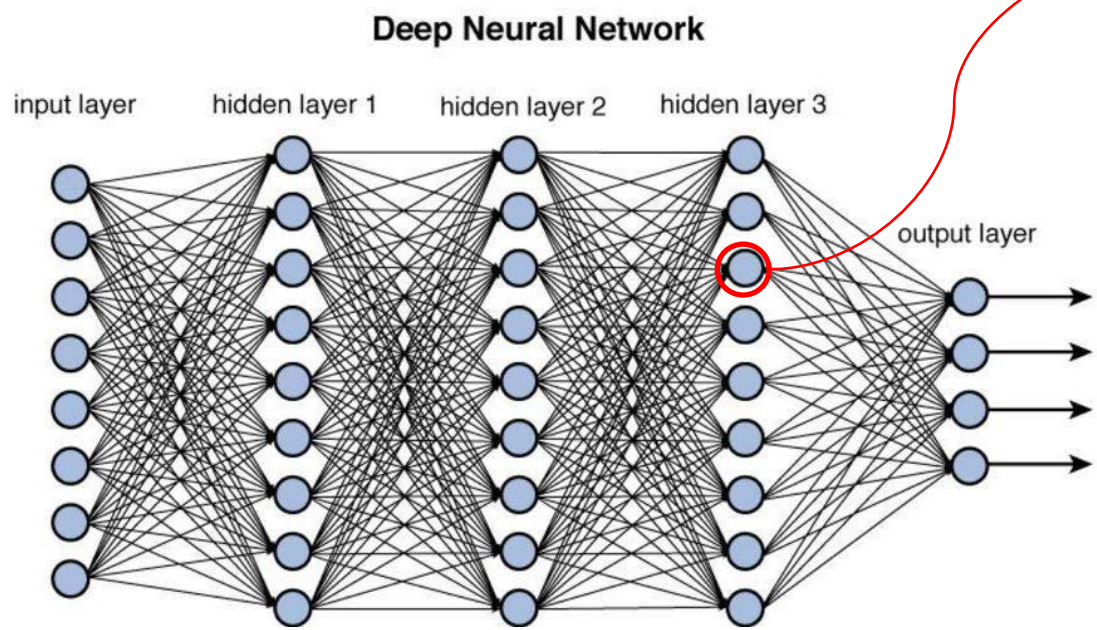
기댓값의 재밌는 성질

X and Y are random variables.

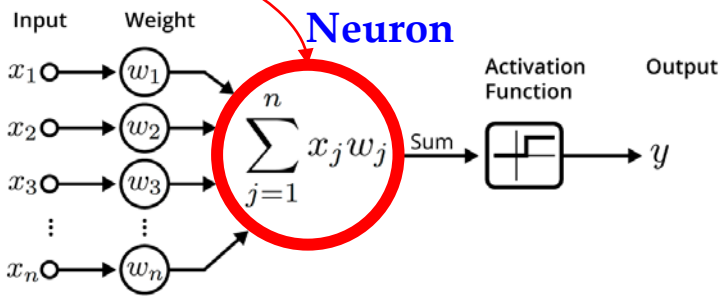
- $E[X+Y] = E[X] + E[Y]$
- $E[XY] = E[X] \cdot E[Y]$
- $E[aX+b] = aE[X] + b$
- $E\left[aX^2 + bX + c\right] = aE\left[X^2\right] + bE[X] + c$, where $a, b, c \in \mathbb{R}$
- $E[aX+bY] = aE[X] + E[Y]$, where $a, b, c \in \mathbb{R}$

딥러닝 & 기댓값

- **딥러닝에도 기댓값 개념이 있나요?**
  - 아직 먼 이야기인 것은 맞습니다. 파라미터, 신경망 개념을 알아야 합니다. ㅠ
  - 하지만 기댓값이 어떻게 쓰이는지 맛보기는 할 수 있습니다.



이미지 출처: https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964



An illustration of an artificial neuron. Source: Becoming Human.

이미지 출처: https://www.freecodecamp.org/news/deep-learning-neural-networks-explained-in-plain-english/

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathbb{E}_{x \sim P(x)} \left[\mathbb{E}_{y \sim P(y|x)} [\ln P(y|x;\theta)] \right]$$

수고하셨습니다 ..^^..