# 확률분포와 딥러닝

(Probability Distribution & Deep Learning)

소프트웨어 꼰대 강의

노기섭 교수 (kafa46@cju.ac.kr)

# 확률변수(Random Variable) 복습

#### ■ Random variable 복습

# 맵핑 함수

 $X: \Omega \to E$ 

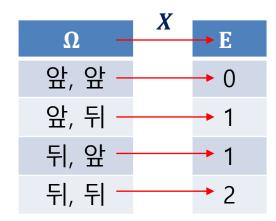
Ω: sample space

E: measurable space

#### 동전 던지기



X: 동전 2개를 동시에 던졌을 때 뒷면 개수



### 용어 정리

#### Definition (online wikipeida: <u>확률분포</u>)

- 확률 변수가 특정한 값을 가질 확률을 나타내는 함수를 의미한다. Measurable Space Random variable -3-2-10 1 2 Sample Space e≈2.72 다양한 확률 분포 (입출력 패턴으로 이해) Uniform Distribution Normal Distribution Standard Normal Distribution uniform.rvs(size=10000.loc=10.scale=20) norm.rvs(size=10000,loc=10,scale=100) Continuous 300 200 150 100 400 200 200 100 50 Total number of 10 15 20 25 30 -200 0 possible random outcomes **Probability** Gamma Distribution **Exponential Distribution** Beta Distribution Continuous gamma.rvs(a=5,size=10000) expon.rvs(scale=1,loc=0,size=10000) **Function** 400 800 600 400 200 300 200 200 100 • 입력: 확률변수 값 0.2 0.4 0.6 0.8 10 15 • 출력: 확률 Binomial Distribution Poisson Distribution Bernoulli Normal Distribution bemoulli.rvs(size=10000,p=0.6) poisson.rvs(mu=3.size=10000) • 표현:  $p = f_X(x)$ Discrete 3000 6000 2000 1500 1000 2000 4000 2000 1000 500 10 0.5 1 1.5 #30DayChartChallenge - statistics - 2021/04/09 Datasets created using scipy.stats twitter.com/vivekparasharr | github.com/vivekparasharr | vivekparasharr.medium.com

# **Types of Probability Distribution**

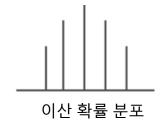
#### ■ 퉁 쳐서 한번에 표기할 때

$$p = f_X(x)$$
 또는  $p = f(x)$ 

여기서 x는 Random variable (확률 변수의 값)

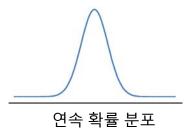
#### ■ 확률분포 종류 $\rightarrow$ Random variable 값(x)에 따라 구분

- 이산형 (Discrete) 값 입력
  - · Probability Mass Function
  - '확률 질량 함수'라고 부르기도 함  $p(x) = f_X(X = x)$



- 연속형 (Continuous) 값 입력
  - · Probability Density Function
  - ・ '확률 밀도 함수'라고 부르기도 함

$$p(a \le x \le b) = \int_a^b f_X(x) dx$$



# 확률 분포 맛보기

### ■ 베르누이 분포

- Random variable
  - 1  $\rightarrow$  success
  - $0 \rightarrow \text{fail}$

$$p = P[success] = P[X = 1]$$

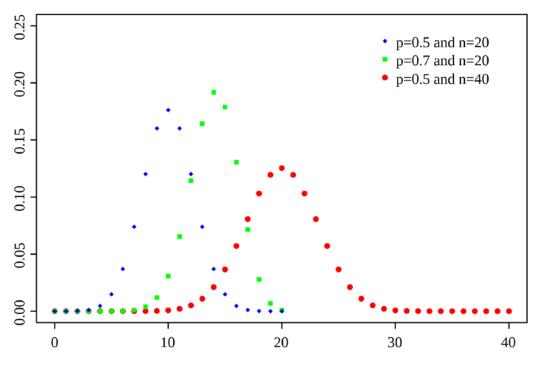
#### - 확률 함수

$$f_X(x;p) = p^x(1-p)^{1-x}, \qquad x = 0 \text{ or } 1$$

- · 매개변수 p가 주어졌을 때 확률변수 X 값이 x일 확률
- 기댓값 (expected value) E[X] = p
- 분산 (variance) V[X] = p(1-p)

# 이항 분포 (Binomial Distribution)

#### 이항 분포 (Binomial Distribution)



참고: 베르누이 분포는 파라미터 n = 1로 고정한 special case라고 생각하면 됨

매개변수 n과 p가 주어졌을 때 n번 시행 중에 k번 성공할 확률

$$X \sim B(n, p)$$

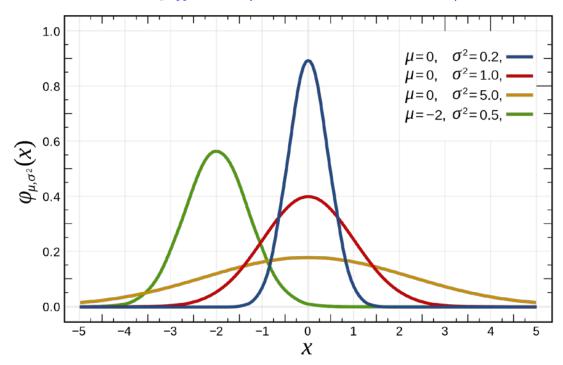
$$f_X(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$
, where  $\binom{n}{k} = \frac{n!}{k! (n-k)!}$ 

기댓값 (expected value) E[X] = np

분산 (variance) 
$$V[X] = np(1-p)$$

## 정규 분포 (Normal Distribution

#### 정규 분포 (Normal Distribution)



참고: 정규 분포를 '가우스 분포', '벨 커브'라고 부름 평균이 0이고 분산이 1인 정규분포를 '표준 정규 분포'라고 합니다.

매개변수  $\mu$ 와  $\sigma$ 가 주어졌을 때 확률변수 X의 값이 x일 확률

$$X \sim N(\mu, \sigma^2)$$

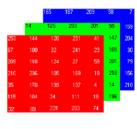
$$f_X(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

기댓값 (expected value)  $E[X] = \mu$ 

분산 (variance) 
$$V[X] = \sigma^2$$

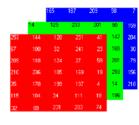
### 데이터셋과 확률 분포?





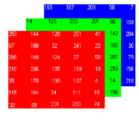
$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \end{bmatrix}$$





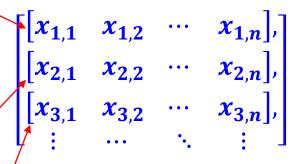
$$\begin{bmatrix} x_{2,1} & x_{2,2} & \cdots & x_{2,n} \end{bmatrix}$$





$$\begin{bmatrix} x_{3,1} & x_{3,2} & \cdots & x_{3,n} \end{bmatrix}$$

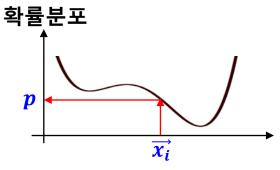
### 확률 분포가 어디에 있다는 거죠?



확률변수: 이미지의 픽셀 값

확률함수:  $f_X(X = \overrightarrow{x_i}, L = l_i; \theta)$ 

(L: label)



### 딥러닝과 확률 분포

- 딥러닝과 확률 분포라는게....
  - 도대체 딥러닝과 확률 분포가 어떤 관계가 있는 건가요?
- 데이터가 있을 때,
  - 그 데이터를 정확히 표현하는 확률 분포가 존재
  - 그러나 신(神)이 아닌 이상 우리는 알 수 없음.



이미지 출처: https://www.cnet.co.kr/view/?no=20220916111031

요런 분포일 경우 정답은 확률적으로 OOO 이다. [ $x_{1,1} x_{1,2} \cdots x_{1,n}$ ]

- 데이터를 잘 학습하면 정답을 모사(imitate, 흉내내는)하는 확률 분포를 찾을 수 있을 것임.
- 딥러닝을 활용하여 그 확률 분포를 찾자!



수고하셨습니다 ..^^..