Some Thoughts on Data Science

by Prof. Seungchul Lee
Industrial AI Lab
http://isystems.unist.ac.kr/
POSTECH

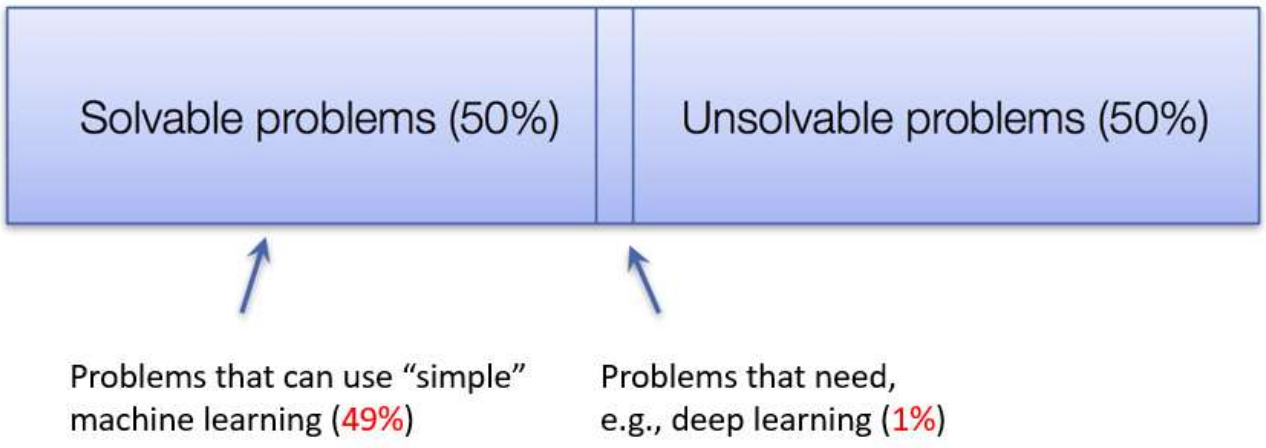
Table of Contents

- I. 1. Get Involved in Data Science Now
- II. 2. Study Materials

1. Get Involved in Data Science Now

- Most researchers will be interested in how data science and machine learning techniques can be applied to their domains
- but you will need to spend substantial time learning the domain itself

Data Problems We Would Like to Solve



Solving with Deep Learning

- When you come up against some machine learning problem with “traditional” features (i.e., human-interpretable characteristics of the data),
  - do not try to solve it by applying deep learning methods first
  - Instead, use
    - linear regression/classification,
    - linear regression/classification with non-linear features, or
    - gradient boosting methods

- If you really want to squeeze out a 1-2% improvement in performance, then you can apply deep learning
  - However, it’s also undeniable that deep learning has made remarkable progress for structured data like images, audio, or text

What About “Superhuman” Machine Learning

- It’s a common misconception that machine learning will outperform human experts on most tasks
  - No, it is supervised learning
  - Cannot be better than your training data

- In reality, the benefit from machine learning often doesn’t come from superhuman performance in most cases,
  - it comes from the ability to scale up expert-level performance extremely quickly

Dealing with Impossible Problems

- You’ve built a tool to manually classify examples, run through many cases (or had a domain expert run through them), and you get poor performance

- What do you do?
  - You do not try to throw more, bigger machine learning algorithms at the problem

- Instead you need to change the problem by:
  - 1) changing the input (i.e., the features),
  - 2) changing the output (i.e., decomposing it to smaller sub-problems)

Chainging Input (i.e., Adding Features)

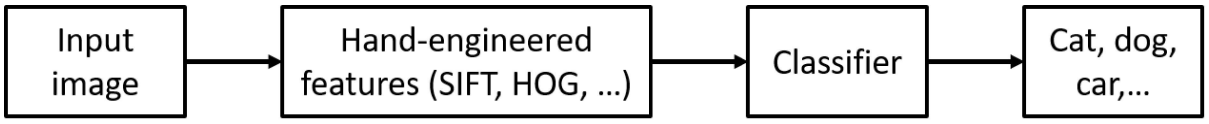
- Adding more data is good, but:
  - 1) Do spot checks (visually) to see if this new features can help you differentiate between what you were previously unable to predict
  - 2) Get advice from domain experts, see what sorts of data source they use in practice (if people are already solving the problem)

Changing Output (i.e., Changing the Problem)

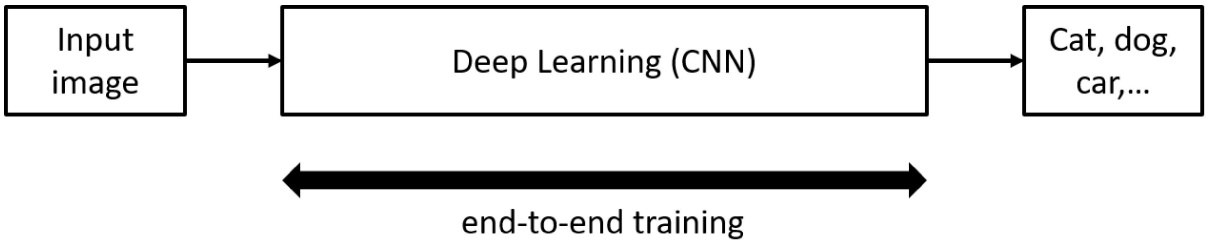
- Just make the problem easier!
- Decompose it to smaller sub-problems

Machine Learning vs. Deep Learning

- State-of-the-art until 2012



- Deep supervised learning



- Hyperparameters
  - Learning rate
  - # of iterations
  - # of hidden layers
  - # of hidden units
  - Choice of activation functions

2. Study Materials

Deep Learning for ME

- 딥러닝은 인공지능 연구자보다 여러분에게 더 필요할 수 있습니다.
- 새로운 기술을 어디에 적용해 볼 수 있을지 고민하세요.

인공지능 어떻게 공부할 것인가?

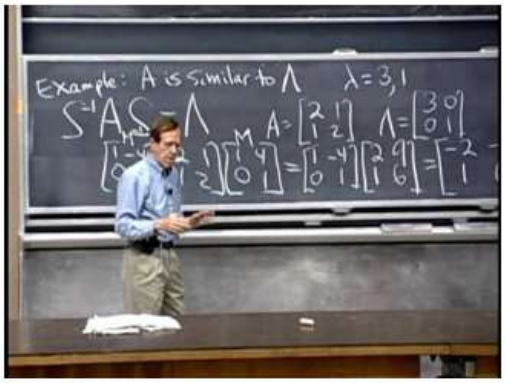
- Deep learning 으로 인공지능을 처음 공부하면 안된다.
- Linear algebra, Optimization, Statistics, Probability, Machine Learning
  - Then deep learning
- (Numerical or Scientific) Computer Programming
  - MATLAB or Python
  - 개념, 수식, 코드

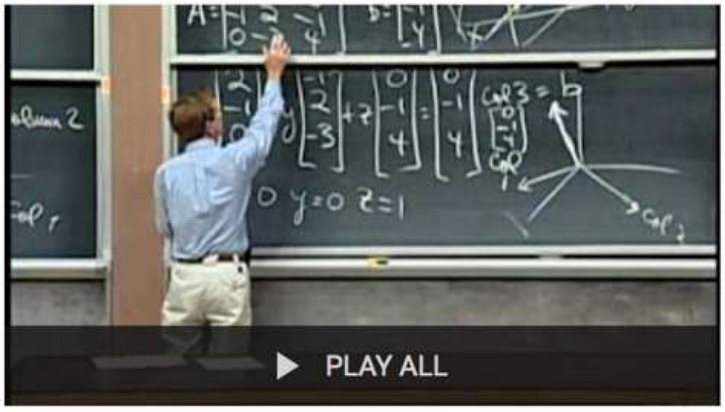
유용한 공부 자료

강의 대부분의 내용은 아래 연구자분들의 자료를 선택적으로 취합해서 만들어졌습니다.

1) Linear Algebra

- Gilbert Strang from MIT
- https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/ (https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/)
- YouTube

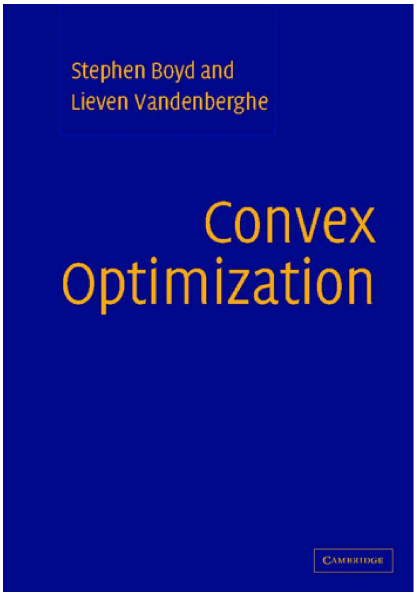


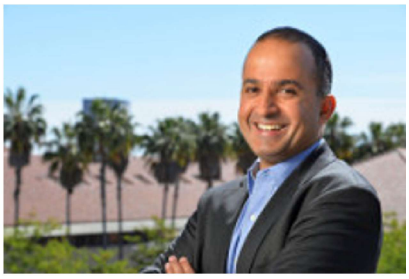


Gilbert Strang lectures on Linear Algebra (MIT)

2) Optimization and Linear Systems

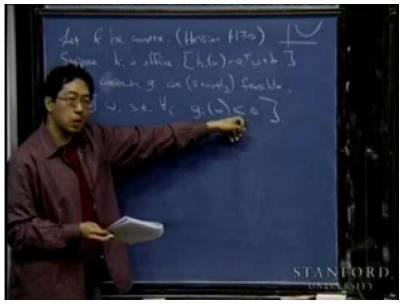
- Stephen Boyd and Sanjay Lall from Stanford
  - Linear Dynamical Systems
  - Linear Control Systems
  - Convex optimization
  - Textbook
  - http://stanford.edu/~boyd/ (http://stanford.edu/~boyd/)
  - https://lagunita.stanford.edu/courses/Engineering/CVX101/Winter2014/about (https://lagunita.stanford.edu/courses/Engineering/CVX101/Winter2014/about)



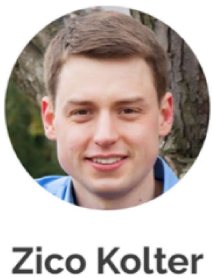


3) Machine Learning

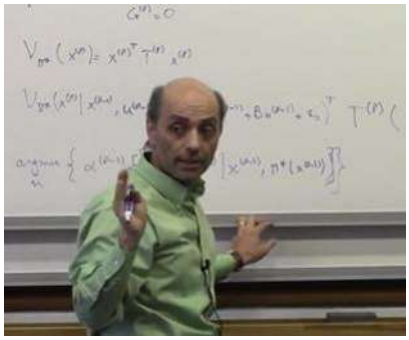
- CS229 - Machine Learning
  - Prof. Andrew NG from Stanford
  - Different from coursera
  - YouTube and lecture note
  - https://see.stanford.edu/Course/CS229 (https://see.stanford.edu/Course/CS229)



- Artificial Intelligence
  - Prof. Zico Kolter from CMU
  - YouTube and lecture note
  - http://zicokolter.com/ (http://zicokolter.com/)



- Learning theory
  - Prof. Reza Shadmehr from Johns Hopkins Univ.
  - YouTube
  - http://www.shadmehrlab.org/ (http://www.shadmehrlab.org/)

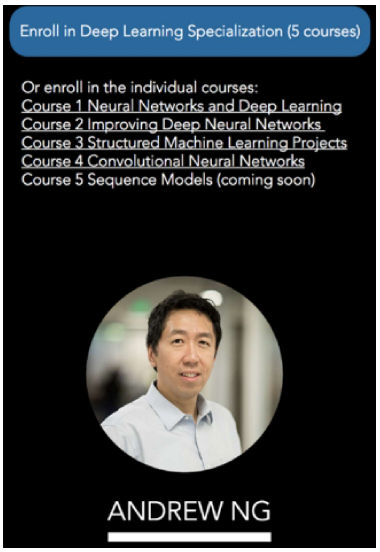


- Artificial Intelligence
  - Prof. Patrick Henry Winston from MIT
  - YouTube
  - https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/ (https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/)

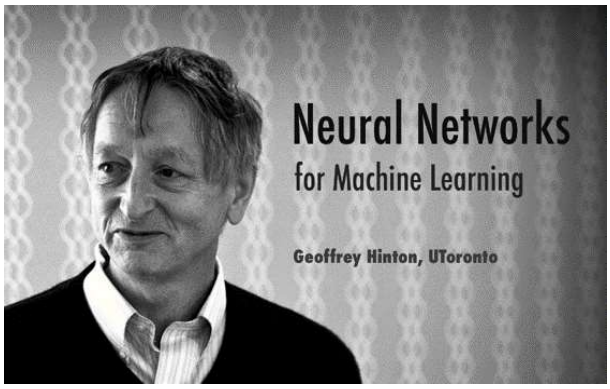
- Learning from data
  - Prof. Yaser Abu-Mostafa from Caltech
  - https://work.caltech.edu/telecourse.html (https://work.caltech.edu/telecourse.html)

4) Deep Learning

- Prof. Andrew Ag from Stanford
  - Coursera
  - http://deeplearning.ai/ (http://deeplearning.ai/)



- Neural Networks for Machine Learning
  - https://www.coursera.org/learn/neural-networks# (https://www.coursera.org/learn/neural-networks#)

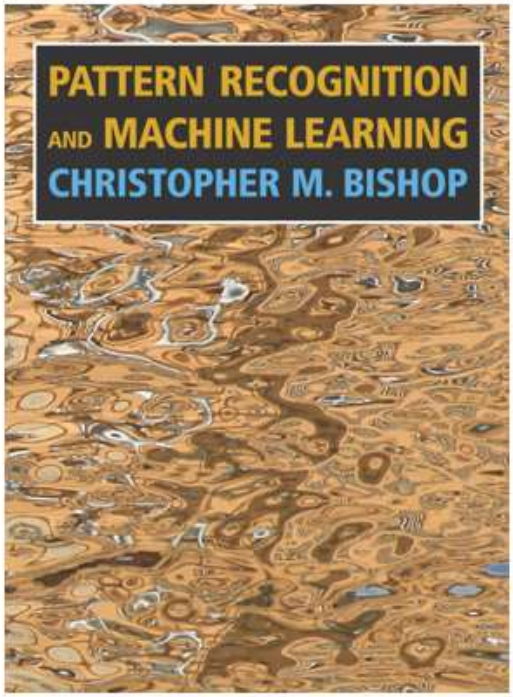


University Lectures on Deep Learning

- Stanford
  - CS231n: Convolutional Neural Networks for Visual Recognition
  - http://deeplearning.stanford.edu/tutorial/ (http://deeplearning.stanford.edu/tutorial/)
- CMU
- NYU
- MIT
- Toronto

Books

- Pattern Recognition and Machine Learning by Christopher Bishop
- https://www.microsoft.com/en-us/research/people/cmbishop/ (https://www.microsoft.com/en-us/research/people/cmbishop/)



한국어 강좌

- 김성훈 교수님, 홍콩과기대 (현재 네이버 AI 팀)
- https://hunkim.github.io/ml/ (https://hunkim.github.io/ml/)

- 문일철 교수님, 카이스트
- http://kaist.edwith.org/machinelearning1_17 (http://kaist.edwith.org/machinelearning1_17)
- https://www.youtube.com/playlist?list=PLt9QR0WkC4WVszuogbmIIHIIQ2RMI78RC (https://www.youtube.com/playlist?list=PLt9QR0WkC4WVszuogbmIIHIIQ2RMI78RC)

In [1]:

```
%%javascript
$.getScript('https://kmahelona.github.io/ipython_notebook_goodies/ipython_notebook_toc.
js')
```