

2024-1

# Multivariate Data Analysis

Youjin Shin  
(yj.shinn@catholic.ac.kr)



# About Professor

Youjin Shin (신유진)

- Assistant professor(Department of Data Science, Department of Artificial Intelligence)
- Ph.D. in Computer Science, the State University of New York at Stonybrook
- Research Interest: Medical AI, AI-driven prediction & anomaly detection
- Email: yj.shinn@catholic.ac.kr
- Homepage: <https://sites.google.com/view/adslab0>
- Office: T405, 02-2164-5524



**ADSLAB**

AI-driven Data Science Lab

# About Professor

## Medical AI

### DO YOU KNOW THE DIFFERENT TYPES OF SKIN CANCER?



**Basal Cell**  
Most common form of skin cancer



**Squamous Cell**  
Second most common form of skin cancer

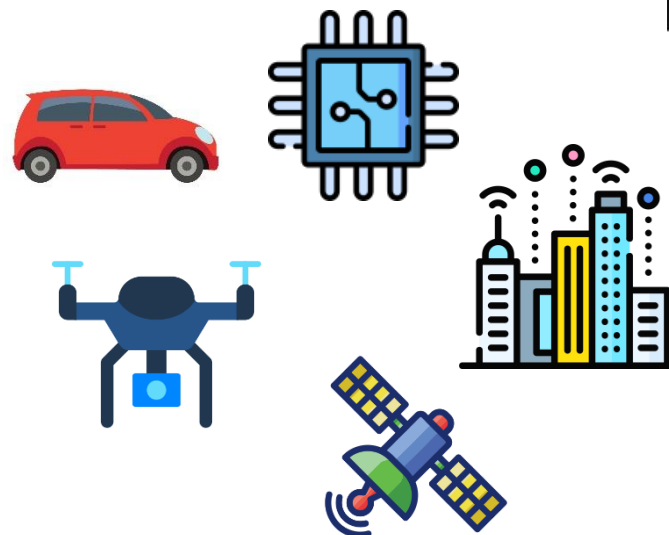
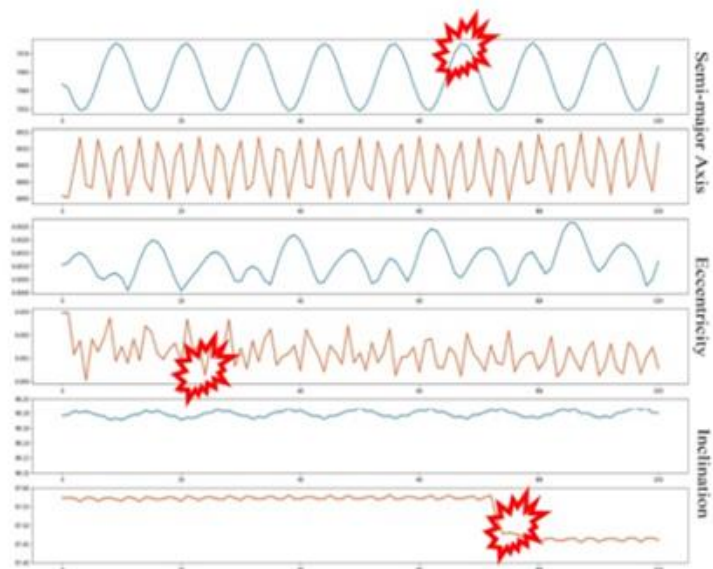


**Melanoma**  
Most serious form of skin cancer



골다공증성 척추압박골절 환자

## Anomaly Detection & Prediction



Others..



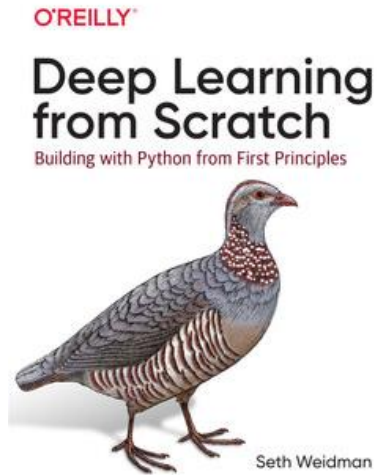
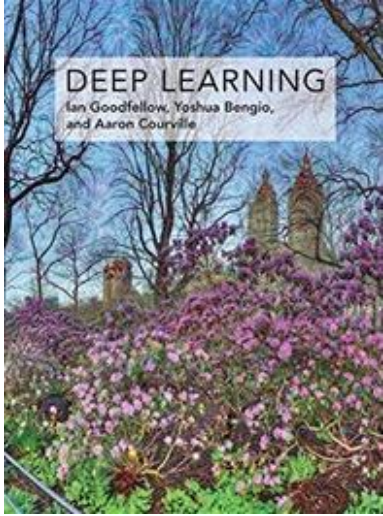
# Learning Objectives

---

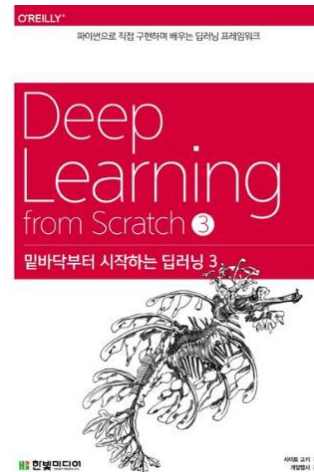
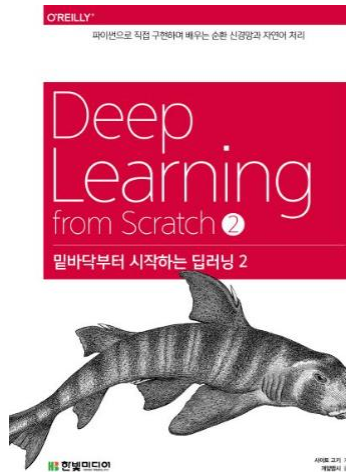
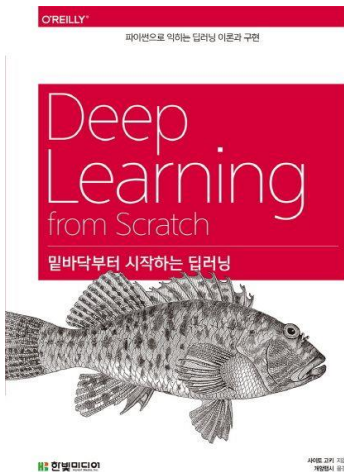
## Multivariate Data Analysis

- This course is designed to provide learners with practical skills in data science using the Python programming language. Through engaging projects and real-world applications, participants will acquire a comprehensive understanding of data analysis, machine learning, artificial intelligence, and data visualization.
- Firstly, students will develop a strong proficiency in Python, particularly for data manipulation and analysis. Building on this foundation, learners will delve into the fundamental concepts of machine learning, applying these principles to real-world datasets. Through hands-on experience, participants will gain practical expertise in data analysis techniques and effective communication of their findings. The course also emphasizes the application of acquired skills in real-world scenarios through engaging projects, fostering a practical and problem-solving-oriented approach to data science.

# Books



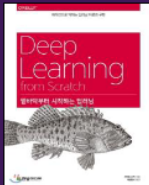
- Deep Learning (Author: Ian Goodfellow)
- Deep Learning from Scratch (O'REILLY)
- Deep Learning Series (O'REILLY)
- 모두의 딥러닝 (길벗)
- 딥러닝 텐서플로 교과서(길벗)





## 첫 번째 이야기 밑바닥부터 시작하는 딥러닝 ①

### 합성곱 신경망(CNN)과 이미지 처리



파이썬 기초   파셉트론   신경망 학습   역전파

합성곱 신경망   계산 그래프   이미지 인식

딥러닝

라이브러리나 프레임워크에 의존하지 않고, 딥러닝의 핵심을 '밑바닥부터' 직접 만들어보며 즐겁게 배울 수 있는 본격 딥러닝 입문서다. 술술 읽힐 만큼 쉽게 설명하였고, 역전파처럼 어려운 내용을 '계산 그래프' 기법으로 시각적으로 풀이했다. 무엇보다 작동하는 코드가 있어서 직접 돌려보고 요리조리 수정해보면 어려운 이론도 명확하게 이해할 수 있다. 딥러닝에 새롭게 입문하려는 분과 기초를 다시금 정리하고 싶은 현업 연구자와 개발자에게 최고의 책이 될 것이다.

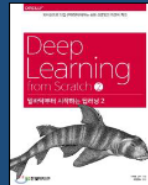
- 1 헬로 파이썬**  
파이썬 언어와 넘파이, 넷플릿을 소개하고 사용법을 익힌다.
- 2 파셉트론**  
신경망의 기원이 되는 알고리즘이라 할 수 있는 파셉트론을 설명하고 파셉트론을 사용해 간단한 문제를 풀어본다.
- 3 신경망**  
파셉트론의 한계를 극복해주는 신경망의 개념을 설명하고, 입력 데이터의 정체를 식별하는 신경망을 구현한다.
- 4 신경망 학습**  
훈련 데이터를 이용해 신경망을 학습시킨다. 또한 학습 지표로서의 손실 함수, 최적의 손실 함수를 찾는 경사법을 알아본다.
- 5 오차역전파법(역전파)**  
가중치 매개변수의 기울기를 효율적으로 계산하는 오차역전파법을 '계산 그래프'라는 시각적 기법으로 설명한다.
- 6 학습 관련 기술들**  
매개변수 갱신, 가중치의 초기값, 배치 정규화, 과대적합(오버피팅), 드롭아웃, 하이퍼파라미터 최적화 등의 기술을 배운다.
- 7 합성곱 신경망(CNN)**  
이미지 인식, 음성 인식 등에 활용되는 합성곱 신경망의 메커니즘을 자세히 설명하고 파이썬으로 구현해본다.
- 8 딥러닝**  
층을 깊게 쌓은 심층 신경망(딥러닝)의 특징, 풀어야 할 과제, 가능성을 이야기한다. 또한 다양한 첨단 딥러닝 기법을 소개한다.

### 리뷰어들의 목소리

- "탄탄한 구성으로 신경망과 딥러닝 전반을 쉽게 설명해주고, 특히 역전파는 제가 본 교재 중 가장 쉽게 서술했습니다."  
안홍철 에셋클래스자산운용 자산운용사
- "수학과 프로그래밍 언어에 익숙하지 못한 의로 관련 분들에게도 이 책은 제목처럼 '밑바닥부터' 딥러닝을 이해시켜줄 것입니다."  
임형택 안과전문의
- "이 책의 원서는 일본 TV에 소개될 정도로 인기입니다. 한국에서도 이 책이 AI 보급에 중요한 기록제가 될 것 기대합니다."  
최성훈 Tomomi Research Inc.
- "딥러닝 공부를 시작하시는 모든 분께 추천합니다."  
서민구 구글코리아 소프트웨어 엔지니어

## 두 번째 이야기 밑바닥부터 시작하는 딥러닝 ②

### 순환 신경망(RNN)과 자연어 처리



순환 신경망   자연어 처리   시계열 데이터

word2vec   LSTM   seq2seq

어텐션

1편에서 다루지 못한 순환 신경망(RNN)을 자연어 처리와 시계열 데이터 처리에 사용하는 딥러닝 기술에 초점을 맞춰 살펴본다. 총 8개 장으로 구성되며, 전체를 하나의 이야기처럼 순서대로 읽도록 꾸몄다. 1편에서 배운 내용을 요약한 '신경망 복습'을 첫 장에 배치하여 신경망과 파이썬 지식을 어느 정도 갖춘 분이라면 1편을 읽지 않았더라도 무리 없이 따라올 수 있도록 배려했다.

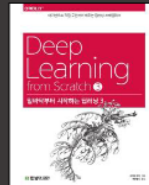
- 1 신경망 복습**  
1편의 내용을 요약하고, 구현 규칙과 달라진 점을 소개한다.
- 2 자연어와 단어의 분산 표현**  
컴퓨터에게 말을 이해시킨다는 것의 뜻과 방법들을 알아보고, 파이썬으로 텍스트를 처리하는 연습을 해본다.
- 3 word2vec**  
주요 기본 기법(신경망)으로 단어의 분산 표현을 얻어내는 '단순한' word2vec를 구현해본다.
- 4 word2vec 속도 개선**  
Embedding과 네거티브 샘플링을 도입해 '진짜' word2vec를 완성한다.
- 5 순환 신경망(RNN)**  
4장까지 살펴본 피드포워드 신경망의 문제점을 짚어보고, 순환 신경망으로 그 문제를 해결한다.
- 6 게이트가 추가된 RNN**  
5장의 기본 RNN에 게이트를 추가하여 장기 의존 관계를 학습시킨다(LSTM).
- 7 RNN을 사용한 문장 생성**  
언어 모델을 사용해 '이가 글을 쓰게' 해보고, seq2seq를 구현하여 하나의 시계열을 다른 시계열로 변환해본다.
- 8 어텐션**  
어텐션을 적용해 seq2seq를 개선하여 효과를 확인해보고, 어텐션의 구조를 코드 수준에서 이해해본다.

### 리뷰어들의 목소리

- "널리 사용되는 딥러닝 모델을 직접 구현하면서 기본 원리를 체득하는 체험형 입문서입니다."  
윤영선 한남대학교 정보통신공학과 교수
- "쉽게 풀이한 이론과 친절하게 설명한 코드를 따라가다 보면 자연어 처리 분야의 발전 과정을 자연스럽게 이해할 수 있습니다."  
양민혁 현대모비스 빅데이터팀
- "저와 같은 어린 학생들이 '과연 할 수 있을까?' 망설이고 있다면, 주저 없이 이 책을 집어 드세요."  
김경수 봉일천고등학교 2학년
- "자연어 처리 기본서는 이 책의 전과 후로 나뉘 겁니다."  
성민서 홍익대학교 컴퓨터공학과 4학년

## 세 번째 이야기 밑바닥부터 시작하는 딥러닝 ③

### 동적 계산 그래프와 딥러닝 프레임워크

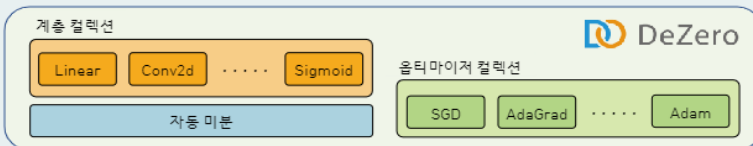
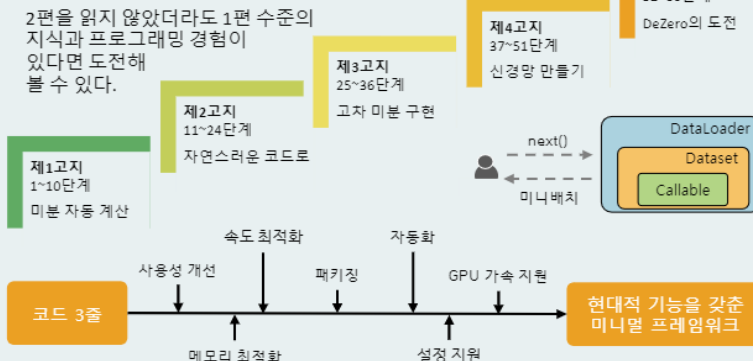


자동 미분   변수와 함수   동적 그래프 (Define-by-Run)

고차 미분   옵티마이저   계층

데이터 로더   전처리

파이토치와 텐서플로 같은 현대적인 프레임워크가 제공하는 공통 기능을 갖춘 미니멀 딥러닝 프레임워크를 밑바닥부터 구현한다. 딥러닝 프레임워크 안은 놀라운 기술과 재미있는 장치로 가득하다. 3편의 목표는 그것들을 밖으로 꺼내어 제대로 이해시키는 것이다. 다른 누군가가 만들어놓은 도구를 사용하지만 해서는 도달하기 어려운 깊은 깨달음을 얻어갈 것이다.



### 리뷰어들의 목소리

- "이 책을 학습하고 나면 텐서플로나 파이토치 같은 프레임워크의 코드를 직접 보는 것도 한결 수월해질 것입니다."  
박진형 XL8 공동 창업자, 전 애플 엔지니어
- "딥러닝을 대략적으로만 알고 있는 분들이 부족했던 부분을 채워 한층 완벽해질 수 있도록 꼼꼼하고 자세하게 설명합니다."  
안상준 '파이썬 딥러닝 파이토치' 저자
- "현재 주류인 Define-by-Run의 구조와 인공지능의 기본 설계 두 마리 토끼를 모두 설명해낸 책입니다."  
시한 VAIS 인공지능 커뮤니티 운영진
- "설레는 모험으로 가득한, 마치 잘 짜여진 RPG 게임을 한 레벨씩 정복하는 듯한 몰입감이 있습니다."  
전두용 성운대 교양학부 교수

# Weekly Schedule

3월

SUN	MON	TUE	WED	THU	FRI	SAT
					1	2
3	4 개강	5	6 수강변경 기간	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

4월

SUN	MON	TUE	WED	THU	FRI	SAT
	1	2	3	4	5	6
7	8 개강	9	10 선거일	11	12	13
14	15	16	17	18	19	20
21	22 중간고사	23	24	25	26	27
28	29	30				

# Weekly Schedule

5월

SUN	MON	TUE	WED	THU	FRI	SAT
			1	2	3	4
5	6 어린이날 대체휴일	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

6월

SUN	MON	TUE	WED	THU	FRI	SAT
						1
2	3 프로젝트 발표	4	5 프로젝트 발표	6	7	8
9	10 프로젝트 발표(보강)	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						



# Grading

- Evaluation Criteria :

Midterm exam	30 %
Final Project	40 %
Assignment	20 %
Attendance	10 %
Total	100 %

- Grading (Absolute evaluation) :

Score	Grade
95 - 100	A+
90 - 95	A0
85 - 90	B+
80 - 85	B0
75 - 80	C+
70 - 75	C0
65 - 70	D+
60 - 65	D0
Below 60	F

# Quiz & Exam Policy

---

- The quiz consists of 10 simple questions.
  - yes/no choice
  - multiple choice
  - short answer
  - one-sentence explanation
- Midterm and final project:
  - Midterm exam day : 4/22 (MON) @ B346
  - Open-book Test😊
  - Project Presentation days : 6/3,5,10 @ B346

# Attendance Policy

---

- In 5 minutes after class starts : Attend
  - In 30 minutes after class starts : Late (3 Lates = 1 Absent)
  - After 30 minutes : Absent
- 
- Basically, we will use the cyber campus attendance system.
  - However, when I call your names one day and you are not actually in the classroom, 3 absences will be counted.