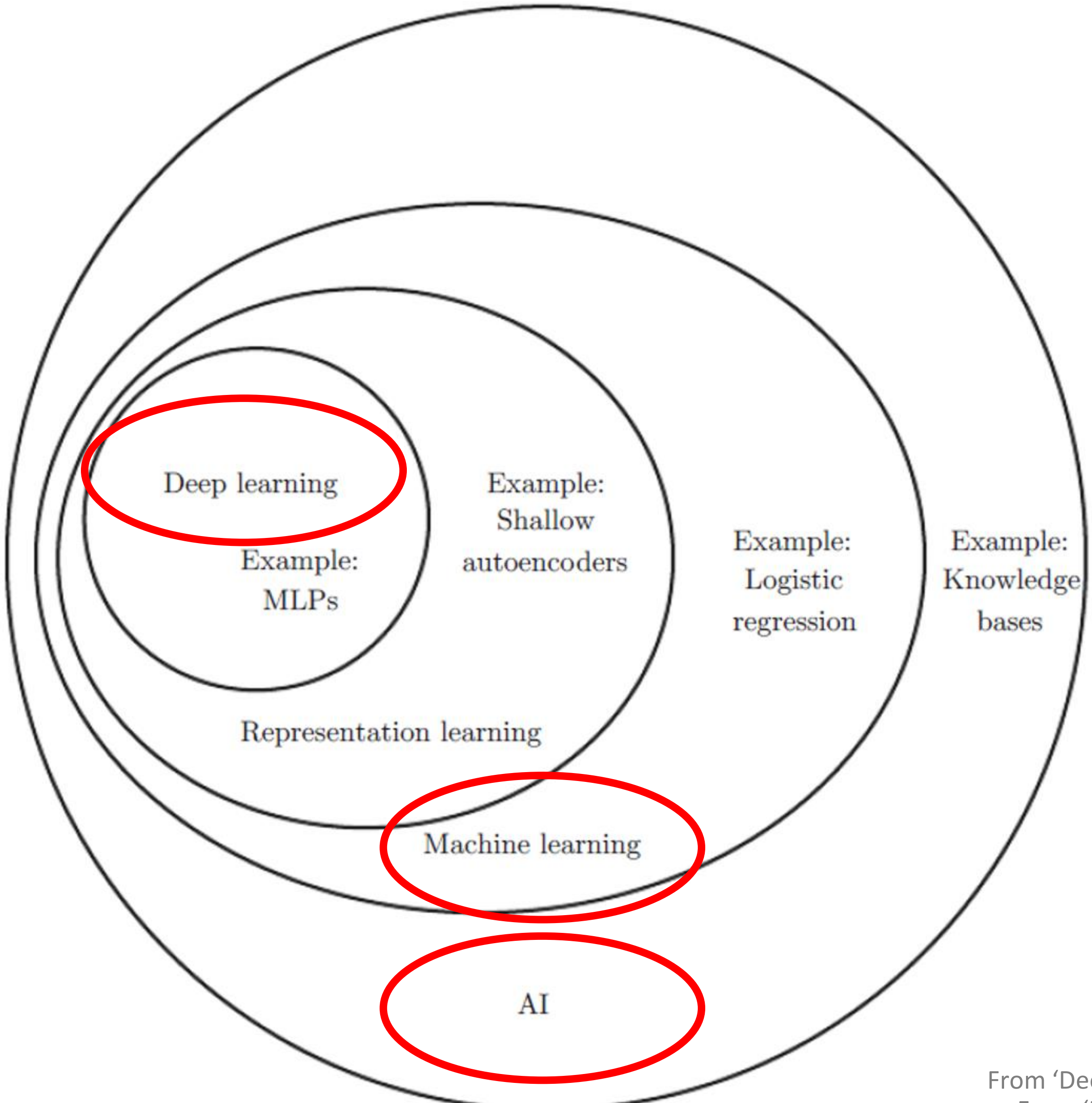
On the Current Status of AI Technology

2018.04.013

Seoul National University
Applied Data Science Lab.

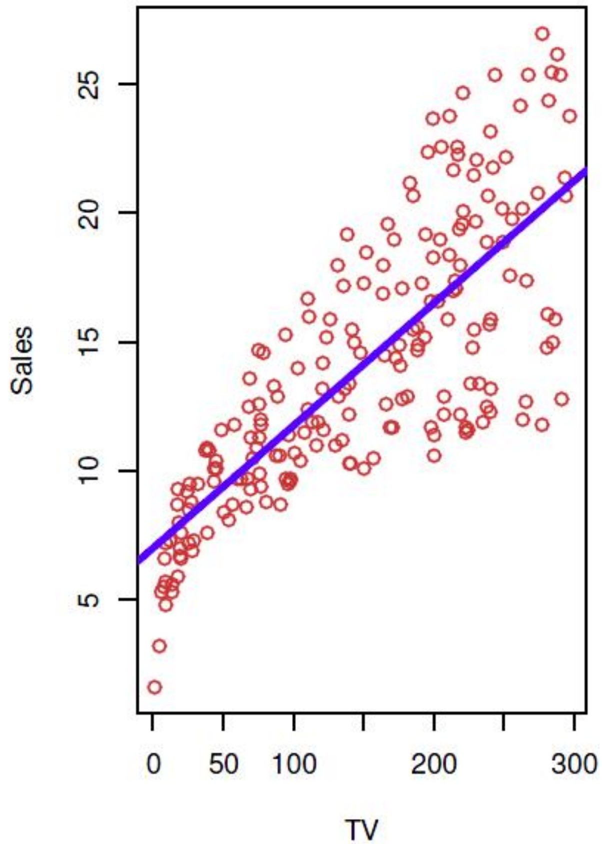
Prof. Wonjong Rhee, Ph.D., IEEE Fellow

AI, Machine Learning, Deep Learning



From 'Deep Learning' Book
From 'Deep Learning' Book

Linear Regression (선형회귀)



$$y = a + b \times x$$

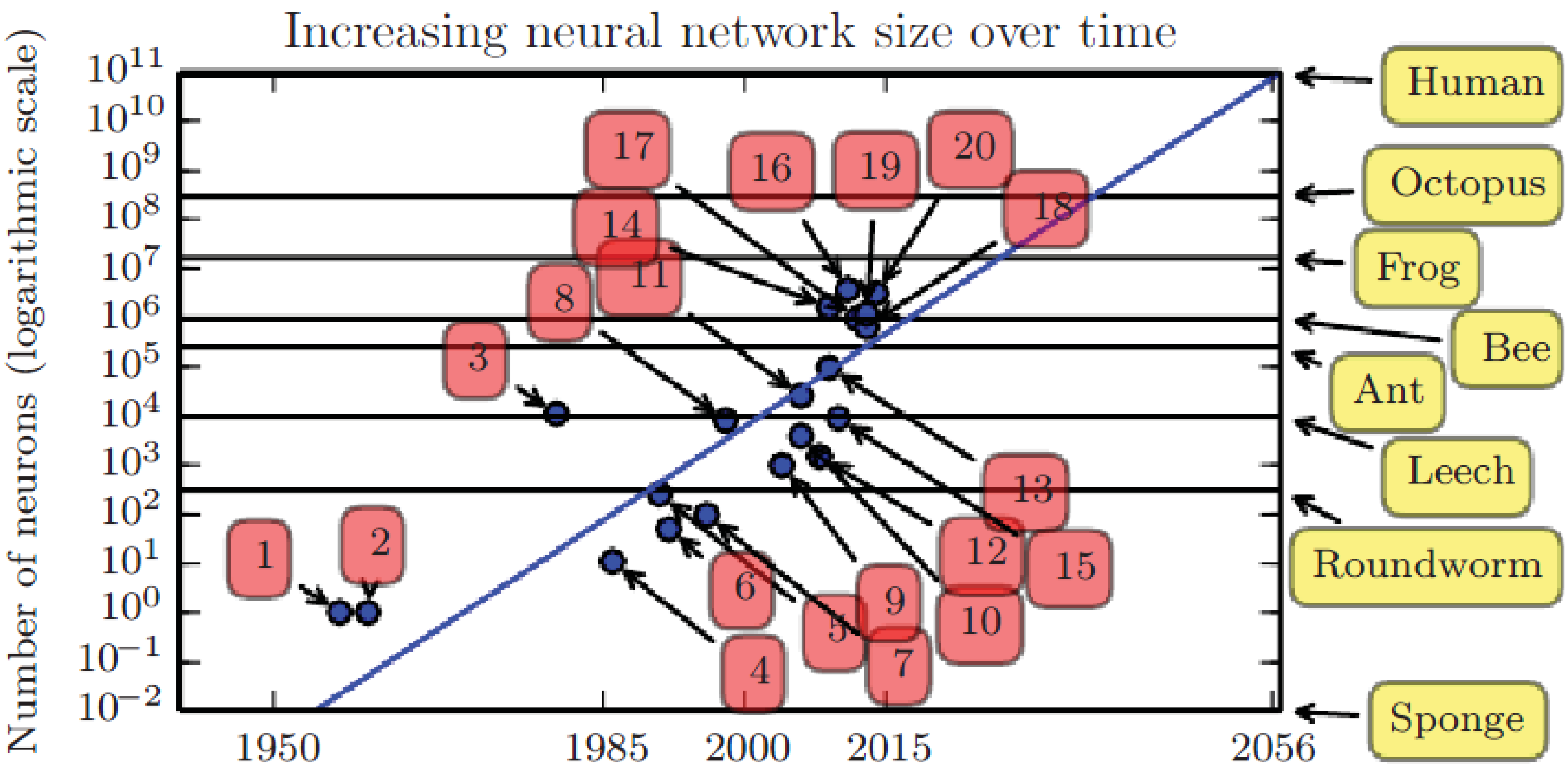
Deep Learning

A woman is throwing a frisbee in a park.

$$\cancel{y = a + b \times x}$$

Millions of parameters!
(& non-linearity)

Number of Parameters



From 'Deep Learning' Book

It’s Not Perfect

A woman holding a clock in her hand.

A man wearing a hat and
a hat on a skateboard.

A woman is sitting at a table
with a large pizza.

A man is talking on his cell phone
while another man watches.

This neural net can do only image captioning.
Fails often.

People in Financial Sector

- Searching for a golden algorithm?
- To beat all in the market?
- Deep Learning?

Contents

- NFL Theorem
- Feature Engineering
- Power of DL
- Current Status

NFL Theorem

No Free Lunch Theorem

IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 1, NO. 1, APRIL 1997 67

No Free Lunch Theorems for Optimization

David H. Wolpert and William G. Macready

***Abstract*—A framework is developed to explore the connection between effective optimization algorithms and the problems they are solving. A number of "no free lunch" (NFL) theorems are presented which establish that for any algorithm, any elevated performance over one class of problems is offset by performance over another class. These theorems result in a geometric interpretation of what it means for an algorithm to be well suited to an optimization problem. Applications of the NFL theorems to information-theoretic aspects of optimization and benchmark measures of performance are also presented. Other issues addressed include time-varying optimization problems and *a priori* "head-to-head" minimax distinctions between optimization algorithms, distinctions that result despite the NFL theorems' enforcing of a type of uniformity over all algorithms.**

***Index Terms*— Evolutionary algorithms, information theory, optimization.**

information theory and Bayesian analysis contribute to an understanding of these issues? How *a priori* generalizable are the performance results of a certain algorithm on a certain class of problems to its performance on other classes of problems? How should we even measure such generalization? How should we assess the performance of algorithms on problems so that we may programmatically compare those algorithms?

Broadly speaking, we take two approaches to these questions. First, we investigate what *a priori* restrictions there are on the performance of one or more algorithms as one runs over the set of all optimization problems. Our second approach is to instead focus on a particular problem and consider the effects of running over all algorithms. In the current paper

No free lunch theorems for optimization

DH Wolpert, WG Macready - IEEE transactions on evolutionary …, 1997 - ieeexplore.ieee.org

… Page 7. WOLPERT AND MACREADY: **NO FREE LUNCH** THEOREMS FOR OPTIMIZATION 73 we can choose an algorithm for which the calculation is tractable. **Theorem** 3: For any algorithm, the fraction of cost func- tions that result in a particular histogram is …

☆ 99 Cited by 6297 Related articles All 44 versions

A Simplified Example

- Coin toss game

  0 0 1 1 1 0 0 1 1 1 ?

  Next – 0 or 1?

- Algorithms
  - Alg1: Random
  - Alg2: If 3 consecutive 1's predict 1, otherwise random
  - Alg3: If 3 consecutive 1's predict 0, otherwise random
  - …

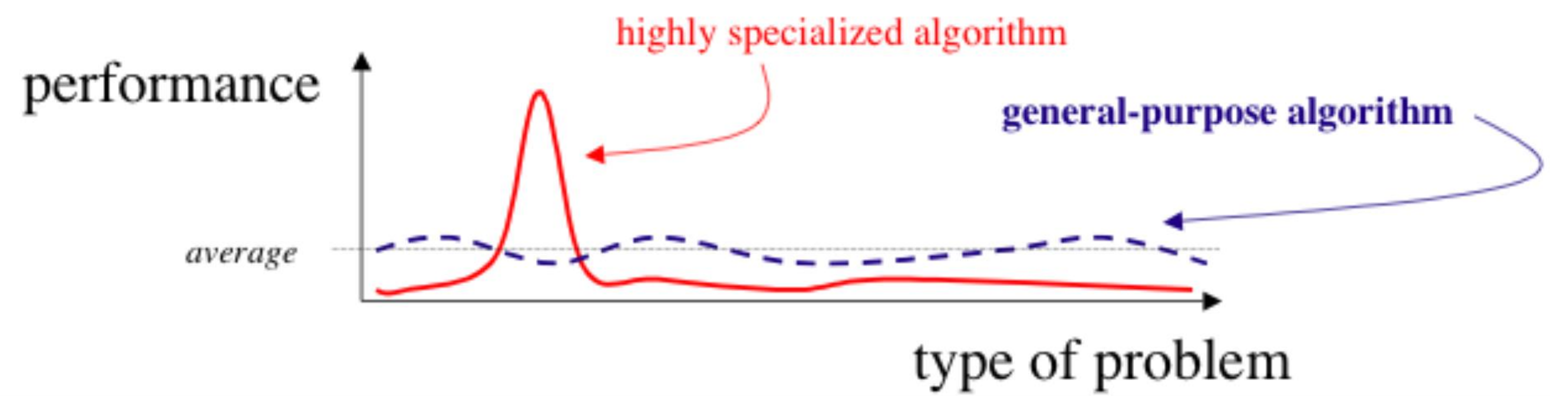
- Should one algorithm preferred over others?

No If All Are Equally Possible

0 0 0
0 0 1
0 1 0
0 1 1
1 0 0
1 0 1
1 1 0
1 1 1

- Given the first two outputs, predict the third
- All algorithms perform exactly the same on the average

The Teaching



- Domain knowledge can improve performance
- At the cost of generality

Figure from https://medium.com/@LeonFedden/the-no-free-lunch-theorem-62ae2c3ed10c

Knowledge + Data

- Use of knowledge in machine learning
  - (Deep) neural networks? Decision tree? Bayesian models?
  - Try many and choose one?
  - If too many tried over the same data, statistical learning theory teaches that overfitting will occur…
- The thing is… Machine learning can use only data.
  It is human's job to bring knowledge.
- Overcoming limitations of models

0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 1 1 1

➔ Count of '1 1 1'

Feature Engineering

Confidential and Proprietary – SNU ADS Lab.

Feature Extraction from Data

- Models are limited
  - Linear, SVM, NN, etc.
- Patterns and rules are very unique for each domain
- If one can extract the right features
  - Model does not need to be very smart

From Wiki

The process of feature engineering[6] [edit]

1. Brainstorming or Testing features;
2. Deciding what features to create;
3. Creating features;
4. Checking how the features work with your model;
5. Improving your features if needed;
6. Go back to brainstorming/creating more features until the work is done.

Feature Engineering

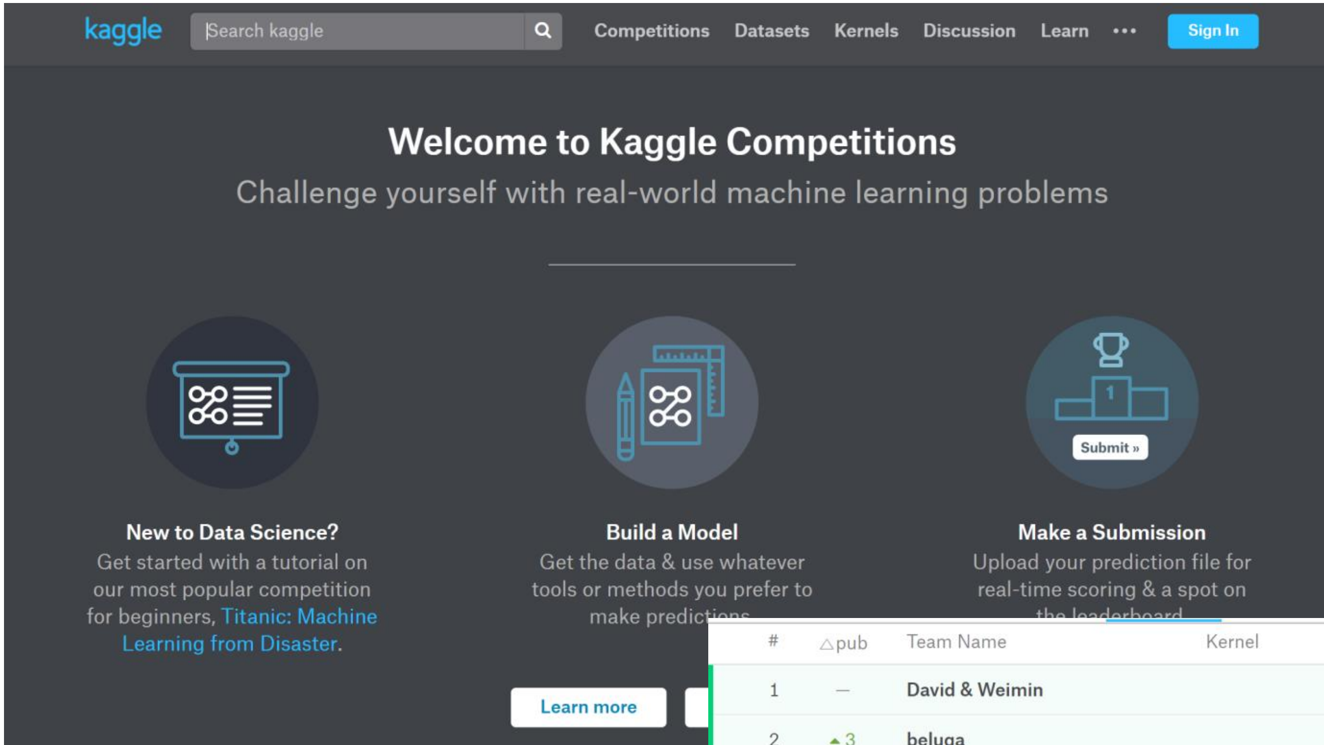
- Done by highly skilled people in each domain

- Often the deciding factor of the performance

- Are you a quant?

  What's your secret sauce?

kaggle Competition



#	△pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	David & Weimin			0.0822	118	3mo
2	▲3	beluga			0.0855	24	3mo
3	▲3	Evgeny Nekrasov			0.0857	66	3mo
4	—	Mark Rippetoe witnesses			0.0868	106	3mo
5	▼3	Kohei and Medrr			0.0888	114	3mo
6	▲3	AzAkhtyamov			0.0910	60	3mo
7	▲7	Juan Zhai 卷毛			0.0930	60	3mo
8	▲3	alijs			0.0981	15	3mo
9	▲529	Troy Retter			0.1046	5	3mo
10	▲29	ubik			0.1051	55	3mo
11	▲20	VictorHBD			0.1075	14	3mo

The Labor of Feature Engineering

- Become ‘skilled’
- Try over many data sets
- Repeat
- Deploy
- Monitor if the secret sauces are still working

Domain Experts

- Use one's best knowledge for extracting features (rules) from the data

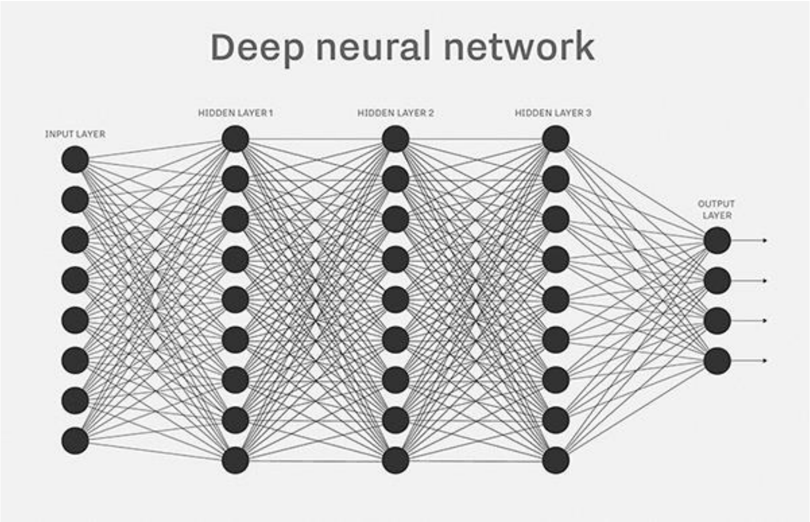
- Typically heuristics dominate
  - Any smart person can do it

- But, algorithms can be used, too
  - With proper training

Power of DL

Confidential and Proprietary – SNU ADS Lab.

Deep Learning



- It’s a model (deep version of neural network)
- Unlike other models, so powerful that it can model any function *f*(x)
- Considering its millions of parameters, it does not overfit as much as it should

- Overall effect

  ➔ Automatic feature engineering

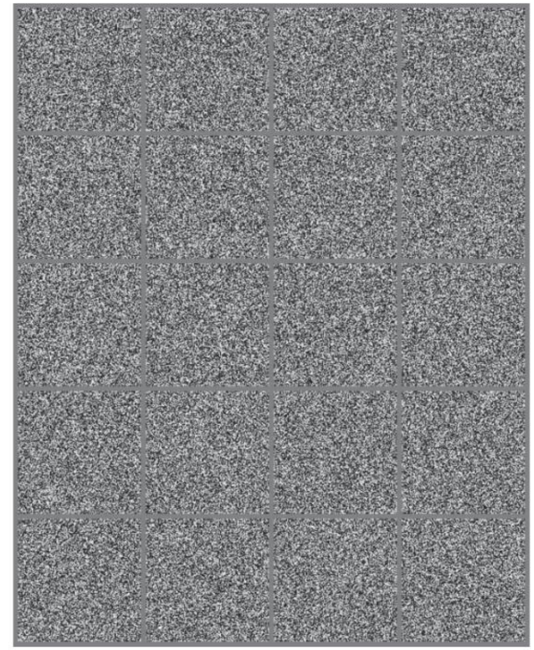
Deep Learning

- Very good at extracting features and encoding them
  - Called representation
- Automated feature engineering
  - Finally less labor
- Deep learning itself needs domain knowledge for tuning
  - Called hyper-parameter optimization
  - But less effort, less domain knowledge: your secret sauce might be defeated in a few days
- But, no guarantee that it will beat human
  - Depends on domain

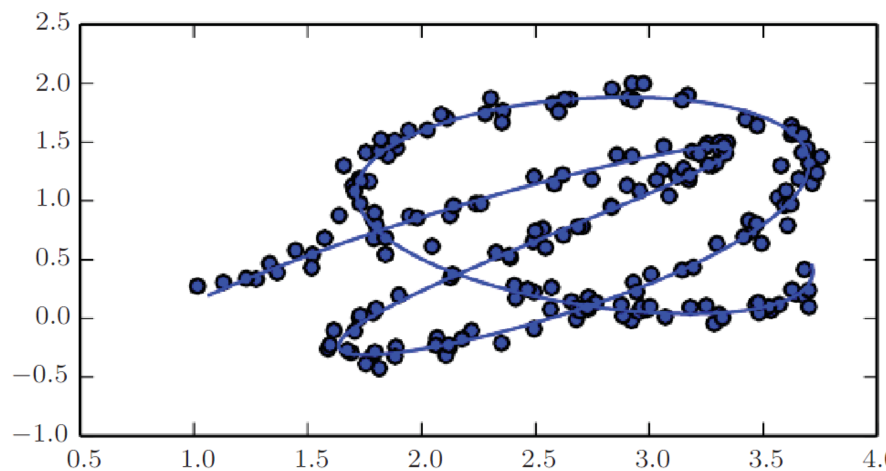
Representation Learning

- Truly random distribution example
  - Each pixel 256 possible values
  - 3 channels (Red, Green, Blue)
  - 1000^2 pixels

  ➔ $256^{3\times1,000,000}$ combinations



- Concept of "Manifold" in learning
  - Quite different from statistical noise problems



Yoshua Bengio
Aaron Courville

Manifold

- Some high dimensional problems have strong structure

  ➔ Deep learning will outperform any other

- Image, language

- Low dimensional and/or statistical noise problems

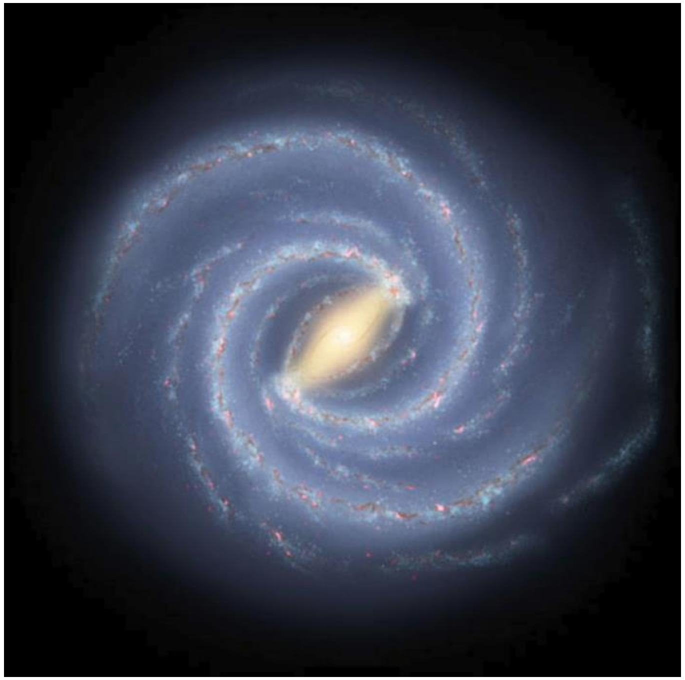
  ➔ No reason for DL to outperform

- What's your problem like?

Current Status

Today’s Research

- If you wait enough time, all algorithm engineers will be working on deep learning
- As of today, black hole of research
- (AI winter again?)



When So Many Smart People Work on DL

- No matter how hard you study, technology changes faster than you
- Many new tricks
- The new standard
  - If a trick is found, look for domains/applications

Back to Finance

- Massive data

- Stock market
  - Price is a reflection of all people's trades

  - Many random variables added → Gaussian distribution?

    OR

  - Can it have a strong structure as in image or language?

Back to Finance

- Active investing is not the only problem
- Where should we apply the new tricks?
- My view
  - There will be many successful cases
  - Often small improvements over what people have already done well, but big improvements, too
  - Will take some time for human (not machine) to figure out where to apply what trick
- It's an exciting time

  (until the next AI winter?)

Thank you!

Confidential and Proprietary – SNU ADS Lab.