Sampling Representation & Monte Carlo Approximation in deep learning

소프트웨어 꼰대 강의

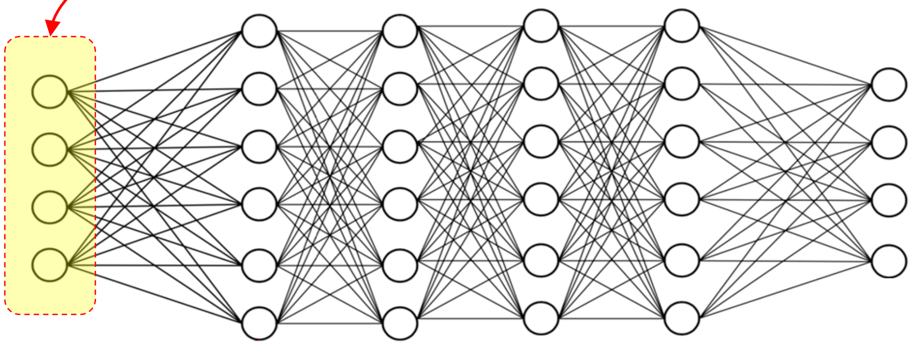
노기섭 교수

(kafa46@cju.ac.kr)

딥러닝 공부하다 보면 보게 되는 것들……

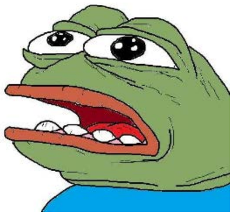
입력 데이터

X?? ***x??*** $\boldsymbol{x_i}$***??***



일반적인 딥러닝 구조

$$\mathbb{E}_{X\sim P(X)}[f(x)] = \sum_{x\in\chi} P(X=x)\cdot x$$



이건 또 뭐냐?

이번 강의를 들으면 확률 분포에서 샘플링 하는

표현법과 의미를 이해할 수 있게 됩니다^^

확률 분포로부터의 샘플링

Prerequisites

- Random variable (확률 변수) → https://youtu.be/iTxTGBOhzCA
- Probability function (확률 함수)
- Probability distribution (확률 분포)

https://youtu.be/qpbIKg21mvI

샘플링(or dataset)에서 자주 사용하는 표현법

$X \sim P$ 또는 $X \sim P(X = x)$

Random variable X는 확률 분포(확률 함수) $P(x)$와 같이 분포되어 있다.

Random variable X는 확률 분포 $P(x)$에서 샘플링 하였다.

- 아래 수식을 보게 되었다면, 어떻게 해석할까요?

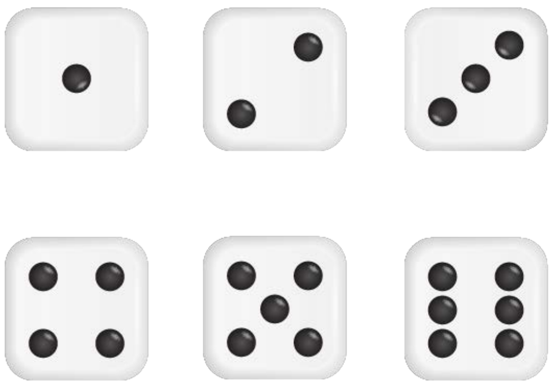
$$\mathbb{E}_{X \sim P(X)}[f(x)] = \sum_{x \in \chi} P(X = x) \cdot x$$

다음 슬라이드에서
예제를 통해 살펴보겠습니다 ^^

Sample Space (Ω)가 명확한 경우(복습)

주사위 던지는 실험

- Random variable ***X***: 주사위를 던졌을 때 나온 값



$\Omega \xrightarrow{X} E$

Ω	E
1	1
2	2
3	3
4	4
5	5
6	6

하지만 정확한
Sample space 모를 경우는?

딥러닝에 사용할 dataset일 경우와 동일…
몬테 카를로 근사부터 다음 슬라이드에서
예제를 통해 살펴보겠습니다 ^^

$$\mathbb{E}_{X\sim P(X)}[f(x)] = \sum_{x\in\chi} P(X=x)\cdot f(x)$$

$$= \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = \frac{21}{6} = 3.5$$

여기서 $f(x) = x$

Monte Carlo Approximation

- 1940년대에 '존 폰 노이만'과 '스타니슬라프 울람'에 의해 발명
- 룰렛 게임과 동일한 무작위 특성을 공유한다는 점 때문에 모나코의 유명한 도박장의 이름을 따서 명명

자료출처: https://aws.amazon.com/ko/what-is/monte-carlo-simulation/

Goal

Approximate $\boxed{\mathbb{E}_{X\sim P(X)}[f(x)]}$

Intractable to compute exactly in efficient way!!

Definition

If $x_i, x_2, \cdots, x_n \sim P(x)$ i.i.d. *then*

$$\boxed{\hat{\mu}_n = \frac{1}{n}\sum_{i}^{n} f(x_i)} \quad \textit{is a basic } \text{Monte Carlo } \textit{estimator}$$

$$\textit{of } \mathbb{E}_{X\sim P(X)}[f(x)]$$

It is just a sample mean

Remarks

Unbiased estimator(샘플의 평균은 전체 평균과 같다.)

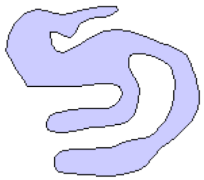
$$\boxed{\mathbb{E}[\hat{\mu}_n] = \mathbb{E}[f(x)]}$$

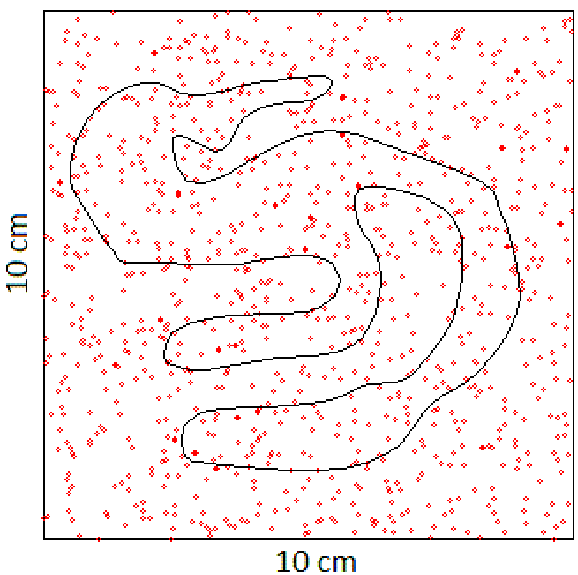
Consistent estimator(샘플이 많으면 오차가 적어진다.)

$$\boxed{\hat{\mu}_n \xrightarrow{P(x)} \mathbb{E}[f(x)] \quad as \quad n \to \infty \text{ (for all } \epsilon > 0,\ P(|\hat{\mu}_n - \mathbb{E}[f(x)]| < \epsilon \to 1)}$$

Example: Monte Carlo Approximation

- **다음 도형의 넓이를 구해 주세용 ^^.**
  - 조건
    - 데이터는 많이 구할 수 있다.
    - 도형을 생성하는 함수 $f(x)$ 알 수 있다.
    - 정확한 넓이는 존재할 것이다. 그러나 구하는 공식은 정확히 모른다.
  - 해결책





- Random variable X 는 확률 분포(함수) $P(X)$를 따른다고 가정
- Random 데이터 n 개를 모은다.

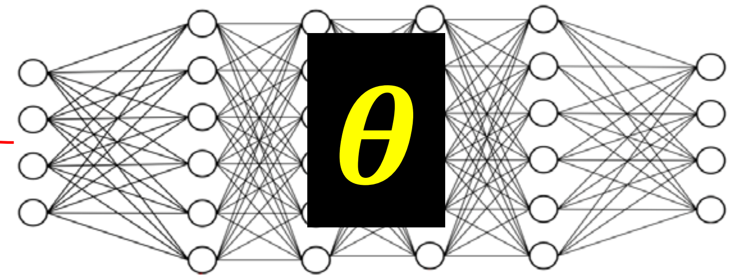
$$D = \{x_i \mid i = 1, 2, \cdots, n\}, \quad x_i \sim P(X)$$

- $f(x)$를 이용하여 n 개 데이터 중에서 도형 내부에 있는 개수를 센다.
- 도형의 면적 $= \dfrac{\text{도형 내부에 있는 데이터의 개수}}{\text{전체 데이터의 개수}} \times$ 직사각형 넓이

$$\mathbb{E}_{X \sim P(X)}[f(X)] \approx \frac{1}{n} \sum_{i=1}^{n} f(x_i), where\ x_i \sim P(X)$$

실제 딥러닝 데이터는 어떻게 처리할까요?

parameter $\boldsymbol{\theta}$



$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \mathbb{E}_{X \sim P(x)}[\log P(y|x;\theta)]$$

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$$

	carat	cut	color	clarity	depth	table	price
1	0.52	Ideal	D	VS2	61.4	56	1664
2	0.5	Very Good	F	SI1	62.3	60	1250
3	0.61	Ideal	G	VVS2	61.6	54	2242
4	0.36	Premium	G	VS2	62.5	58	756
5	0.7	Very Good	E	VS2	63.5	54	2889
6	0.56	Ideal	F	VS1	61.7	56	2016
7	1.19	Premium	E	I1	60.2	61	3572
8	0.52	Ideal	F	IF	60.6	57	2575
9	0.38	Ideal	E	IF	62.7	55	1433
10	0.51	Ideal	E	VS2	62.1	54	1608

Monte Carlo Approx.

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \log P(y_i|x_i;\theta)$$

$$\hat{\theta} = -\underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \log P(y_i|x_i;\theta)$$

References

참고 자료

- Monte Carlo Approximation (몬테 카를로 근사)
  - 이론: https://www.youtube.com/watch?v=7TybpwBlcMk&t=8s
  - 실습: https://youtu.be/x3b_DAEmZnU?si=M074vK5Tz_XAWuSf
- Standard notation for [sampling from] a conditional probability (조건부 확률에서의 샘플링 표기)
  - https://stats.stackexchange.com/questions/179182/standard-notation-for-sampling-from-a-conditional-probability

수고하셨습니다 ..^^..