Stochastic Process (확률 과정)

소프트웨어 꼰대 강의

노기섭 교수

(kafa46@cju.ac.kr)

Contents

Stochastic Approach

Stochastic??

In Machine Learning (https://machinelearningmastery.com/stochastic-in-machine-learning/)

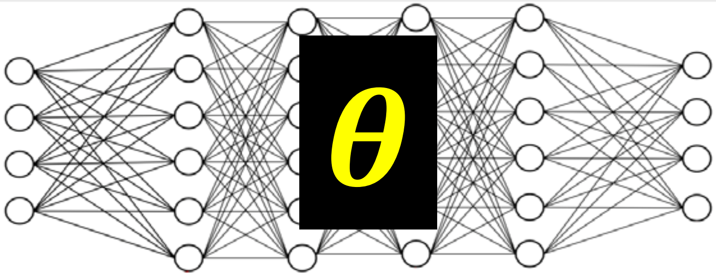
- The outcome involves some randomness and has some uncertainty.
  - Stochastic is a synonym for random and probabilistic.
  - ML algorithms are stochastic because they explicitly use randomness during optimization.

Stochastic Approach

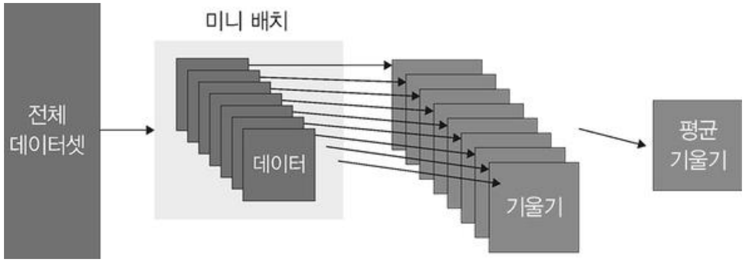
- Outcomes are by-products of chance and randomness
- Outcomes are that causation is probabilistic rather than inevitable or predetermined.
  - https://www.insidehighered.com/opinion/blogs/higher-ed-gamma/2023/04/21/value-stochastic-thinking

딥러닝의 학습 방법

parameter $\boldsymbol{\theta}$



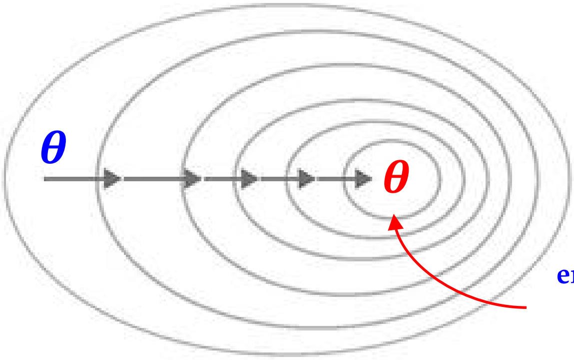
$$\hat{\boldsymbol{\theta}} = -\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\textbf{argmax}}\, \mathbb{E}[\log P(error;\boldsymbol{\theta})]$$

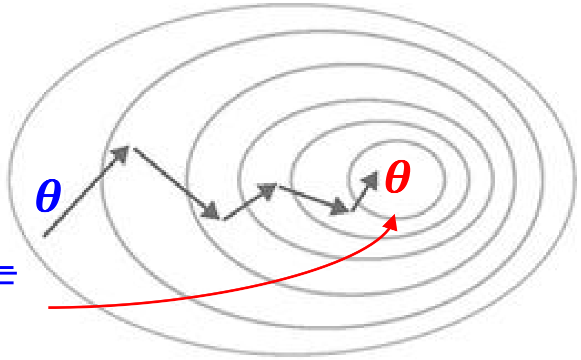


이번 강의를 들으면 Stochastic의 의미를 이해할 수 있게 됩니다.^^

Batch Gradient Descent (BGD)
전체 dataset에 대한 오류를 한번에 계산

Stochastic Gradient Descent (SGD)
미니배치 dataset에 대한 오류를 각각 계산





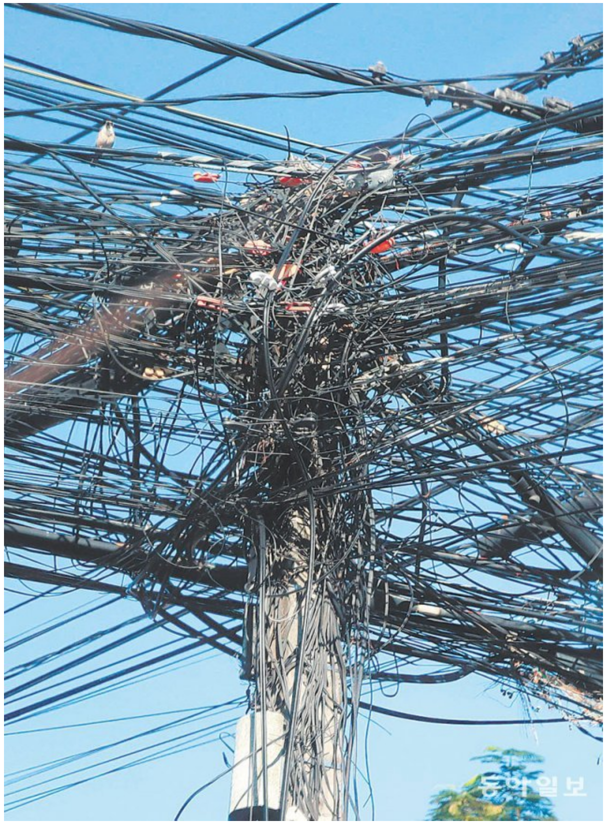
error 최소화 하는 파라미터 집합 (정답)

복습: 확률의 곱셈?

- **어떤 일이 일어날 확률 p**
- **어떤 일이 일어나지 않을 확률 $1-p$**
- **두 개의 확률을 곱한다는 의미**
  - 두 사건이 독립(independent)
  - 두 사건의 확률이 동시에 발생할 확률로 정의
  - (Example)
    - 사건 A가 발생할 확률: $p(A) = 0.5$
    - 사건 B가 발생할 확률: $p(B) = 0.4$
    - 사건 A와 B가 동시에 발생할 확률: $p(A\&B) = 0.5 \times 0.4 = 0.2$
  - 특징: 동시에 발생할 확률은 각각의 확률값 보다 작아진다.
- **가정사항**
  - 완전히 독립이라는 가정
  - 위험성: 완전히 독립이 아닌 경우도 독립이라고 생각하는 경우

그림으로 살펴보는 Stochastic

이해하기 복잡한 세상



이미지 출처: https://www.donga.com/news/article/all/20221013/115946826/1

미래에 일어날 일을 한 가지로 예측할 수는 없음

동시에 던지는 실험: 경우의 수는 4가지

실험의 결과 → 던져봐야 알게 됨

Stochastic Process
An ongoing process where the next state might depend on both previous states and some random element!

Stochastic 실습 및 분석

Stochastic vs. Non-stochastic

- Randomness 적용 → Stochastic implementation

```
def rolling_die_2():
    '''return 1에서 6사이 정수 값 중
       랜덤하게 선택된 숫자
    '''
```

- Under determined

```
def rolling_die_1():
    '''return 1에서 6사이 하나의 정수값'''
```

Stochastic - 주사위 눈이 나올 경우의 수

주사위 던지기 실행 with Python

```
import random

def rolling_die_2():
    '''return 1에서 6사이 정수 값 중
       랜던하게 선택된 숫자
    '''
    return random.choice([1, 2, 3, 4, 5, 6])


def run_rolling(test_num=5):
    '''주사위 던지기 실행'''
    result = []
    for i in range(test_num):
        result.append(rolling_die_2())
    print(result)

run_rolling()
```

>>> [5, 1, 6, 1, 4] # 실행 결과 예시

다음과 같을 확률은?

[1, 1, 1, 1, 1] ?

[1, 1, 1, 1, 2] ?

[1, 1, 1, 1, 3] ?

⋮

[6, 6, 6, 6, 6] ?

$$\left(\frac{1}{6}\right)^5 \cong 0.00012$$

Stochastic - 주사위 눈이 나올 확률 계산

■ 주사위 던지기 확률 측정(estimator)

다음과 같을 확률은?

[1, 1, 1, 1, 1] ?
[1, 1, 1, 1, 2] ?
⋮
[6, 6, 6, 6, 6] ?

$$\left(\frac{1}{6}\right)^5 \cong 0.00012$$

```python
import random

def rolling_die_2():
    return random.choice([1, 2, 3, 4, 5, 6])

def die_prob_estimator(target: str, trials:int):
    hit_count = 0
    for _ in range(trials):
        result = ''
        for _ in range(len(target)):
            result += str(rolling_die_2())
        if target == result:
            hit_count += 1
    prob_estimation = round(float(hit_count)/float(trials), 6)
    print(f'Estimated Probability of {target}: {prob_estimation}')

die_prob_estimator('11111', 1000000)
```

- 다음 값을 예측할 수 있는가?
  ➔ 동일한 랜덤 넘버를 생성한다면 가능하다.
- 왜 이론값(정답)과 예측값은 틀린가?
  ➔ 실험을 많이 하면 오차가 줄어든다

```
>>> Estimated Probability of 11111: 0.000133
>>> Estimated Probability of 11111: 0.000119
    :          :         :        :
```

만약 1000번 미만으로 실행하면, 거의 0 나올 것임

Stochastic #2. 생일이 같은 사람이 있을 확률?

- 각각의 생일이 나올 확률이 동일(equally likely) 하다고 가정
- N 명 중에서 최소 2명이 생일이 같을 확률

$$1 - \frac{365!}{365^N \times (365 - N)!}$$

```python
def same_birthday_prob(
    num_people: int,
    num_same: int
    ) -> bool:
    possible_date = range(365)
    birthdays = [0] * 365
    for _ in range(num_people):
        birth_date = random.choice(possible_date)
        birthdays[birth_date] += 1
    return max(birthdays) >= num_same # boolean
```

```python
def birthday_estimator(
    num_people: int,
    num_same: int,
    trials: int) -> float:
    hit_count = 0
    for _ in range(trials):
        if same_birthday_prob(num_people, num_same):
            hit_count += 1
    return hit_count/trials
```

```python
def run_birthday_simulation(
    num_same: int,
    peoples: list) -> None:
    '''같은 생일이 있는지 시뮬레이션 수행'''
    for people in peoples:
        print('---'*10)
        estimated_probability = round(
            birthday_estimator(people, num_same, 1000),
            3
        )
        numer = math.factorial(365)
        denomi = (365**people)*math.factorial(365-people)
        theoretical_prob = round(1 - numer/denomi, 3)
        print(f'{people}명 중 같은 생일이 있을 확률: \n\
시뮬레이션: {estimated_probability}\t\
이론값: {theoretical_prob}')

if __name__=='__main__':
    run_birthday_simulation(2, [10, 20, 40, 100])
    # run_birthday_simulation(3, [10, 20, 40, 100])
    # run_birthday_simulation(4, [10, 20, 40, 100])
```

생일 문제의 재해석

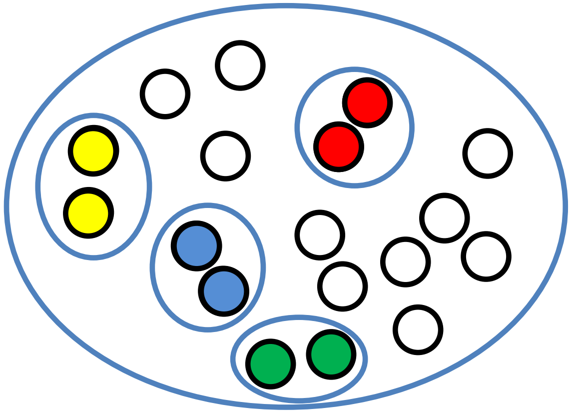
- **최소 2명 생일이 같을 확률 = 1 - 모든 사람의 생일이 다를 확률**
- **최소 3명 생일이 같을 확률??**

1 - (모든 사람 생일이 다를 경우 또는,

- 1쌍 같은 생일을 공유하고
  나머지가 모두 다를 경우 또는,
- 2쌍 생일이 각각 다른 생일을 공유하고
  나머지가 모두 다를 경우 또는,

⋮

- N/2쌍 생일이 각각 다른 생일을 공유하고, 나머지가 모두 다를 경우 또는)



- 수식으로 풀려면 매우 복잡, 때로는 시뮬레이션이 더 편리함 → 확률적(Stochastic) 실행을 통해 결과 도출

- 참고 자료
  - https://youtu.be/-1BnXEwHUok?si=5g9ki8y-Haj4SMWf
  - https://math.stackexchange.com/questions/25876/probability-of-3-people-in-a-room-of-30-having-the-same-birthday

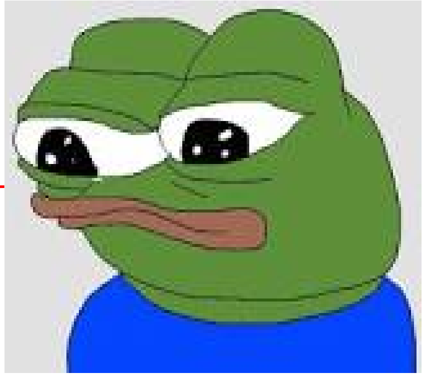
독립 가정을 오류와 Stochastic 접근의 필요

- **프로야구 경기: '이글스', '베어스' → 2개 팀**
- **토요일 경기에 대한 현재까지 승률**
  - 이글스: $\frac{7}{8}$
  - 베어스: $\frac{6}{8}$
- **다음 토요일 경기에 두 팀이 모두 이길 확률**

$$\frac{7}{8}\times\frac{6}{8}=\frac{42}{64}=65.625\%$$

- **다음 토요일 경기에 두 팀이 모두 질 확률**

$$1-65.625\%=34.375\%$$



사건이 많고, 횟수가 증가할수록
더욱 해석하기 어려워진다.

연필 들고 수식
푸는 것보다
좋은 방법은 없나?

- **다음 토요일에 이글스와 베어스가 경기한다면?**
  - 두 팀 중 하나의 팀이 질 확률은: 상대팀이 이길 확률에 영향을 받게 됨 ㅠㅠ
  - 어느 팀이 질 확률은 34.375% 보다는 크게 될 것임.

Stochastic Process

랜덤 과정??

Definition

- **From wiki (https://en.wikipedia.org/wiki/Stochastic_process)**
  - a mathematical object usually defined as a sequence of random variables, where the index of the sequence has the interpretation of time.
  - Stochastic processes are widely used as mathematical models of systems and phenomena that appear to vary in a random manner.
  - 쉽게 말하면, 시간의 흐름에 따라 발생하는 확률적 변화 과정

- **Notation**

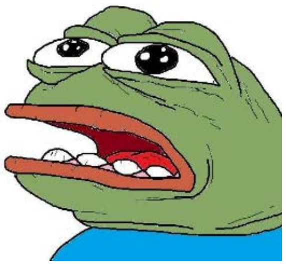
$Random\ Process$: $X(\zeta_i, t_i)$

t : time index

ζ_i: Outcome of random experiment

$X(\zeta)$: Random variable

X : stationary function

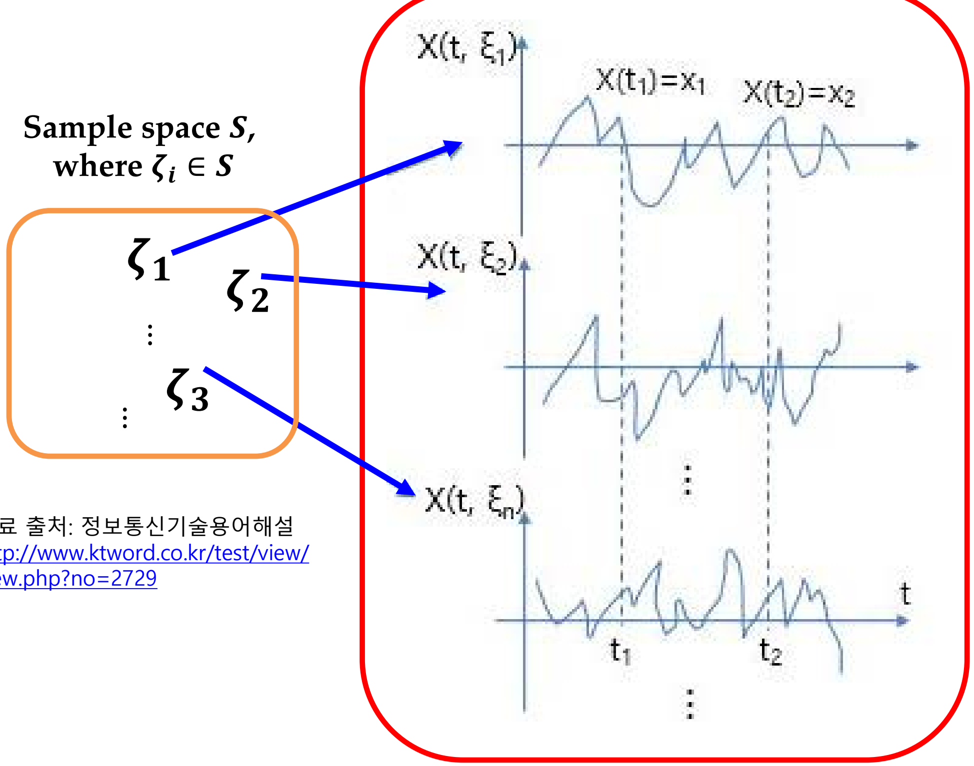


정의가 더 어렵네 ㅠㅠ

Stochastic (Random) Process Ensemble

■ Ensemble (앙상블)

Random process로부터 나타나는 시행 결과들의 모음(집합)



$\{\, X(t, \zeta_i) \mid i = 1, 2, \cdots, \infty \}$

자료 출처: 정보통신기술용어해설
http://www.ktword.co.kr/test/view/view.php?no=2729

High-level Understanding

Stochastic Process

- 일단 함수라고 생각한다 → 랜덤 실험의 결과를 시간에 따른 값으로 맵핑 해주는 함수
- 맵핑 함수라는 점에서 Random variable 과 동일
- 하지만 시간 개념이 추가됨 → 시간 t, 그리고 실험 결과를 표시할 ζ (zeta, '제타'라고 읽음)
  - ζ는 random variable

전압계로 살펴보는 Stochastic Process

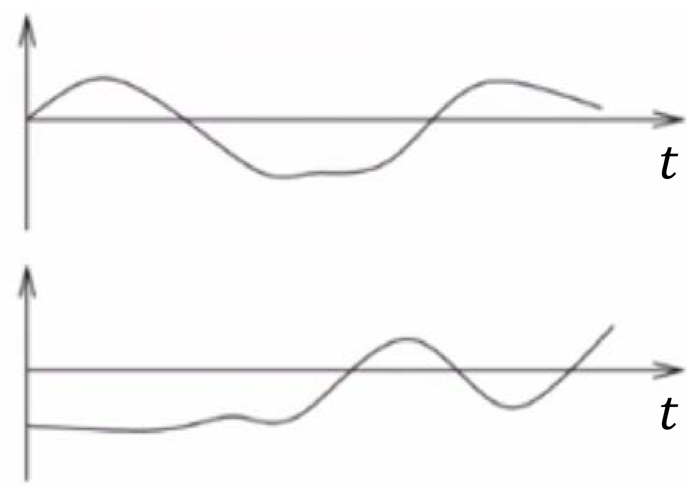
전압계 출력 신호:
원래 전압 + Noise가
혼합되었다 가정

실험 1: $X(\zeta_1) = 0.3$

실험 2: $X(\zeta_2) = -0.2$

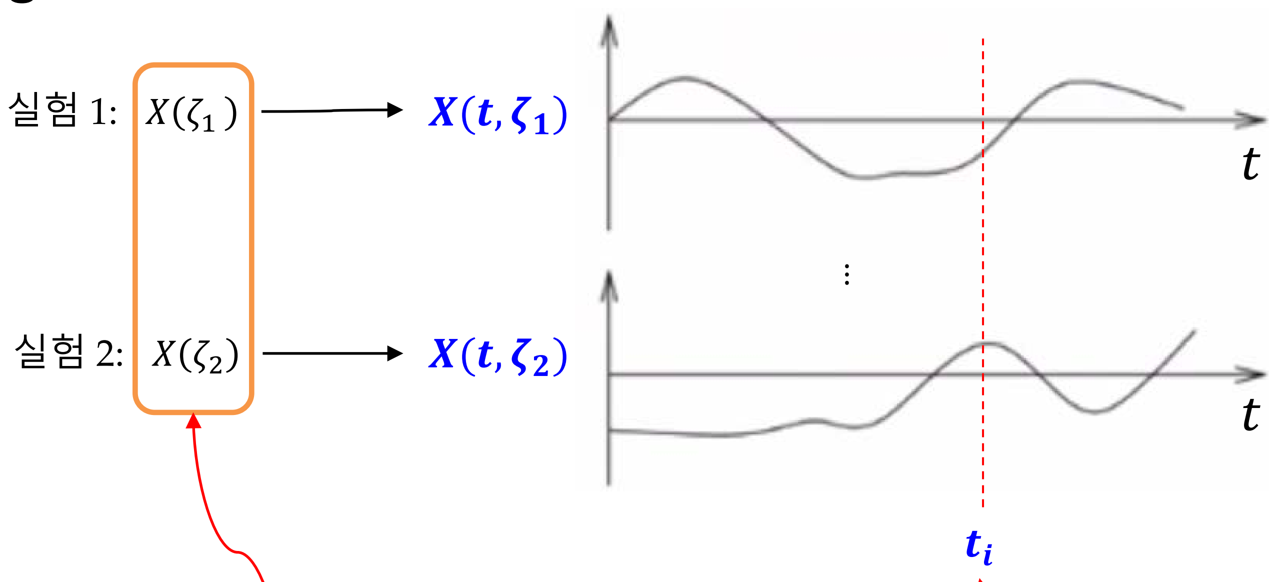
$\boldsymbol{X(t, \zeta_1)}$

$\boldsymbol{X(t, \zeta_2)}$



Detail in Ensemble

특징



Random process에서 시간 ζ_i 고정하면
$X(\zeta_i, t_i)$는 ***deterministic function***

Random process에서 시간 t_i 고정하면
$X(\zeta_i, t_i)$는 ***random variable***

Joint Distribution of Random Process

$X(t_i)$: random variable at a time t_i
characterized by PDF & CDF

$$F(x, t_i) = P[X(t_i) \le x]$$

$$f(x, t_i) = \frac{\delta F(x, t_i)}{\delta x}$$

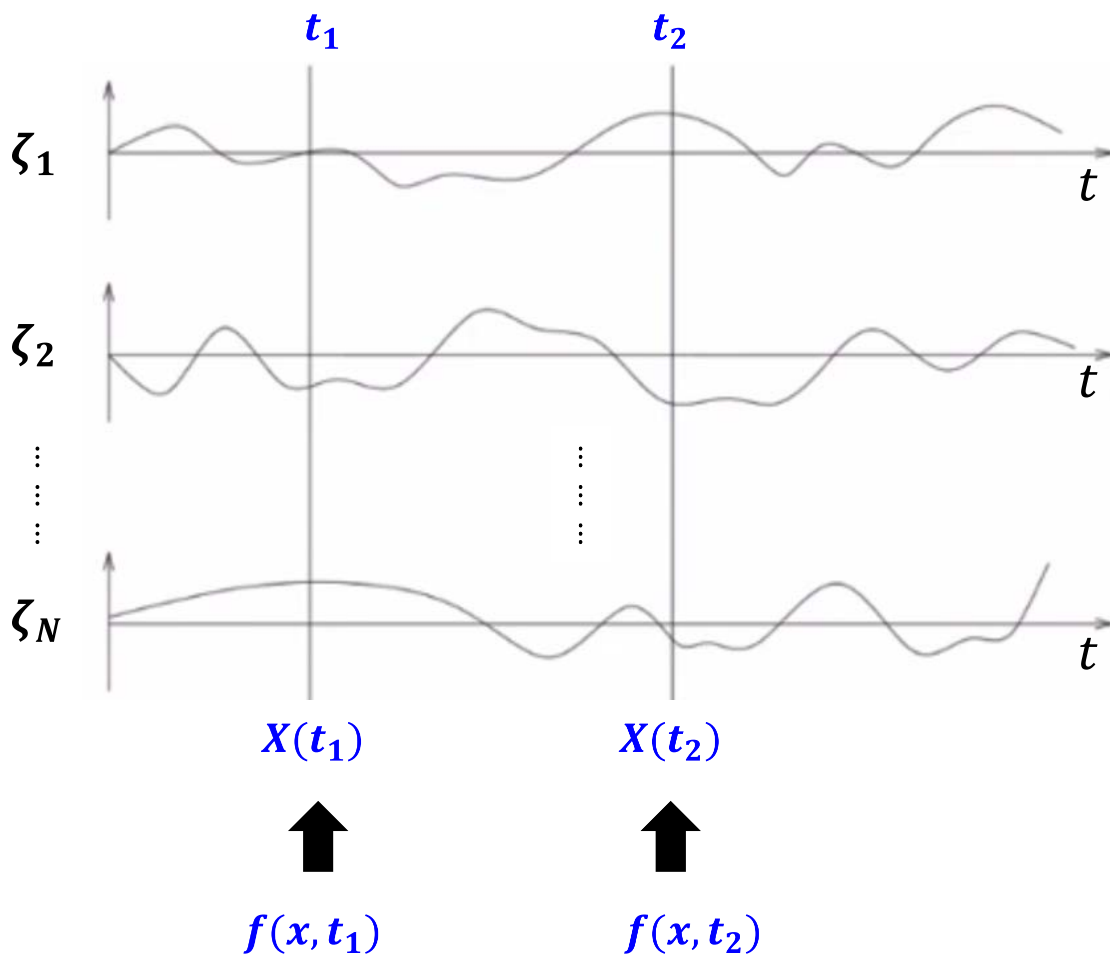
Joint Distribution of Random Process

$$F(x_1, x_2; t_1, t_2)$$

$$= P[X(t_1) \le x_1, X(t_2) \le x_2]$$

$$f(x_1, x_2; t_1, t_2) = \frac{\delta F(x_1, x_2; t_1, t_2)}{\delta x}$$

Ensemble



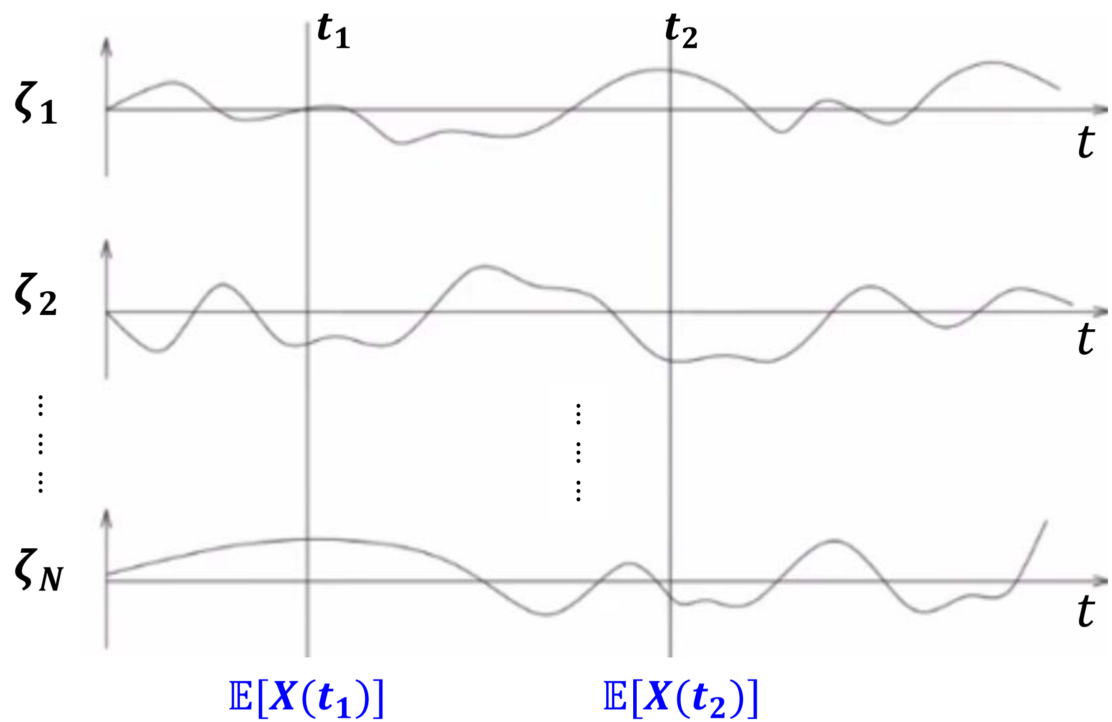
Expectation of Stochastic Process

Definition of expected value
in stochastic (random) process

$$\mu(t) = \mathbb{E}[X(t)] = \int_{-\infty}^{+\infty} x f(x, t)\ dx$$

Notice
- 개별 시간 축에 대한 적분 아님
- 특정 시간을 고정했을 때 각 ζ 값에 대한 적분

고정(fixed value)



Example: Stochastic Process on Exponential Function

$X(t) = e^{-Yt}$
where $Y \sim U(0, x)$ and $t \geq 0$ ➡ Let $g(y) = e^{-yt}$
$\because$ t is fixed (constant) ➡

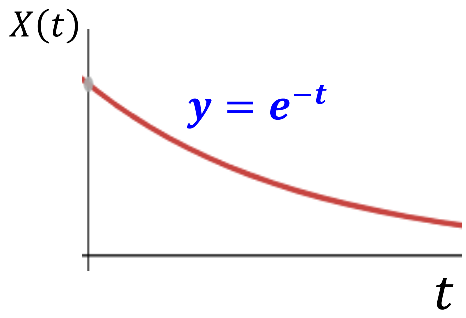
$$\mathbb{E}(X(t)] = \mathbb{E}[e^{-Yt}]$$

$$= \int_0^b g(y) \circ f_Y(y) dy$$

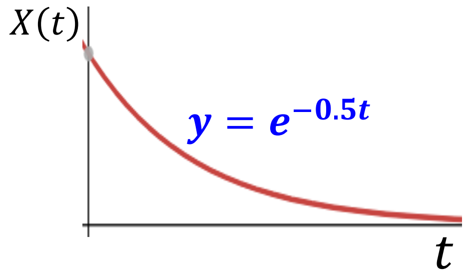
$$= \int_0^b e^{-yt} \times \frac{1}{b} dy$$

$$= \frac{1}{bt} \times \left(1 - e^{bt}\right)$$

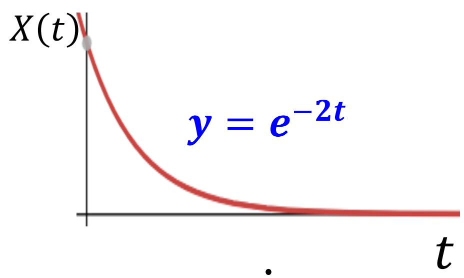
첫 번째 실험에서 얻은 결과 ➡



두 번째 실험에서 얻은 결과 ➡

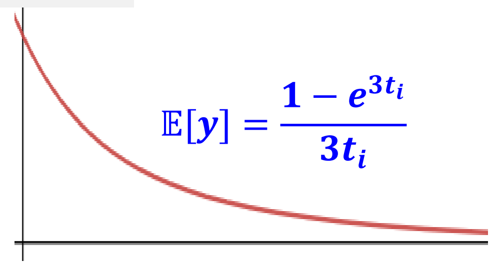


세 번째 실험에서 얻은 결과 ➡



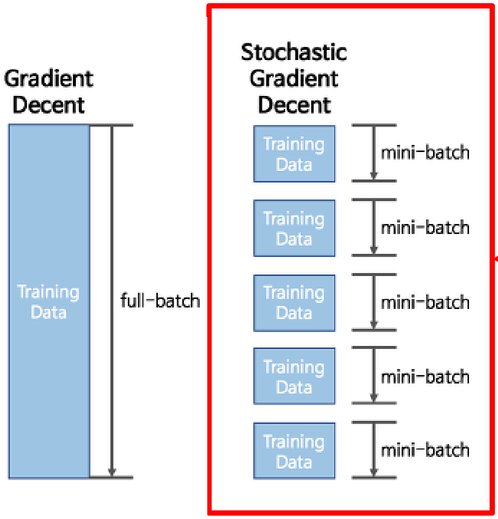
⋮ ⋮

Expectation of Ensemble



딥러닝에서 Stochastic?

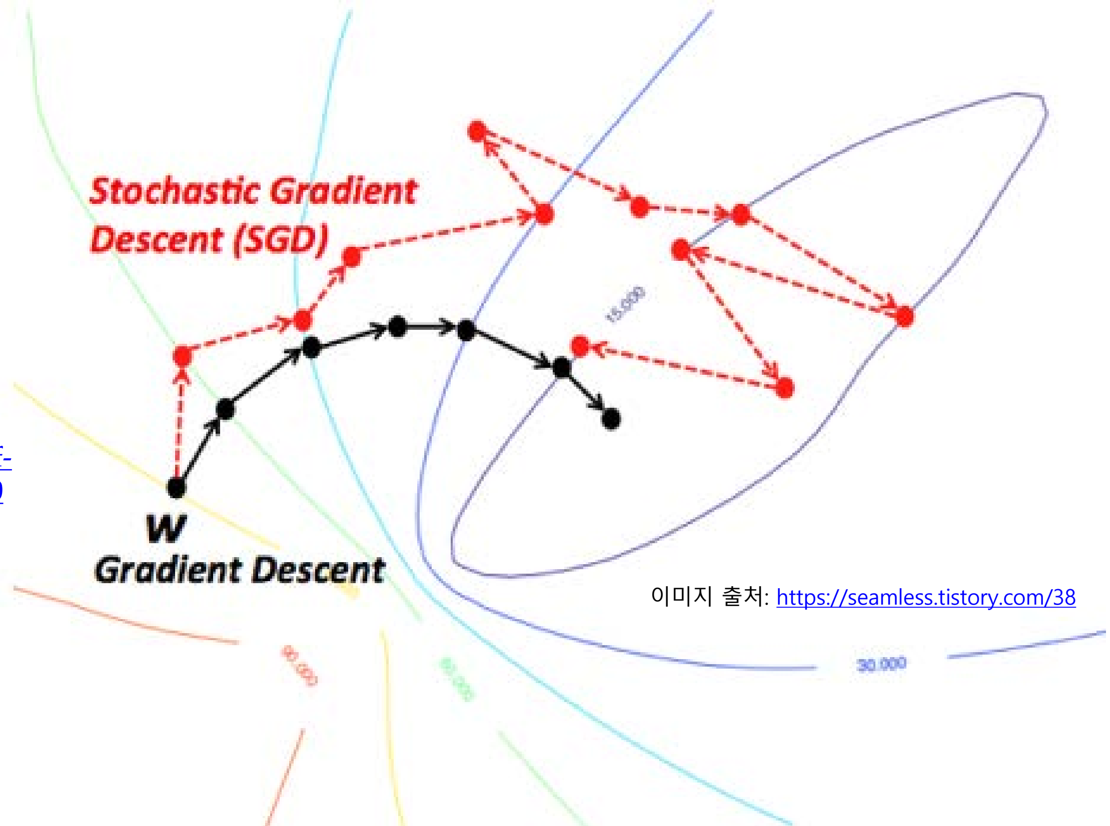
SGD in Deep Learning



Random하게
sub-sampling

그러면…

시간 개념은 어디에?



이미지 출처: https://seamless.tistory.com/38

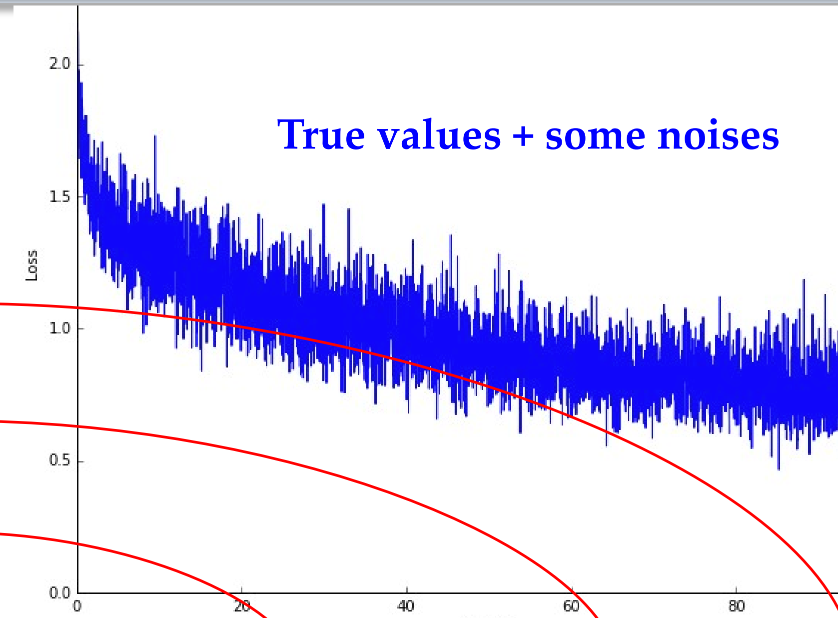
Proof of SGD convergence

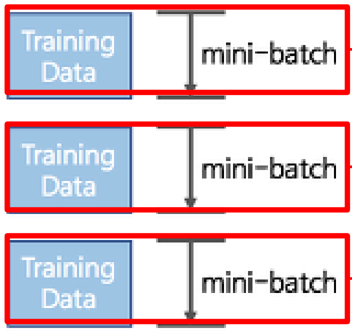
- Blog: https://medium.com/oberman-lab/proof-for-stochastic-gradient-descent-335bdc8693d0

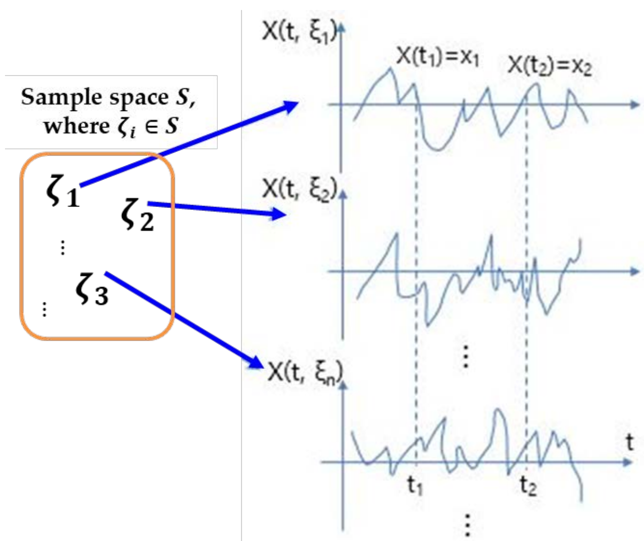
왜 Stochastic 단어가 SGD 앞 글자로 붙었을까?

Stochastic **Gradient Descent**

Stochastic **Process**







	$X(mb_1;\theta_1) = L(\theta_1)$	$X(mb_2;\theta_2) = L(\theta_2)$	⋯
1st epoch : ζ_1	$\boldsymbol{X(t_1) = x_1}$	$\boldsymbol{X(t_2) = x_2}$	⋯
	$X(mb_1;\theta_1) = L(\theta_1)$	$X(mb_2;\theta_2) = L(\theta_2)$	⋯
2nd epoch : ζ_2	$\boldsymbol{X(t_1) = x_1}$	$\boldsymbol{X(t_2) = x_2}$	⋯
	$X(mb_1;\theta_1) = L(\theta_1)$	$X(mb_2;\theta_2) = L(\theta_2)$	⋯
3rd epoch : ζ_3	$\boldsymbol{X(t_1) = x_1}$	$\boldsymbol{X(t_2) = x_2}$	⋯

참고 자료

- **Probability Lecture 9: Stochastic Processes**
  - Geoffrey Messier
    - https://youtu.be/WzFS3g0PM_M?si=iVCwCZ3emEolXed7
    - https://www.youtube.com/playlist?list=PL7sWxFnBVJLUbrCHertPLEqqCyLVnG-tN

수고하셨습니다 ..^^..