Some Thoughts on Data Science

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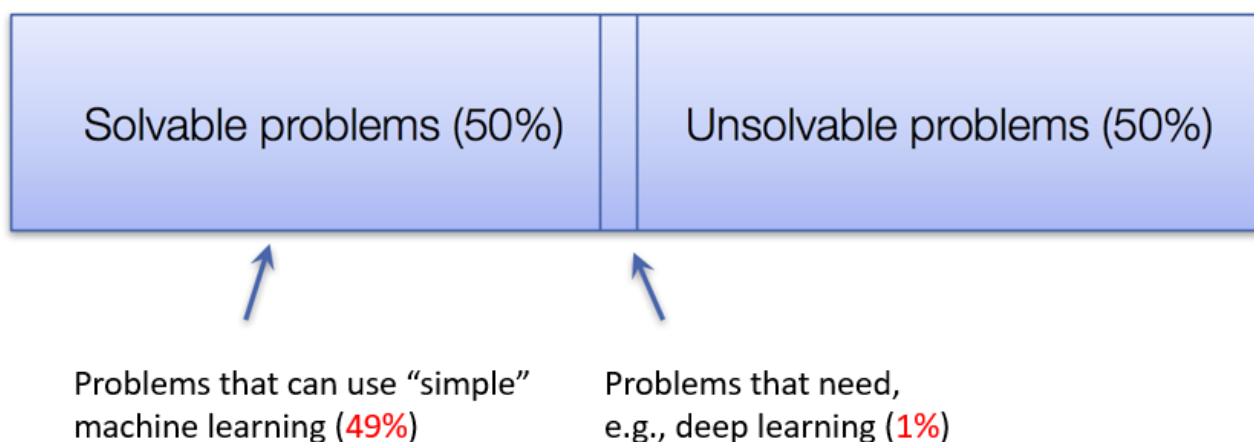http://isystems.unist.ac.kr/
POSTECH

Table of Contents

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

Solvable problems (50%)	Unsolvable problems (50%)

Problems that can use "simple" machine learning (49%)

Problems that need, e.g., deep learning (1%)

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)

Changing Input

- 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)

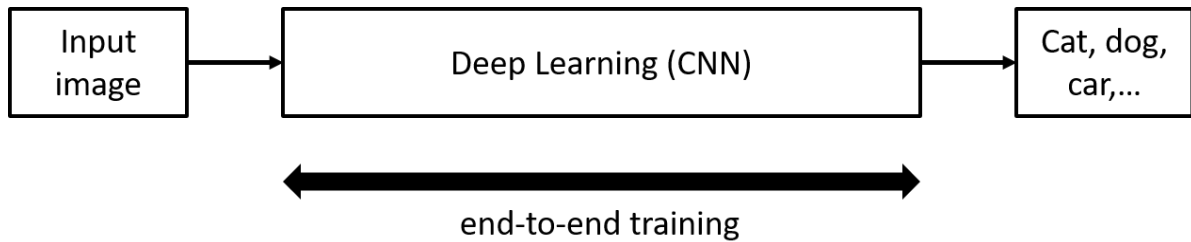**Changing Output (i.e., Changing the Problem)**

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

Machine Learning vs. Deep Learning

- State-of-the-art until 2012

| Input image | → | Hand-engineered features (SIFT, HOG, …) | → | Classifier | → | Cat, dog, car,… |

- Deep supervised learning

| Input image | → | Deep Learning (CNN) | → | Cat, dog, car,… |

← end-to-end training →

- Hyperparameters
    - Learning rate
    - # of iterations
    - # of hidden layers
    - # of hidden units
    - Choice of activation functions
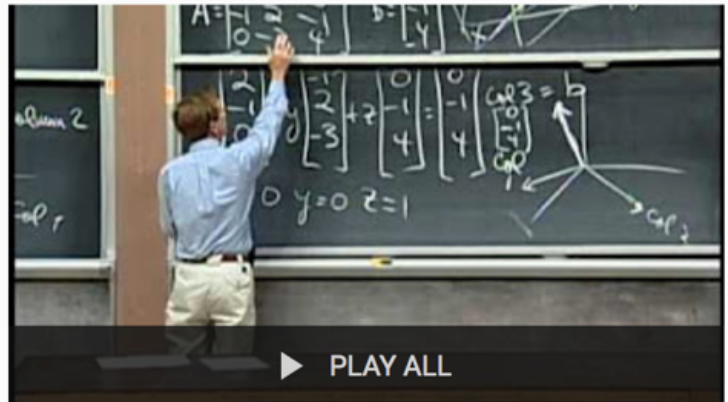
새로운 기술을 어디에 적용해 볼 수 있을지 고민하세요.
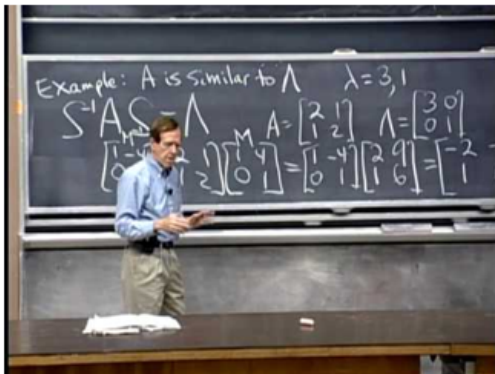
2. Study Materials

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

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)

Convex Optimization
Stephen Boyd and Lieven Vandenberghe
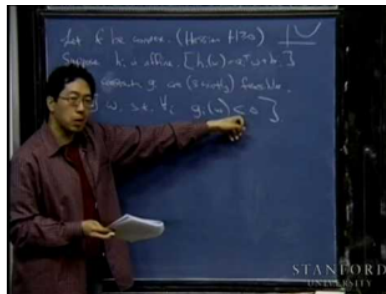
3) Machine Learning

- 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/)
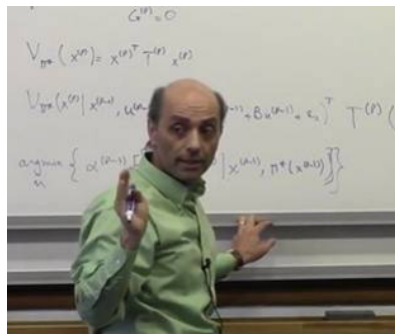

Zico Kolter

Lecture 4: Optimization 1

- 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/)
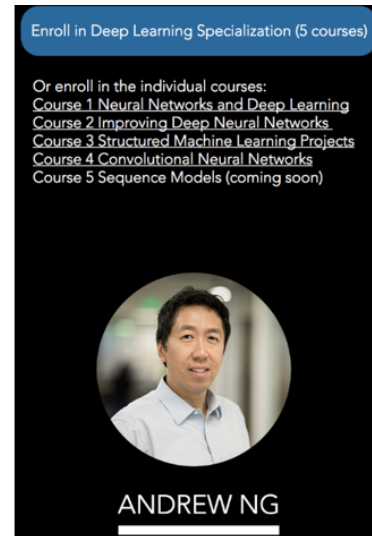

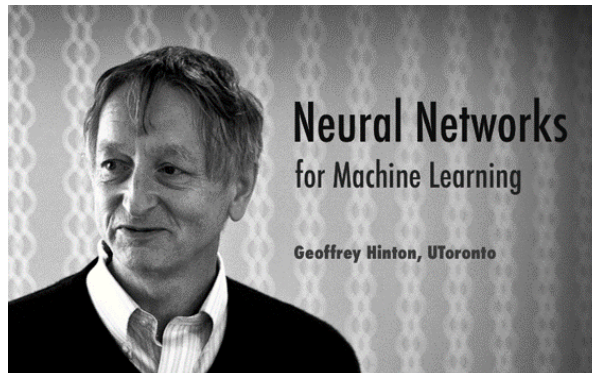- Learning from data
  - Prof. Yaser Abu-Mostafa from Caltech
  - https://work.caltech.edu/telecourse.html (https://work.caltech.edu/telecourse.html)

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#)

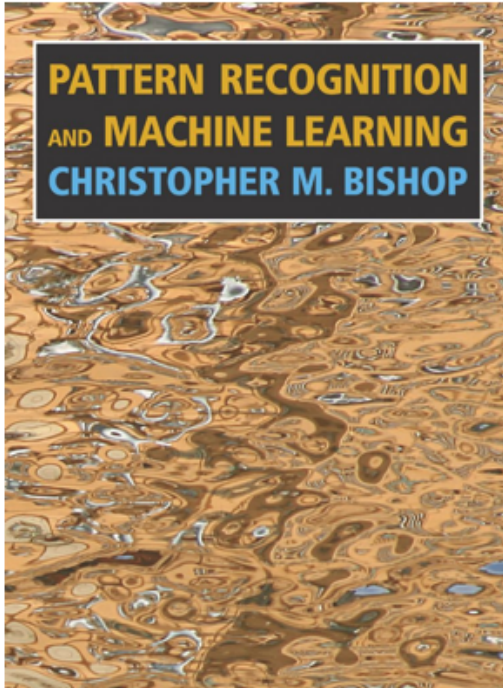**University Lectures on Deep Learning**

- 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/)

Christopher
Bishop

Laboratory Director, Microsoft
Research Cambridge

한국어 강좌

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

모두를 위한 머신러닝/딥러닝 강의

모두를 위한 머신러닝과 딥러닝의 강의

알파고와 이세돌의 경기를 보면서 이제 머신 러닝이 인간이 잘 한다고 여겨진 직관과 의사 결정능력에서도 충분한 데이타가 있으면 어느정도 또는 우리보다 더 잘할수도 있다는 생각을 많이 하게 되었습니다. Andrew Ng 교수님이 말씀하신것 처럼 이런 시대에 머신 러닝을 잘 이해하고 잘 다룰수 있다면 그야말로 "Super Power"를 가지게 되는 것이 아닌가 생각합니다.

더 많은 분들이 머신 러닝과 딥러닝에 대해 더 이해하고 본인들의 문제를 이 멋진 도구를 이용해서 풀수 있게 하기위해 비디오 강의를 준비하였습니다. 더 나아가 이론에만 그치지 않고 최근 구글이 공개한 머신러닝을 위한 오픈소스인 TensorFlow를 이용해서 이론을 구현해 볼수 있도록 하였습니다.

수학이나 컴퓨터 공학적인 지식이 없이도 쉽게 볼수 있도록 만들려고 노력하였습니다.

- 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_no
         tebook_toc.js')
```