

Differentiation

Learning in Deeplearning (딥러닝 학습)

소프트웨어 공대 강의

노기섭 교수

(kafa46@cju.ac.kr)

Course Overview

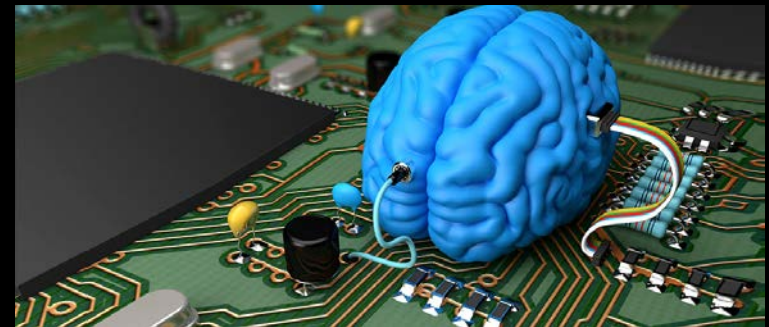
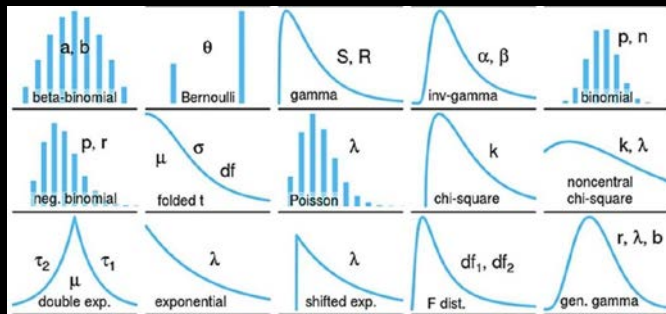
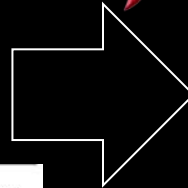
Topic	Contents
01. Orientation 오리엔테이션	Course introduction, motivations, final objectives 과정 소개, 동기부여, 최종 목표
02. Learning in deeplearning 딥러닝 학습	How does the deeplearning learns knowledge from data 어떻게 딥러닝은 데이터로부터 지식을 배우는가?
03. Principle of differentiation 미분의 원리	Basics of differentiation (concepts, notation, operations) 미분 기본지식 (개념, 표기, 연산)
04. Partial differentiation 편미분	Concept & operation of partial differenciation 편미분 개념, 연산
05. Gradient descent 경사 하강법	Concept, interpretation and learning in gradient descent 경사하강 알고리즘 개념, 해석 및 학습
06. Chain rule 연쇄법칙	Concept & operation of chain rule 연쇄법칙 개념 및 연산
07. Matrix differentiation 행렬미분	Partial differentiation in linear system 선형시스템에서의 편미분
08. Back propagation 역전파 학습	The mechanism of back propagation 역전파 학습의 작동 방법
09. Gradient vanishing 기울기 소실	Quick overview on activation function, cause root of gradient vanishing and its counter-measure 활성함수 간단 소개, 기울기 소실 근본원인과 대책

딥러닝이 학습하는 방법

데이터로부터 패턴을 배운다

거기서 거기, 같은 말....

→ 데이터로부터 확률 분포를 배운다



이미지 출처:
<https://knowledge.wharton.upenn.edu/article/democratization-ai-means-tech-innovation/>

[Supervised Version] 사람 어떻게 학습할까?

사람이 학습하는 방법 ^^

202X학년도 대학수학능력시험 문제지

국어 영역

Classification Problem

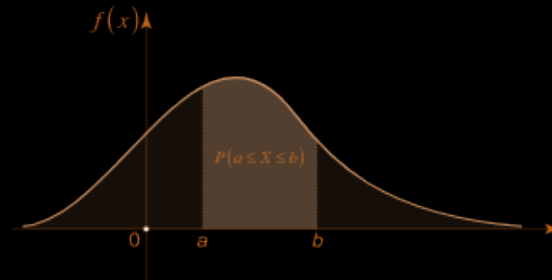
[1~3] 다음 글을 읽고 물음에 답하시오.

독서는 독자가 목표한 결과에 도달하기 위해 글을 읽고 의미를 구성하는 인지 행위이다. 성공적인 독서를 위해서는 초인지가 중요하다. 독서에서의 초인지는 독자가 자신의 독서 행위에 대해 인지하는 것으로서 자신의 독서 과정을 점검하고 조정하는 역할을 한다.

[A] 초인지는 글을 읽기 시작한 후 지속적으로 이루어지는 점검 과정에 동원된다. 독자는 가장 적절하다고 판단한 독서 전략을 사용하여 독서를 진행하는데, 그 전략이 효과적이고 문제가 없는지를 평가하며 점검한다. 효과적이지 않거나 문제가 있다고 판단하면 이를 해결해야 한다. 문제가 무엇인지 분명하지 않은 경우에는 독서 중에 떠오르는 생각들을 살펴보고 그중 독서의 진행을 방해하는 생각들을 분류해보는 방법으로 문제점이 무엇인지 파악할 수 있다. 독서가 중단 없이 이어지는 상태이지만 문제가 발생한 것을 독자가 자신이 인지하지 못하는 경우도 있다. 의도한 목표에 부합하지 않는 방법으로 읽기를 진행하거나 자신이 이해한 정도를 판단하지 못하는 예가 그것이다. 문제 발생 여부의 점검을 위해서는 독서 진행 중간중간에 이해한 내용을 정리하는 방법을 사용할 수 있다.

1. 윗글을 이해한 내용으로 적절하지 않은 것은?

- ① 독서 전략을 선택할 때 독서의 목표를 고려할 필요가 있다.
- ② 독서 전략의 선택을 위해 개별 전략들에 대한 지식이 필요하다.
- ③ 독서 목표의 달성을 위해 독자는 자신의 독서 행위에 대해 인지해야 한다.
- ④ 독서 문제의 해결을 위해 독자는 자신이 사용할 수 있는 전략이 무엇인지 알아야 한다.
- ⑤ 독서 문제를 해결하기 위해 새로 선택한 전략은 점검과 조정의 대상에서 제외할 필요가 있다.



예측
(A)

[0.05
0.15
0.20
0.35
0.25]

출력
(B)

[False
False
False
True
False]

정답
(C)

[False
False
False
False
True]

정답인지 확인(채점)

if $B - C == 0$:
correct!!

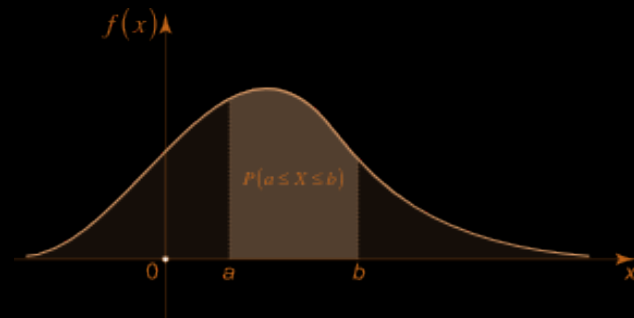
else:
wrong !!

문제를 다시 풀어본다
맞출때까지 계속 풀어본다.

[Supervised Version] 사람의 고득점 전략은?



1. 기출 문제 또는 유사한 문제를 최대한 모은다.
2. 문제를 최대한 풀어본다 (반복적으로!)
3. 문제가 주어졌을 때 정답 확률을 최대화 하는
분포를 배운다.
4. 수능 시험을 보러 간다.



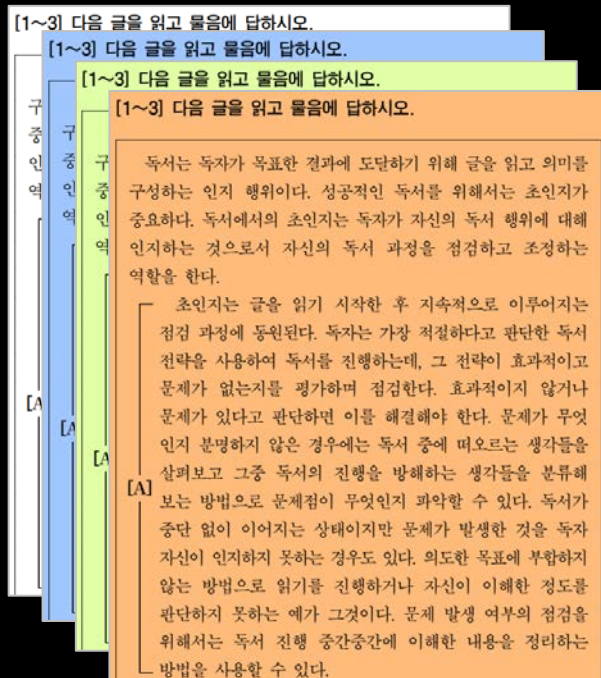
1. 윗글을 이해한 내용으로 적절하지 않은 것은?

1. 윗글을 이해한 내용으로 적절하지 않은 것은?

1. 윗글을 이해한 내용으로 적절하지 않은 것은?

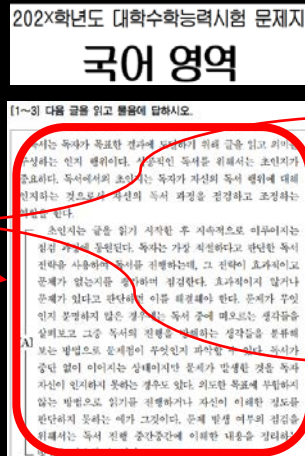
1. 윗글을 이해한 내용으로 적절하지 않은 것은?

- ① 독서 전략을 선택할 때 독서의 목표를 고려할 필요가 있다.
- ② 독서 전략의 선택을 위해 개별 전략들에 대한 지식이 필요하다.
- ③ 독서 목표의 달성을 위해 독자는 자신의 독서 행위에 대해 인지해야 한다.
- ④ 독서 문제의 해결을 위해 독자는 자신이 사용할 수 있는 전략이 무엇인지 알아야 한다.
- ⑤ 독서 문제를 해결하기 위해 새로 선택한 전략은 점검과 조정의 대상에서 제외할 필요가 있다.



[Supervised Version] 컴퓨터는 어떻게 학습할까?

컴퓨터가 학습하는 방법 ^^



Classification Problem

1. **읽글을 이해한 내용으로 적절하지 않은 것은?**
- ① 독서 전략을 선택할 때 독서의 목표를 고려할 필요가 있다.
 - ② 독서 전략의 선택을 위해 개별 전략들에 대한 지식이 필요하다.
 - ③ 독서 목표의 달성을 위해 독자는 자신의 독서 행위에 대해 인지해야 한다.
 - ④ 독서 문제의 해결을 위해 독자는 자신이 사용할 수 있는 전략이 무엇인지 알아야 한다.
 - ⑤ 독서 문제를 해결하기 위해 새로 선택한 전략은 점검과 조정의 대상에서 제외할 필요가 있다.

예측
(A)

$\begin{bmatrix} 0.05 \\ 0.15 \\ 0.20 \\ 0.35 \\ 0.25 \end{bmatrix}$

출력
(B)

$\begin{bmatrix} \text{False} \\ \text{False} \\ \text{False} \\ \text{True} \\ \text{False} \end{bmatrix}$

출력
(C)

$\begin{bmatrix} \text{False} \\ \text{False} \\ \text{False} \\ \text{False} \\ \text{True} \end{bmatrix}$

얼마나 틀렸는지
계산한다.

$$\sum_{i=1}^n |C - A|$$

$$\sum_{i=1}^n \left| \begin{pmatrix} 0.05 \\ 0.15 \\ 0.20 \\ 0.35 \\ 0.25 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right| = \sum_{i=1}^n \left| \begin{pmatrix} 0.05 \\ 0.15 \\ 0.20 \\ 0.35 \\ -0.75 \end{pmatrix} \right|$$

$$= 0.05 + 0.15 + 0.20 + 0.35 + 0.75 = 1.5$$

정답인지 확인

if $B - C == 0$:

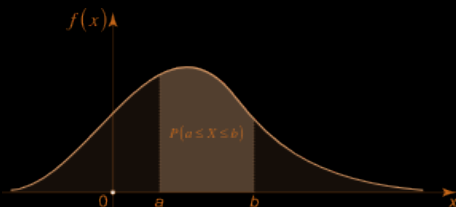
correct!!

else:

wrong π

아! 1.5 만큼 틀렸구나...

틀린 정도를 줄이도록
파라미터를 업데이트 하자!



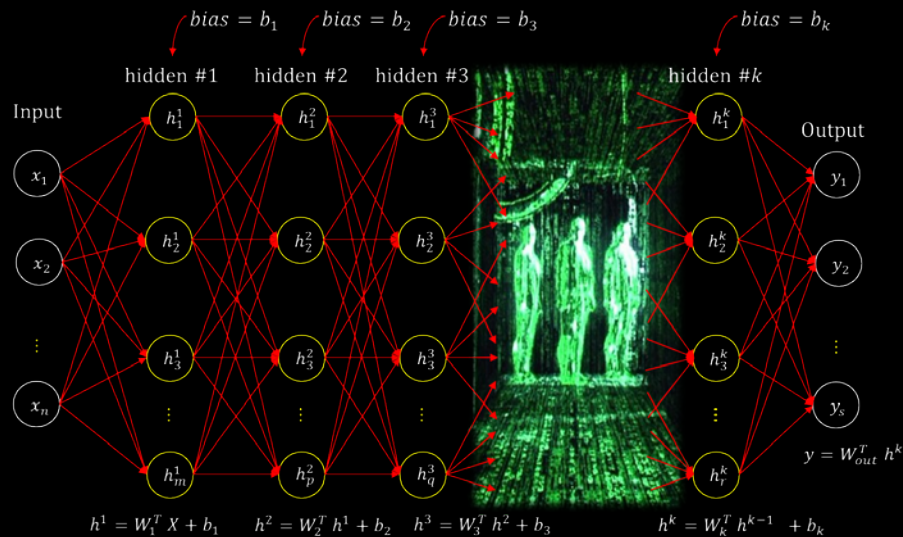
Learning from Dataset

<i>index</i>	<i>x</i>	<i>y</i>
1	<p>[1~3] 다음 글을 읽고 물음에 답하시오.</p> <p>독서는 독자가 목표한 결과에 도달하기 위해 글을 읽고 의미를 구성하는 인지 행위이다. 성공적인 독서를 위해서는 초인자가 중요하다. 독서에서의 초인지는 독자가 자신의 독서 행위에 대해 인지하는 것으로서 자신의 독서 과정을 점검하고 조정하는 역할을 한다.</p> <p>1.윗글을 이해한 내용으로 적절하지 않은 것은?</p>	<p>⑤ 독서 문제를 해결하기 위해 새로 선택한 전략은 점검과 조정의 대상에서 제외할 필요가 있다.</p>
⋮	⋮	⋮

교수님!
장난하세요?



데이터 x 를
숫자로 만들고
입력 계층에
전달한다.



y 가 나올 확률을
계산한다.

얼마나 틀렸는지
계산한다.

틀린 정도를 최소화할 수
있도록 모든 파라미터를
업데이트 한다.

모든 파라미터를 랜덤하게 설정한다.

교수님! 그냥 일일히 업데이트 해도 되는거 아니가요?

교수님! 파라미터 몇 개 밖에 안되면
그냥 업데이트 해도 되는거 아니가요?

파라미터가 2개라면?

고려해야 할 경우의 수 $2! = 2$

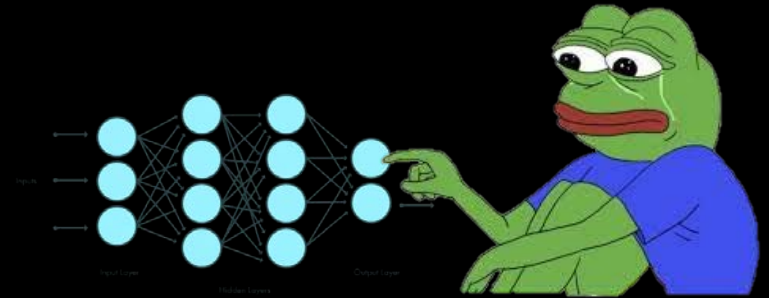
파라미터가 10개라면?

고려해야 할 경우의 수 $10! = 3,628,800$

파라미터가 100개라면?

고려해야 할 경우의 수 100!

= 9332621544394415268169923885626670049071596826438162146859296389521759999322991
560894146397615651828625369792082722375825118521091686400000000000000000000000



BERT large: 3억 4천만 개

T5: 110억 개

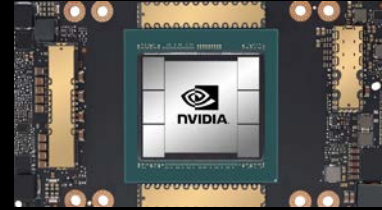
ChatGPT: 1,750억 개

앞으로 더더더더 커질 예정~~

파라미터를 손으로

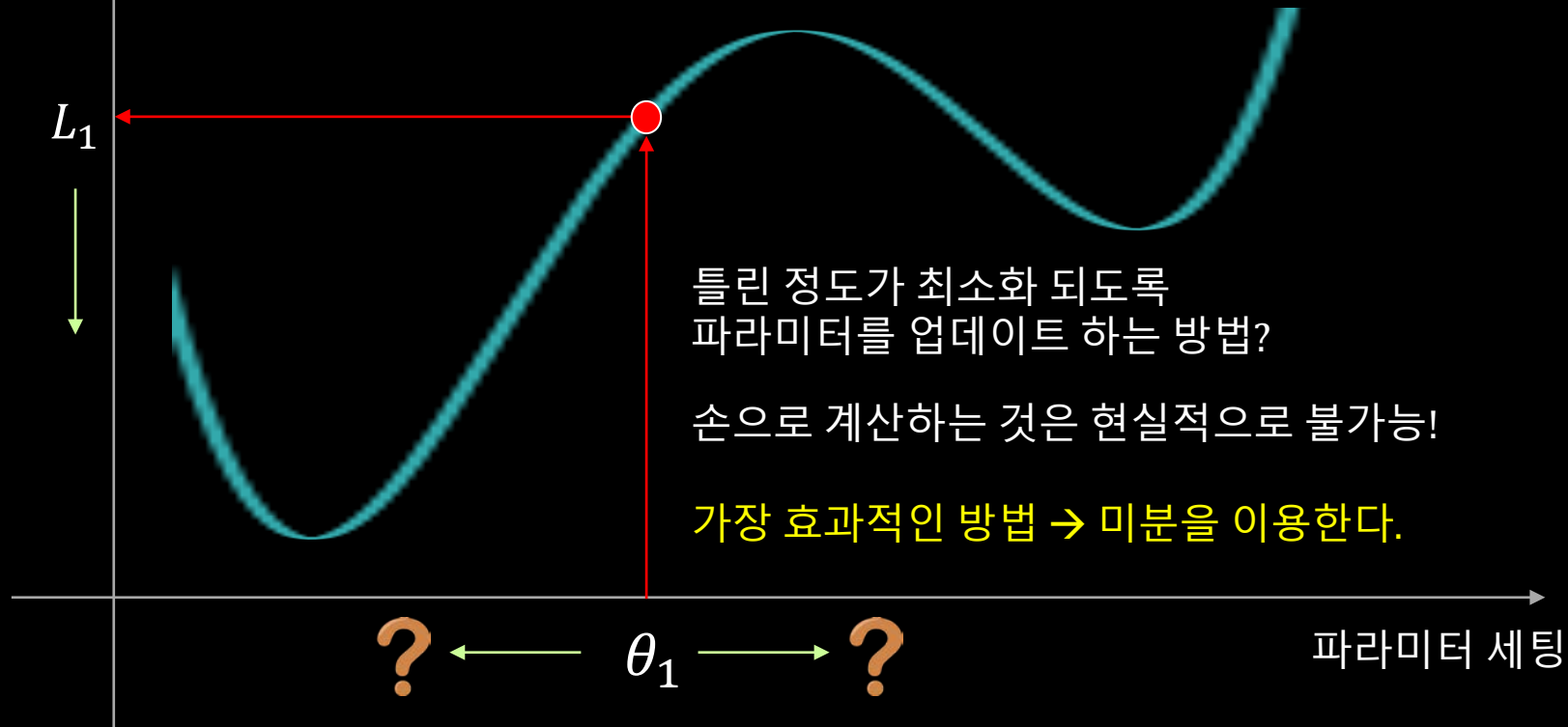
계산하는 것은 Intractable!!!

그나마 제일 합리적인 방법



이 방법도...
사실은... 엄청나게 많은
연산이 필요합니다.

틀린 정도
(Loss)



틀린 정도가 최소화 되도록
파라미터를 업데이트 하는 방법?

손으로 계산하는 것은 현실적으로 불가능!

가장 효과적인 방법 → 미분을 이용한다.

어느 방향으로 얼마나 이동해야 하지?



수고하셨습니다 ..^^..