



# 機械学習による化学反応の予測と設計

瀧川 一学

<https://itakigawa.github.io/>

このスライドはここにあります

<https://itakigawa.github.io/data/jfssa2020.pdf>

- 理化学研究所 革新知能統合研究センター@京阪奈  
iPS細胞連携医学的リスク回避チーム
- 北海道大学 化学反応創成研究拠点 (WPI-ICReDD)



革新知能統合研究センター  
Center for Advanced Intelligence Project

# 今日の話

## 1. 機械学習+化学の面白さとチャレンジ

- 😊 CVとNLPの間の独特設定 + 第一原理がある + 実検証できる
- 🤔 表現学習 + 小サンプル + 強いデータバイアス

## 2. 機械学習の関連トピックの(面白さの)紹介

### 1. 探索と訓練分布外の予測

新たなデータを取る探索戦略 (実験計画・最適化・進化計算)  
外挿・内挿問題、良性の過適合

### 2. 学習の構成性と転移性

注意機構と転移学習・構成的学習

小サンプル学習 (転移学習、メタ学習、計量学習)

### 3. 演繹と帰納の融合

モデルベースの最適化・強化学習、第一原理計算との融合

# 今日の話

## 1. 機械学習+化学の面白さとチャレンジ

- 😊 CVとNLPの間の独特設定 + 第一原理がある + 実検証できる
- 🤔 表現学習 + 小サンプル + 強いデータバイアス

## 2. 機械学習の関連トピックの(面白さの)紹介

### 1. 探索と訓練分布外の予測

新たなデータを取る探索戦略 (実験計画・最適化・進化計算)  
外挿・内挿問題、良性の過適合

### 2. 学習の構成性と転移性

注意機構と転移学習・構成的学習

小サンプル学習 (転移学習、メタ学習、計量学習)

### 3. 演繹と帰納の融合

モデルベースの最適化・強化学習、第一原理計算との融合

# 化学

ある物質を別の物質に

物質A → 物質B

このプロセス  
(目にみえない)

化学

少なくとも分子生物学の  
根幹にある思想

あらゆるものは「化学反応」

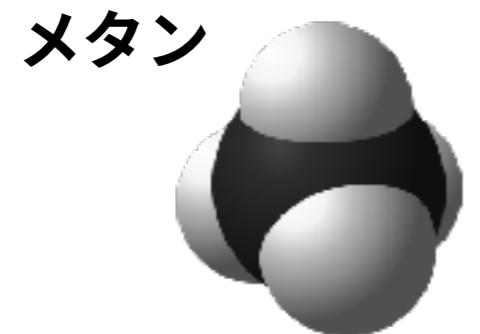
- 身の回りのあらゆる製品の製造
- エネルギーをつくる
- 明るい都市生活
- 農業(土壤改良)
- 環境汚染
- 生命現象
  - 食べ物が美味しい(食・栄養)
  - お肌の調子が良い(美容・健康)
  - ねむい(睡眠・意識・麻酔)
  - 頭がいたい、お腹がいたい(医薬)
  - 病む(疾患)
  - 老いる(老化・長寿)
  - 有限の生(生死・発生・分化)

# もし化学反応が自由自在に設計・制御できたなら...

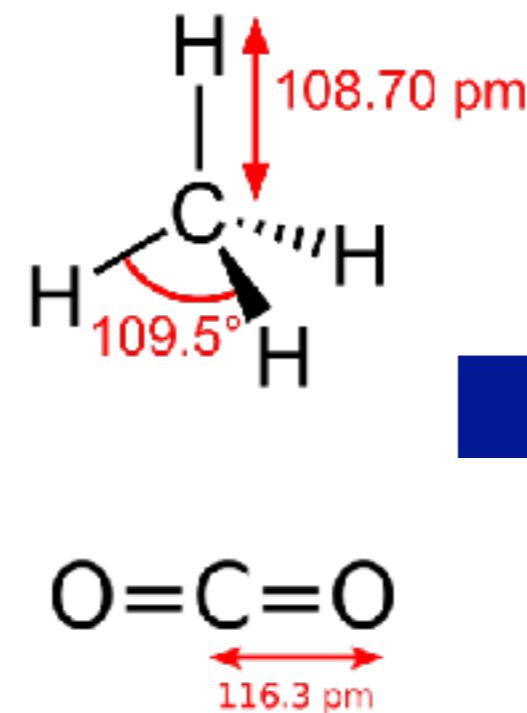
その辺にたくさんある分子

天然ガス  
シェールガス  
メタンハイド  
レート

温暖化



何らかの方法で  
バラバラに...  
何らかの方法で  
組み立てて...



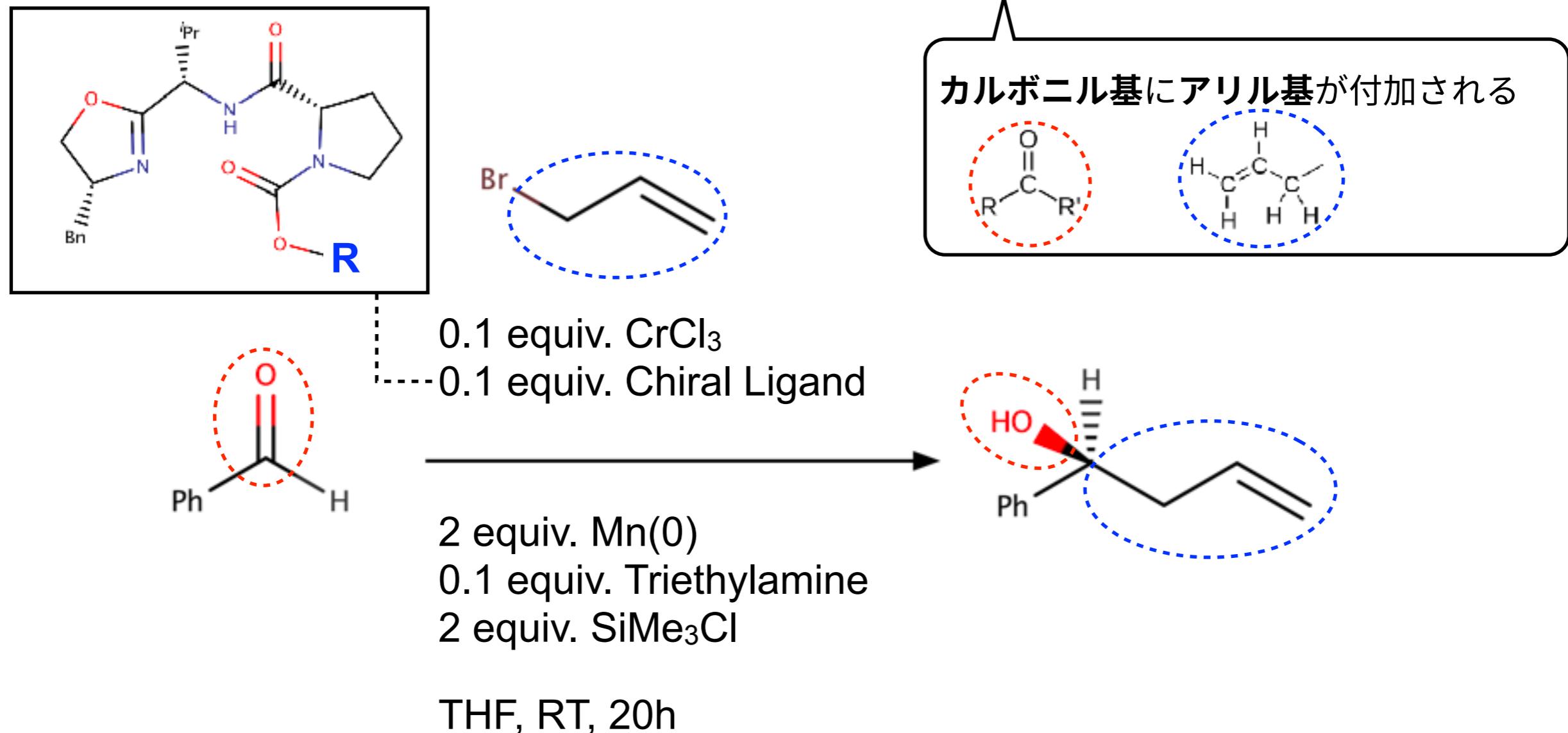
欲しい分子  
欲しい機能

エネルギー  
医薬品  
化学材料  
電子材料  
発光  
農薬  
食品  
化粧品

- 身の回りのものやエネルギーを作るには**化学反応**が必要
- 物理法則の壁**により上のように自由自在な組み替えはできない
  - ①に高いエネルギーか画期的触媒か何かが必要
  - 無理やり①を起こす状況では②を精密に制御にするのが困難  
(C-Hは不活性だし同じC-H結合の特定のものだけを反応させるのも難)

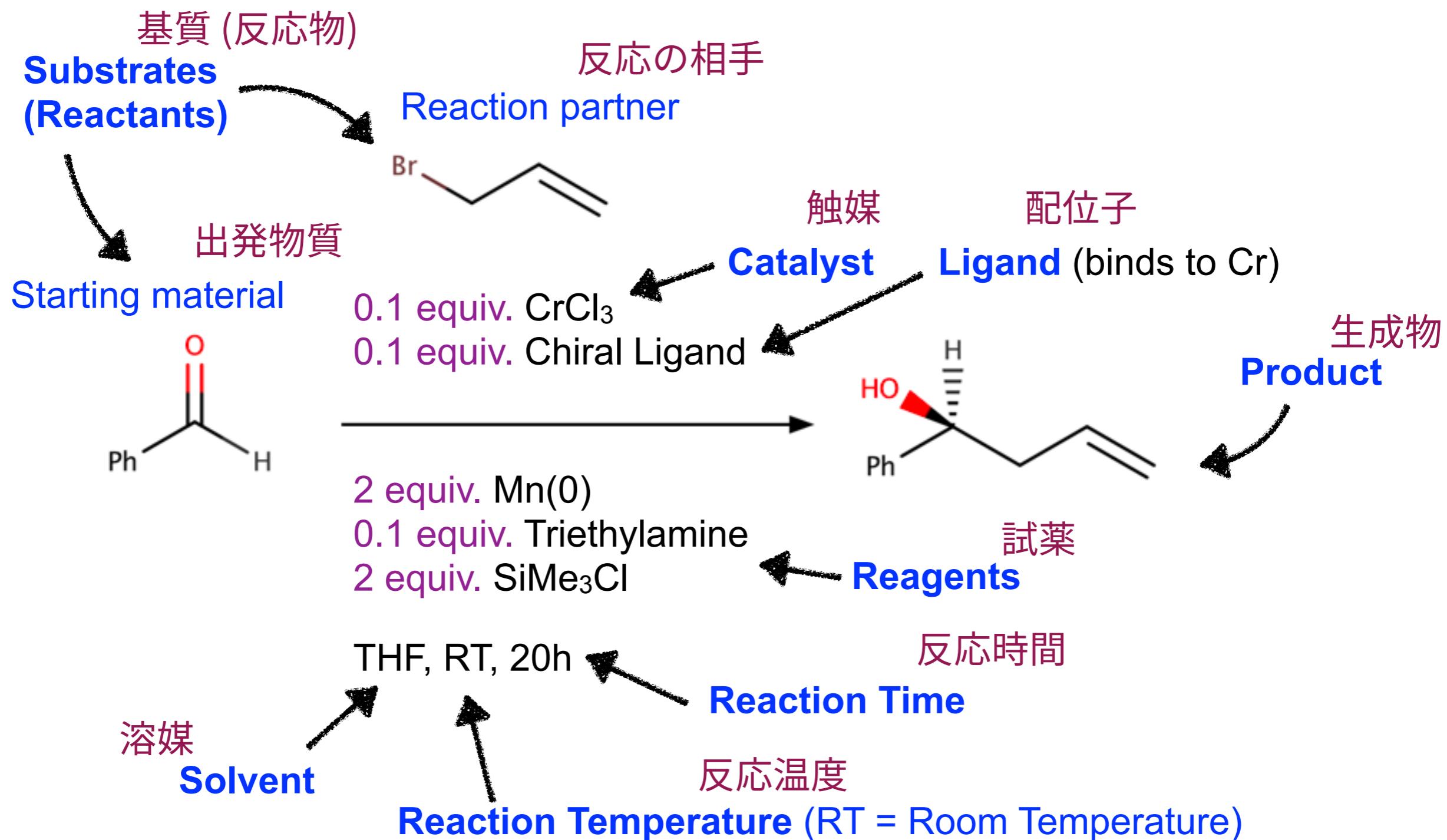
言うは易く行うは難し：不斉合成の一例

## Nozaki-Hiyama-Kishi (NHK) allylation reaction



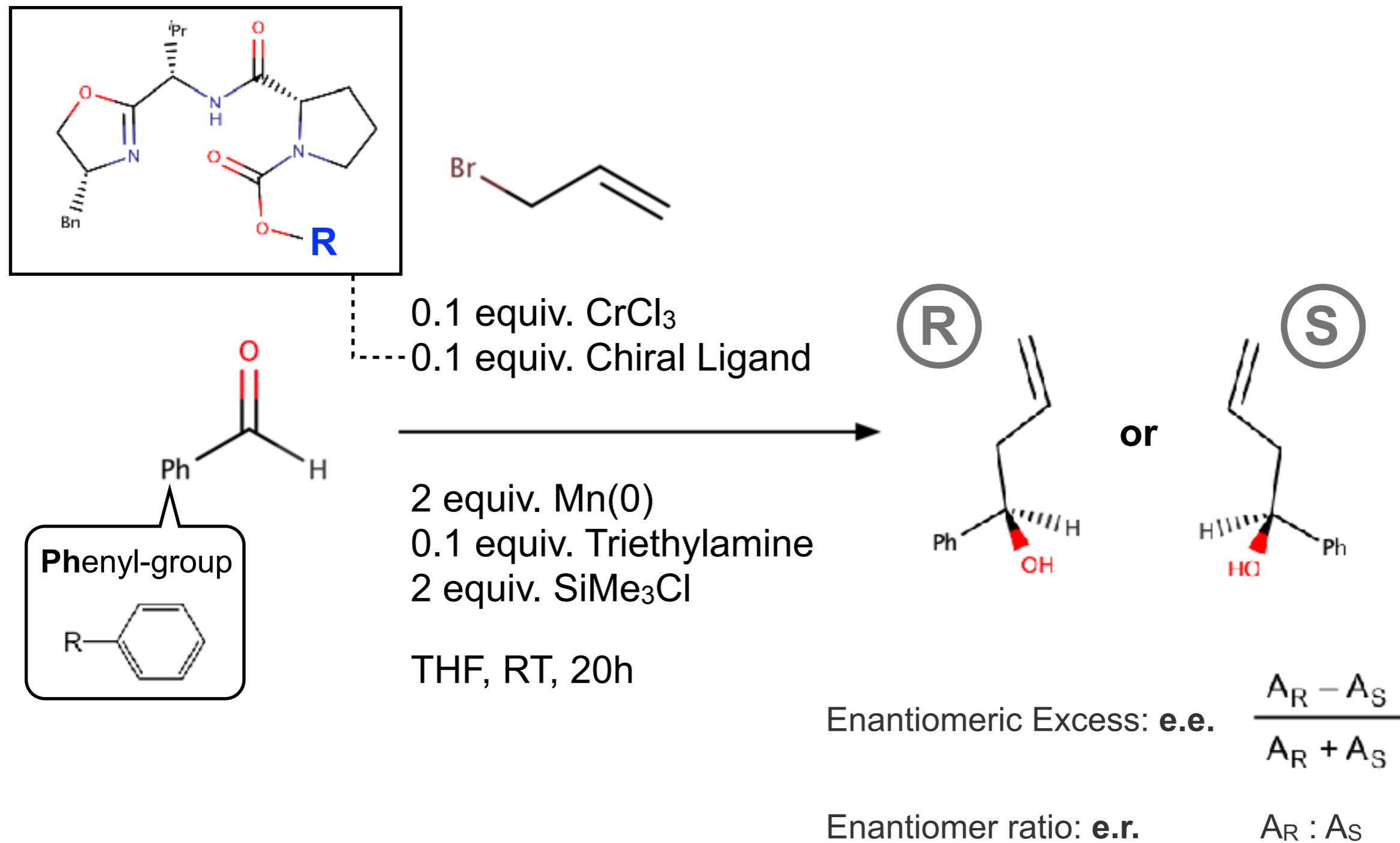
実際に反応を起こすには多数の因子が関与する...

## Nozaki-Hiyama-Kishi (NHK) allylation reaction



立体異性体(R or S)のどちらかだけを選択的に合成したい

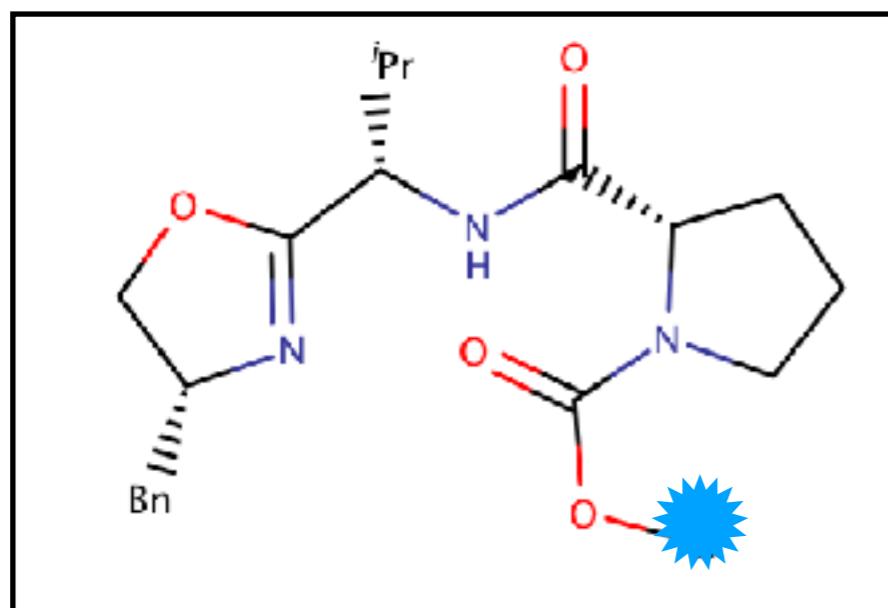
## Nozaki-Hiyama-Kishi (NHK) allylation reaction



# 配位子のいろいろな箇所で置換基 を変えてみたり

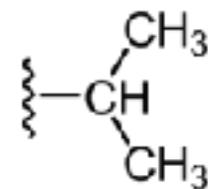
触媒にくっついて実際に仕事をする"配位子"を適切に改変することで高い選択性を目指す

## "Ligands" (配位子)



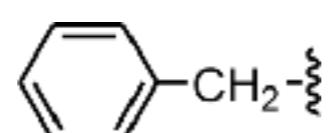
*i*Pr

= isopropyl group



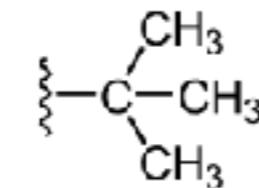
Bn

= benzyl group



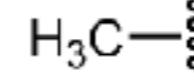
tBu

= tertiary butyl group



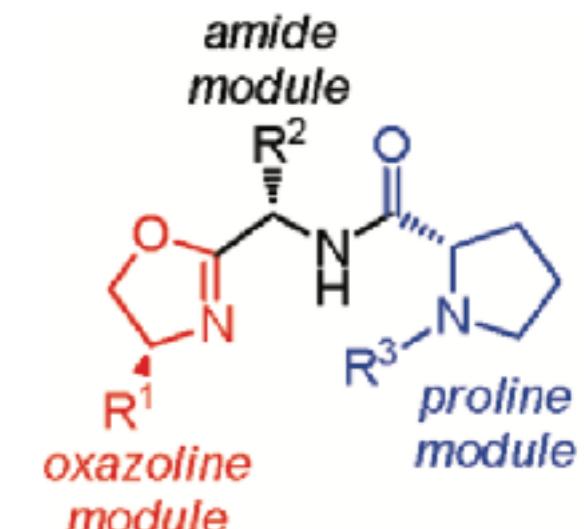
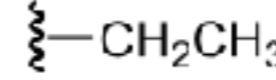
Me

= methyl group



Et

= ethyl group



に入る  
置換基の候補

- Ph
- tBu
- CF<sub>3</sub>
- CH(CH<sub>3</sub>)<sub>2</sub>
- O*i*Pr
- Cl
- CH<sub>3</sub>
- CHO
- CO<sub>2</sub>Me
- COMe
- COOH
- OEt
- NMe<sub>2</sub>
- OMe
- NH<sub>2</sub>
- NO<sub>2</sub>
- OH
- F
- H

# 機械学習+化学の面白さとチャレンジ

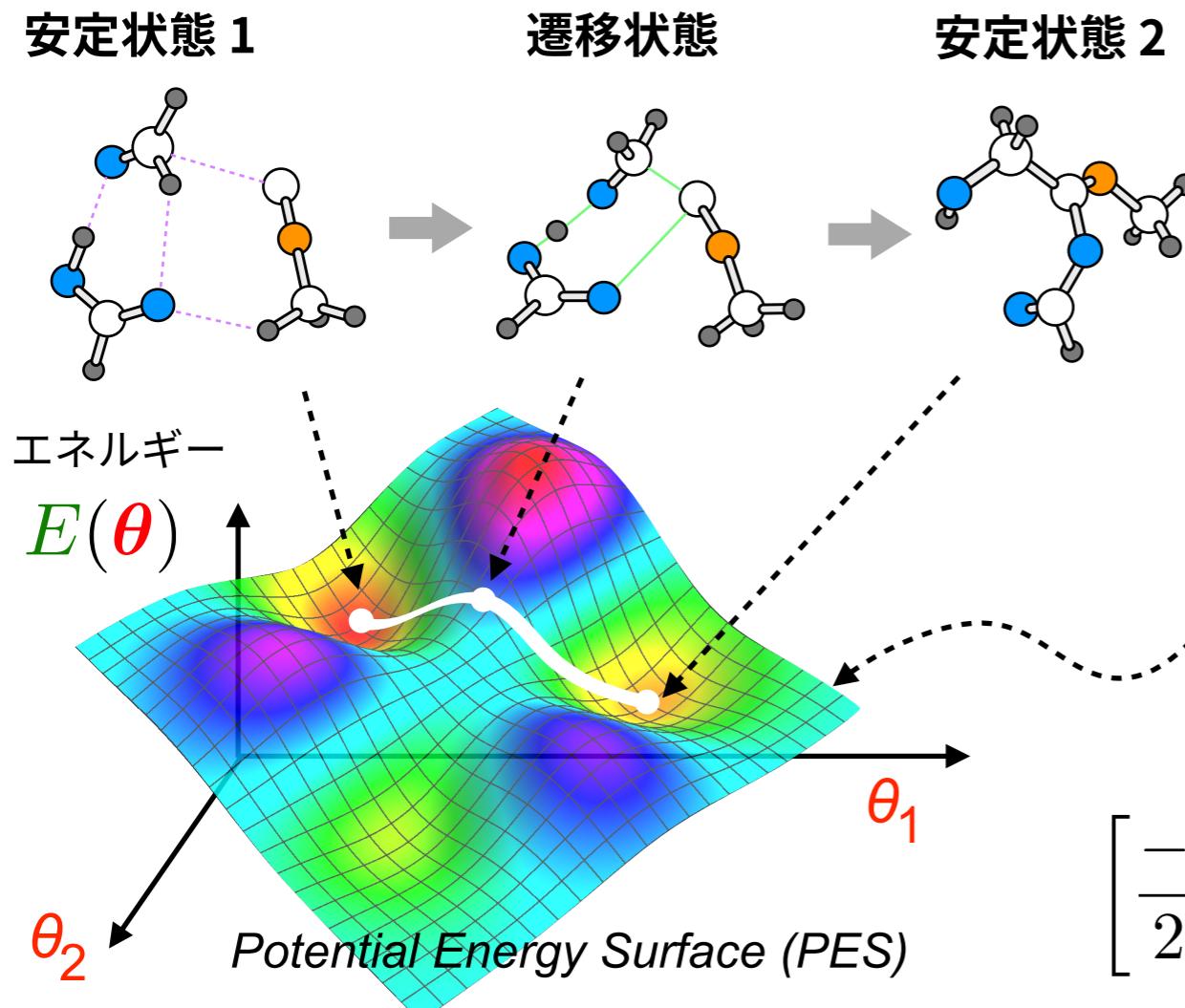
## 😊 面白さ

- CVとNLPの間の独特設定
- 第一原理がある
- 実検証できる

## 🤔 チャレンジ

- 表現学習
- 小サンプル
- 強いデータバイアス

# 機械学習+化学の面白さ



- 入口出口は"離散組合せ的"対象

化合物 = 高々100数種類の元素  
(現在118種)がなす膨大な組合せ

- 化学反応は(自然法則が定める)  
原子や結合の組み替え過程

エネルギーの谷→峠→谷の遷移だが  
このエネルギー面  $E(\theta)$  を求めるには  
毎点で電子に関する方程式の求解が必要

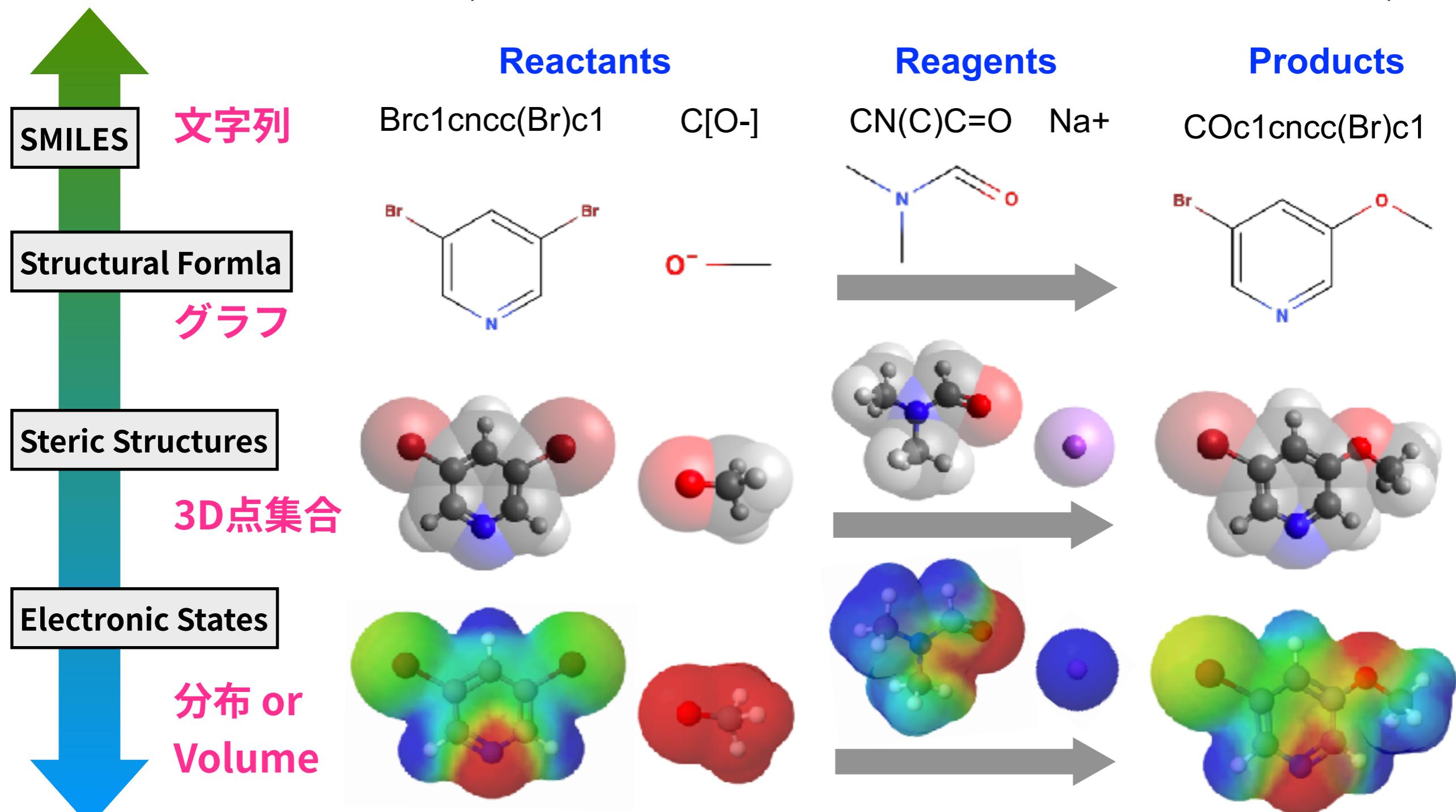
$$\left[ \frac{-\hbar^2}{2m} \nabla^2 + V(\theta) \right] \Psi = E(\theta) \Psi$$

*Schrödinger equation*

- 探索空間が組合せ的に巨大: 理論的に可能な経路の探索も組合せ爆発を起こす
- 計算時間・リソースが大きく計算できる系が限られる: 現実の系では何か妥協が必要
- 現実の化学反応の複雑さと不確定さ: 理論計算に入らない多様な要因が影響
- 現状の理論モデルの単純な仮定や不完全さ: 実験での反応と理論との大きな隔たり

# チャレンジ① 表現学習

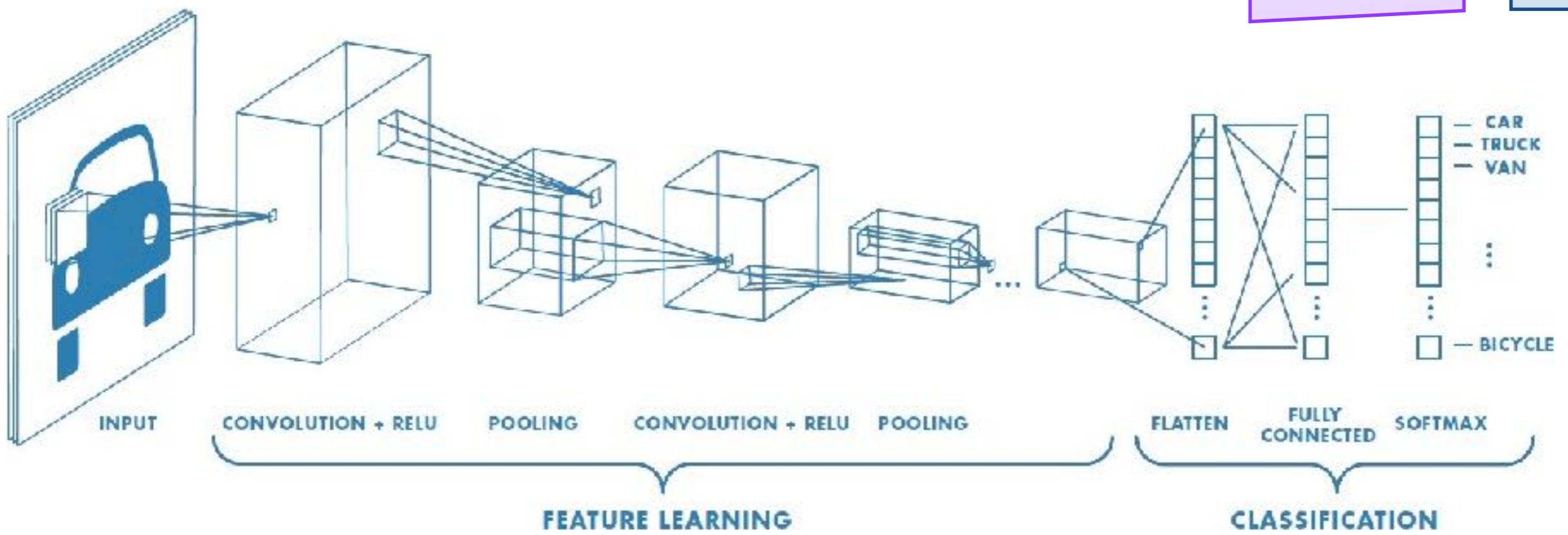
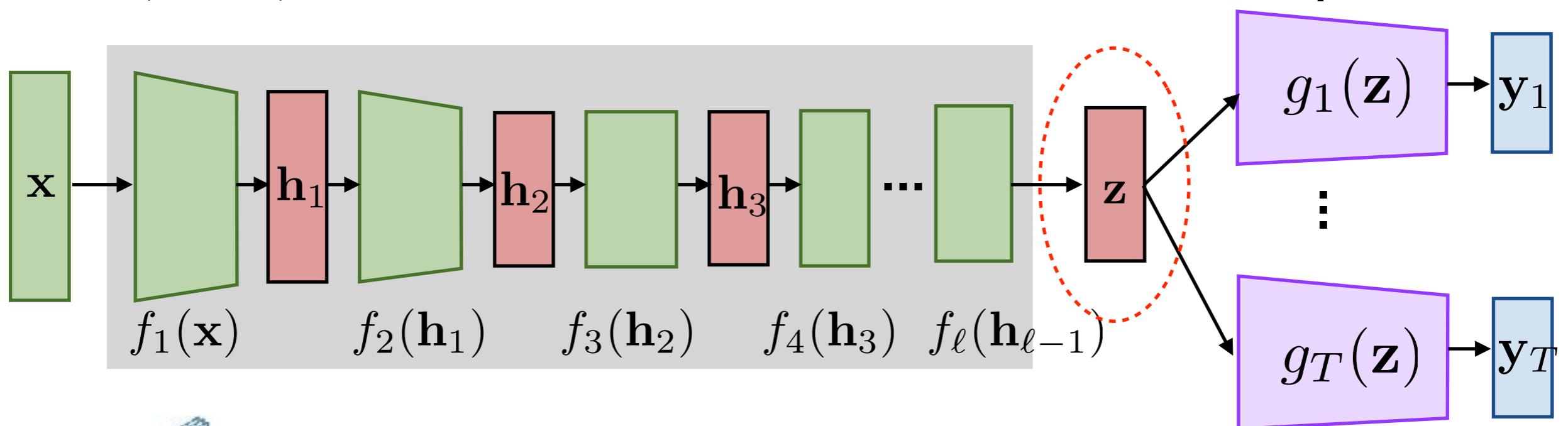
パターン言語として (化学の教科書・データベースにある知識表現)



物理的対象として (量子化学に基づく電子状態計算)

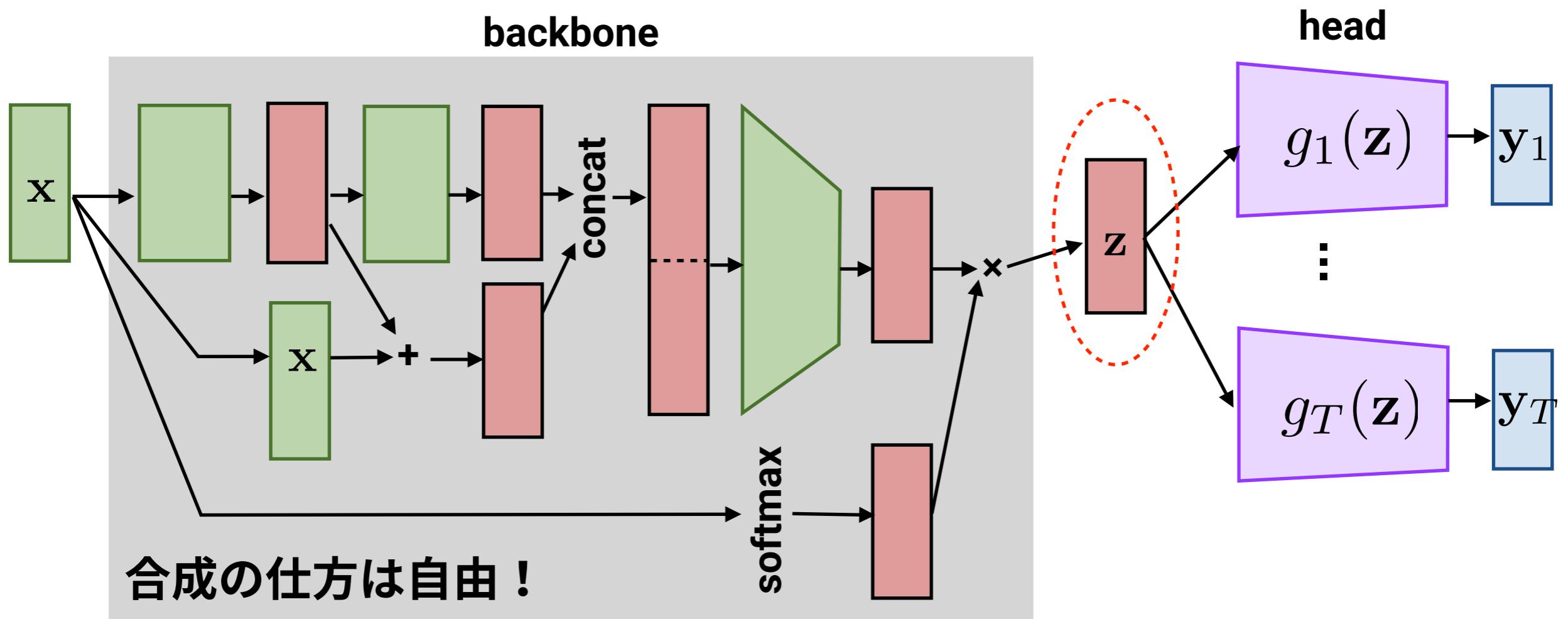
# データからの表現学習

有効な(潜在的)特徴表現 $z$ への変換を学習するbackbone



# 関数の合成に対し(勾配降下で)パラメタの統一的学习が可能

合成関数を計算グラフで保持して、勾配は自動微分(aka backprop)で計算



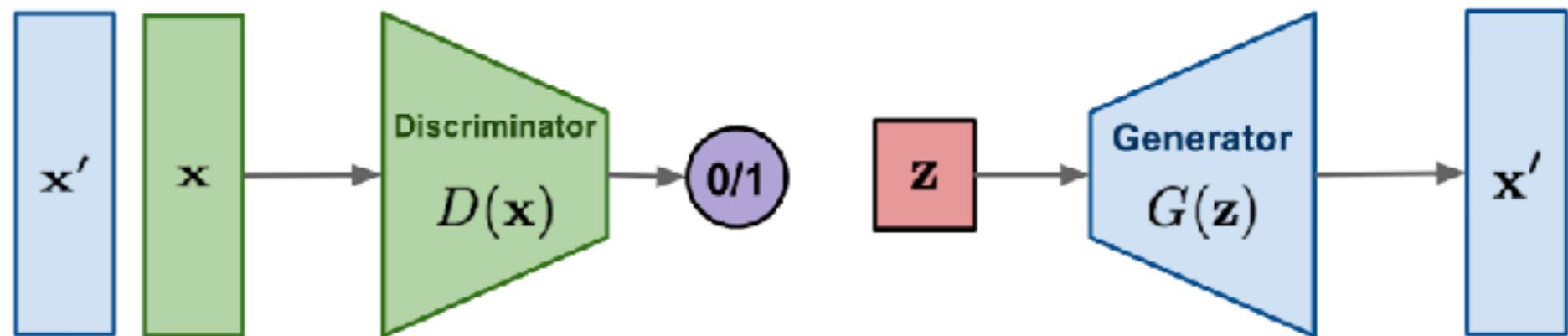
"All numerical computations are ultimately **compositions of a finite set of elementary operations** for which derivatives are known (Verma, 2000; Griewank and Walther, 2008)"

*Automatic differentiation in machine learning: a survey*, JMLR 18 (2018) 1-43

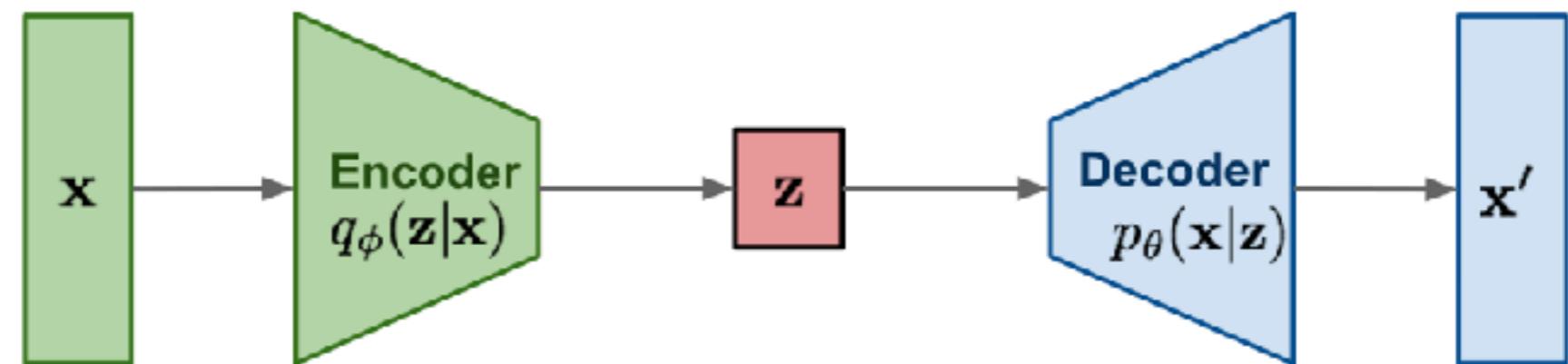
# データからの表現学習

例えば"有効"な表現への変換を学習する点は生成モデルでも同じ

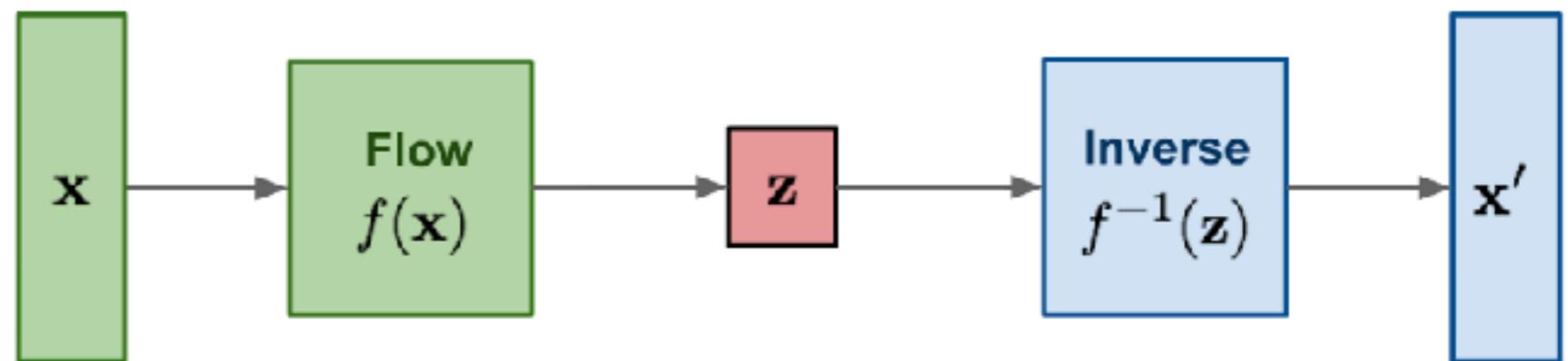
**GAN:** minimax the classification error loss.



**VAE:** maximize ELBO.



**Flow-based generative models:** minimize the negative log-likelihood



# チャレンジ② 小サンプル・強いバイアス

人間が「データセット」作るとどうしても偏ってしまう...

12 SEPTEMBER 2019 | VOL 573 | NATURE | 251

nature

## Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis

Xiwen Jia<sup>1</sup>, Allyson Lynch<sup>1</sup>, Yuheng Huang<sup>1</sup>, Matthew Danielson<sup>1</sup>, Immaculate Lang'at<sup>1</sup>, Alexander Milder<sup>1</sup>, Aaron E. Ruby<sup>1</sup>, Hao Wang<sup>1</sup>, Sorelle A. Friedler<sup>2\*</sup>, Alexander J. Norquist<sup>1,\*</sup> & Joshua Schrier<sup>1,3\*</sup>

Most chemical experiments are planned by human scientists and therefore are subject to a variety of human cognitive biases<sup>1</sup>, heuristics<sup>2</sup> and social influences<sup>3</sup>. These anthropogenic chemical reaction data are widely used to train machine-learning models<sup>4</sup> that are used to predict organic<sup>5</sup> and inorganic<sup>6,7</sup> syntheses. However, it is known that societal biases are encoded in datasets and are perpetuated in machine-learning models<sup>8</sup>. Here we identify as-yet-unacknowledged anthropogenic biases in both the reagent choices and reaction conditions of chemical reaction datasets using a combination of data mining and experiments. We find that the amine choices in the reported crystal structures of hydrothermal synthesis of amine-templated metal oxides<sup>9</sup> follow a power-law distribution in which 17% of amine reactants occur in 79% of reported compounds, consistent with distributions in social influence models<sup>10–12</sup>. An analysis of unpublished historical laboratory notebook records shows similarly biased distributions of reaction condition choices. By performing 548 randomly generated experiments,

that resemble those the medicinal chemists involved have synthesized in the past<sup>19,20</sup>, which use a limited set of reactions<sup>21</sup>, the choice of which is uncorrelated to cost, estimated ease of the synthesis, or the properties of the reactants and products<sup>22</sup>. However, over-representation of a particular experimental choice need not be irrational. For example, 36% of entries in the Protein Data Bank (PDB) report the use of polyethylene glycol additives, which under-represents the true success rate of 59%, and many of these proteins cannot be crystallized using other additives<sup>23</sup>. This suggests that a lack of diversity among crystallization additives in the PDB stems from sub-optimal novelty seeking. Excessively consistent or inconsistent experimental choices that do not mimic the natural distribution of the underlying problem are a signature of anthropogenic influence.

We seek to determine whether there is evidence of bias in reactant choices for organically templated metal oxide syntheses. The incorporation of different organic amines results in compounds with diverse composition, local and extended connectivity, and functionality<sup>9</sup>, so

# データそのものや解析に落とし穴がたくさん...

REPORT

## Predicting reaction performance in C–N cross-coupling using machine learning

Science

Derek T. Ahneman<sup>1</sup>, Jesús G. Estrada<sup>1</sup>, Shishi Lin<sup>2</sup>, Spencer D. Dreher<sup>2,\*</sup>, Abigail G. Doyle<sup>1,\*</sup>

+ See all authors and affiliations

Science 13 Apr 2018:  
Vol. 360, Issue 6385, pp. 186-190  
DOI: 10.1126/science.aar5169

- Main paper <https://doi.org/10.1126/science.aar5169>
- Erratum <https://doi.org/10.1126/science.aat7648>
- Negative comment paper <https://doi.org/10.1126/science.aat860>
- Author's response <https://doi.org/10.1126/science.aat8763>

## Dispute over reaction prediction puts machine learning's pitfalls in spotlight



BY KATRINA KRÄMER | 18 DECEMBER 2018

<https://www.chemistryworld.com/news/dispute-over-reaction-prediction-puts-machine-learnings-pitfalls-in-spotlight/3009912.article>

# バイアスで片付けて良いのか: PULSEと"AI bias"炎上事件

Face Depixelizer "PULSE"についてのコメント炎上でLeCun Twitter停止...

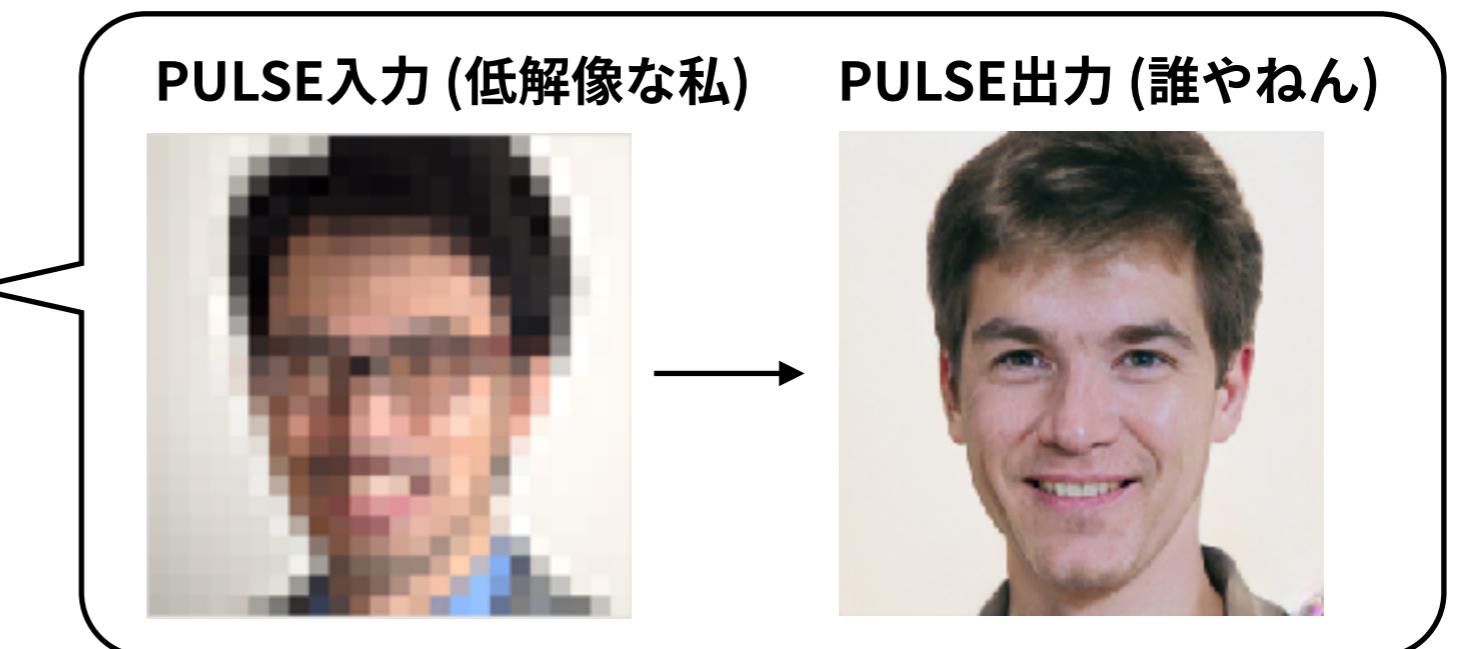
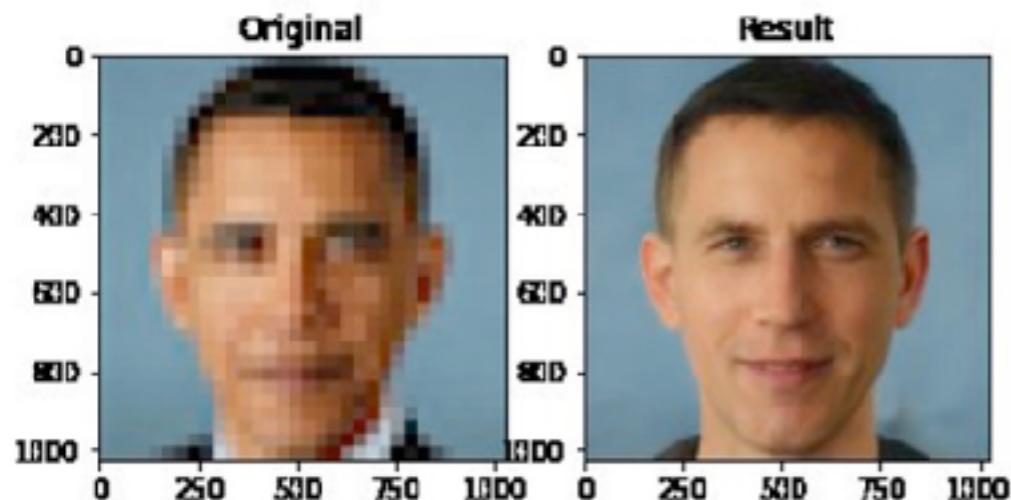
REPORT \ TECH \ ARTIFICIAL INTELLIGENCE

THE VERGE

## What a machine learning tool that turns Obama white can (and can't) tell us about AI bias

A striking image that only hints at a much bigger problem

By James Vincent | Jun 23, 2020, 3:45pm EDT



<https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias>

# 今日の話

## 1. 機械学習+化学の面白さとチャレンジ

- 😊 CVとNLPの間の独特設定 + 第一原理がある + 実検証できる
- 🤔 表現学習 + 小サンプル + 強いデータバイアス

## 2. 機械学習の関連トピックの(面白さの)紹介

### 1. 探索と訓練分布外の予測

新たなデータを取る探索戦略 (実験計画・最適化・進化計算)  
外挿・内挿問題、良性の過適合

### 2. 学習の構成性と転移性

注意機構と転移学習・構成的学習

小サンプル学習 (転移学習、メタ学習、計量学習)

### 3. 演繹と帰納の融合

モデルベースの最適化・強化学習、第一原理計算との融合

# 予測から発見へ：学習の「利用」と「探索」のトレードオフ

実験計画・最適化・進化計算：

より良い結果を得るためにには、新たに調べる対象をどのように選ぶべきだろうか？

= 手堅くオッズの低い馬券を買うか、ハイリスクを承知で万馬券に賭けてみるか？

**Exploitation (利用)**

"学習" 今までのデータから当てられそうなものを試す

既に得られていること(今あるデータ)から考えて、良くなりそうな対象を調べる

- 良い点：手元にあるデータの傾向に再現性があれば当たる論理的根拠がある
- 悪い点：木を見て森を見ず的に重要な鉱脈を見逃し凡庸な微改良にハマるかも

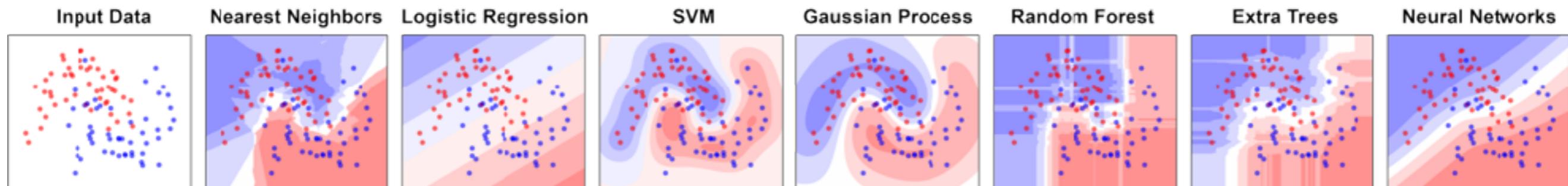
**Exploration (探索)**

"発見" 今までのデータにはない新しいものを試す

今のところ情報がなく不確実だが、もしかしたら良い可能性がある対象を調べる

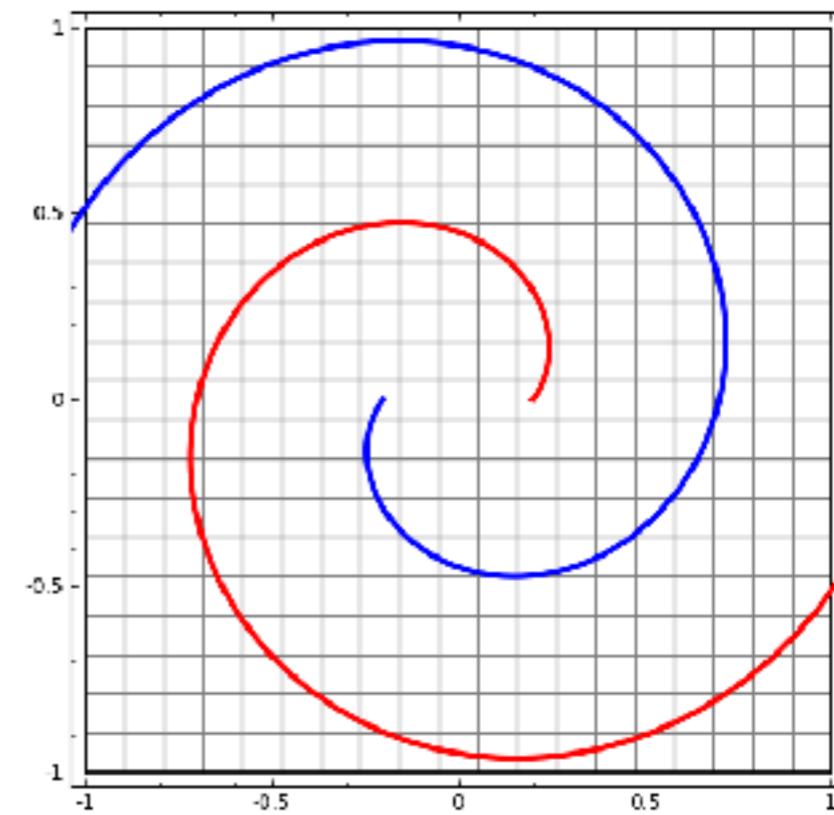
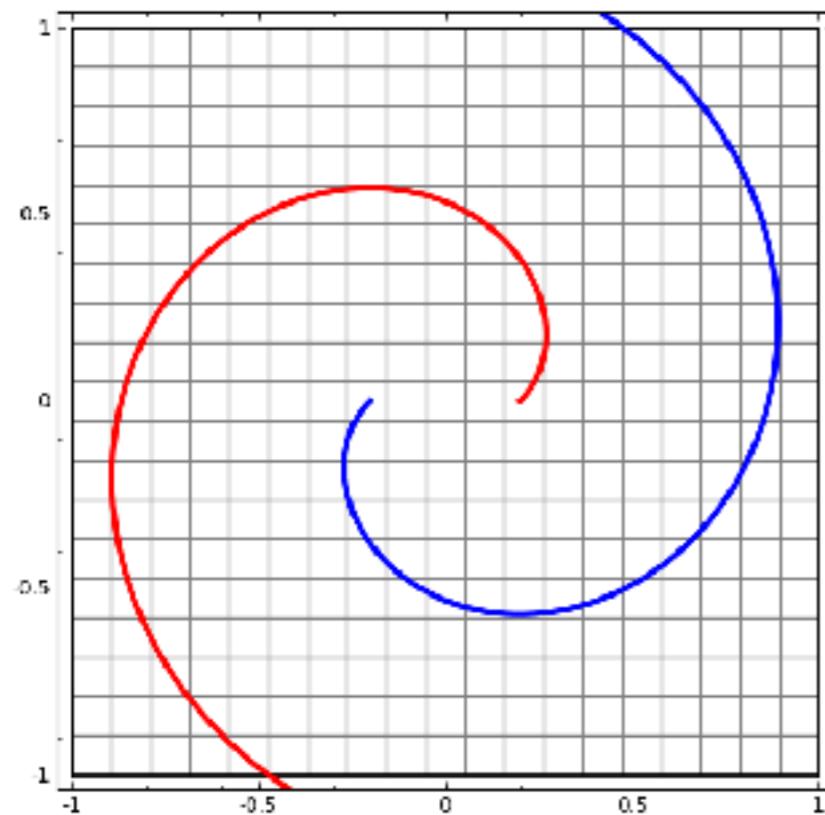
- 良い点：知識・情報の拡充になり、今まで知られていない新たな発見があるかも！
- 悪い点：当たる論理的根拠に乏しく、全く何も得られない可能性も高いかも

# 「利用」 与えられた事例以外へ転移可能な汎用表現の獲得



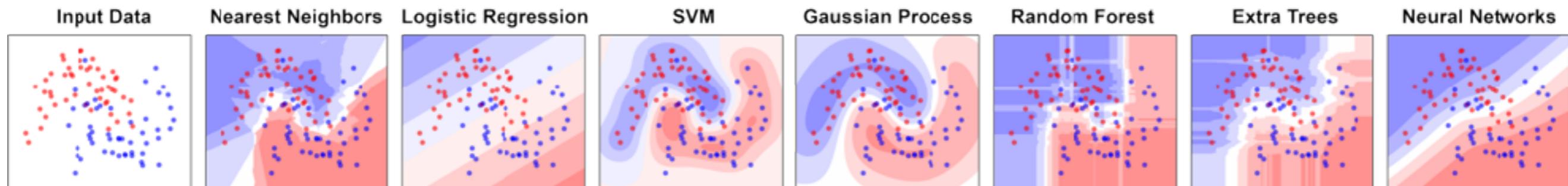
線形分離可能な表現への変換を学習

このタスクは実践的には依然むずかしく  
間違えると元より酷くなりうる..



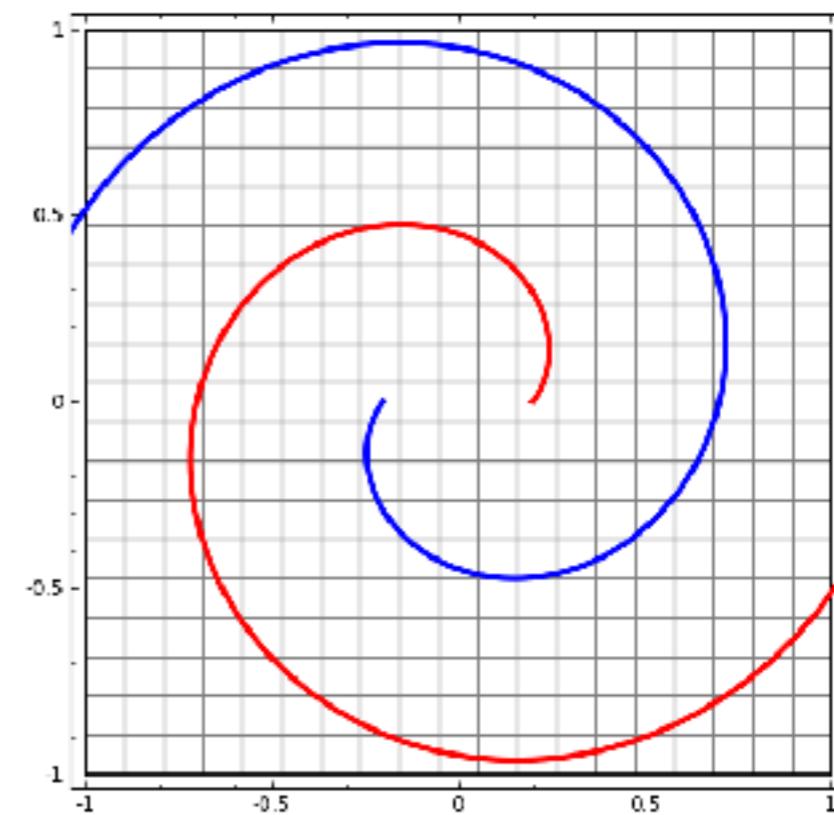
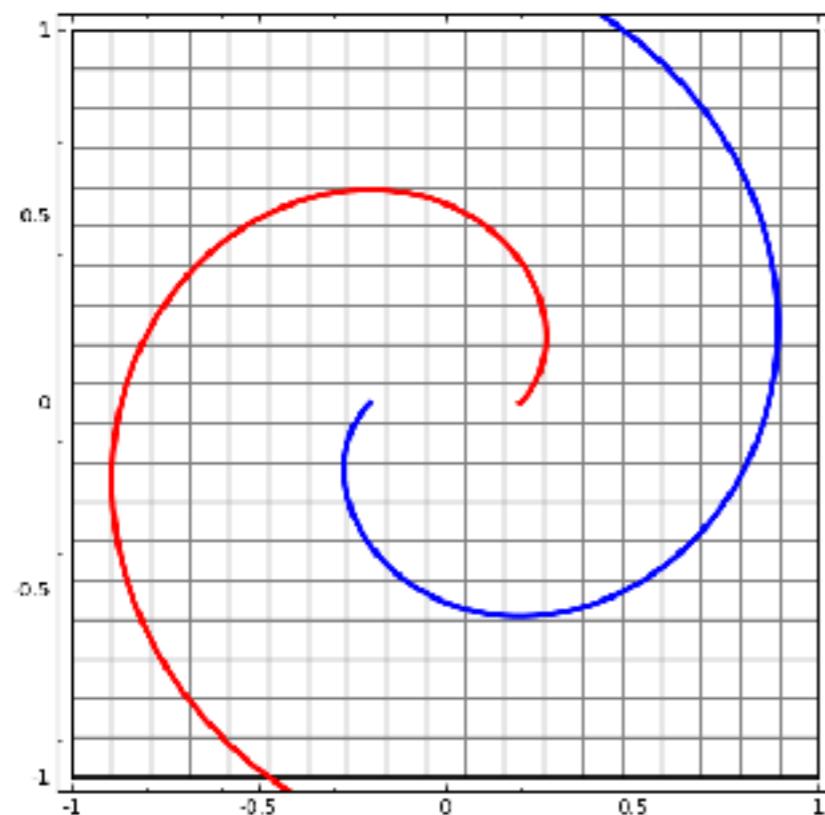
<https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/>

# 「利用」 与えられた事例以外へ転移可能な汎用表現の獲得



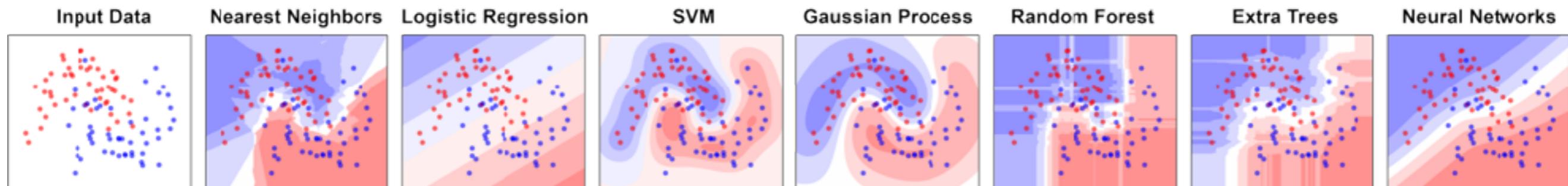
線形分離可能な表現への変換を学習

このタスクは実践的には依然むずかしく  
間違えると元より酷くなりうる..



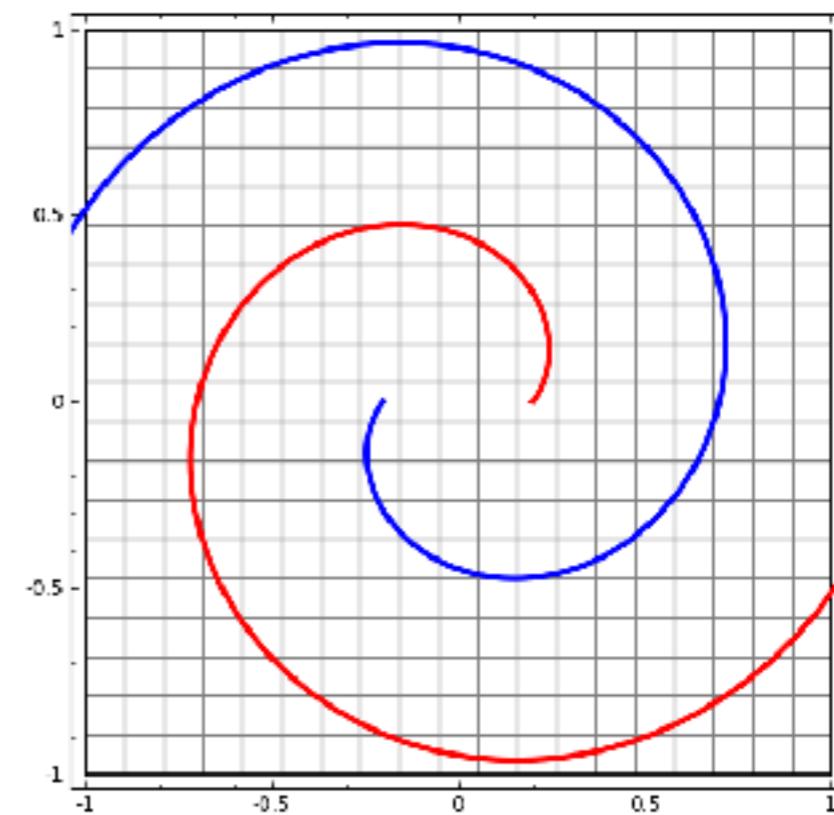
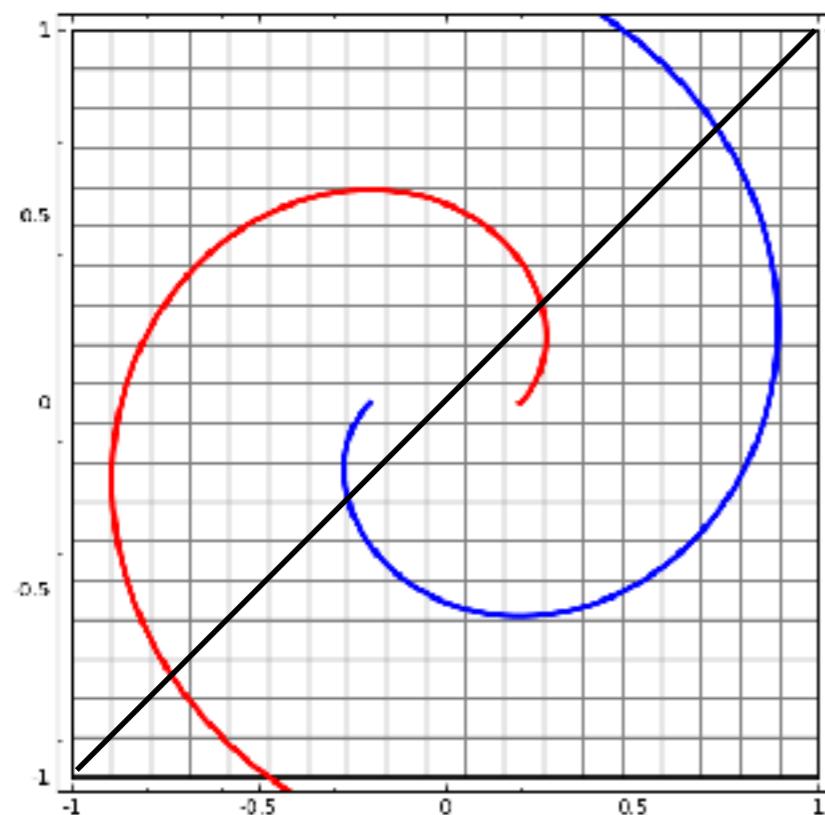
<https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/>

# 「利用」 与えられた事例以外へ転移可能な汎用表現の獲得



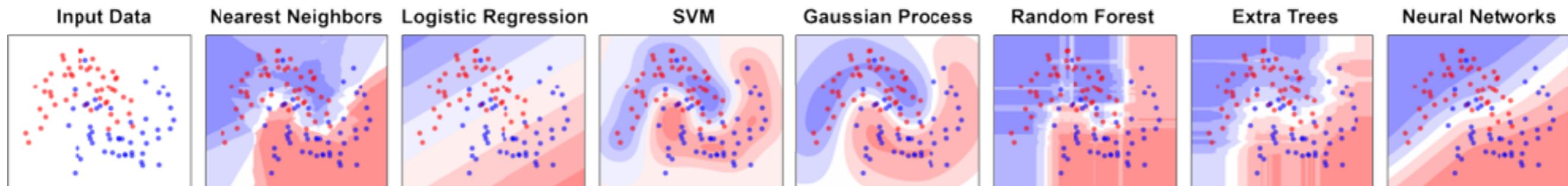
線形分離可能な表現への変換を学習

このタスクは実践的には依然むずかしく  
間違えると元より酷くなりうる..



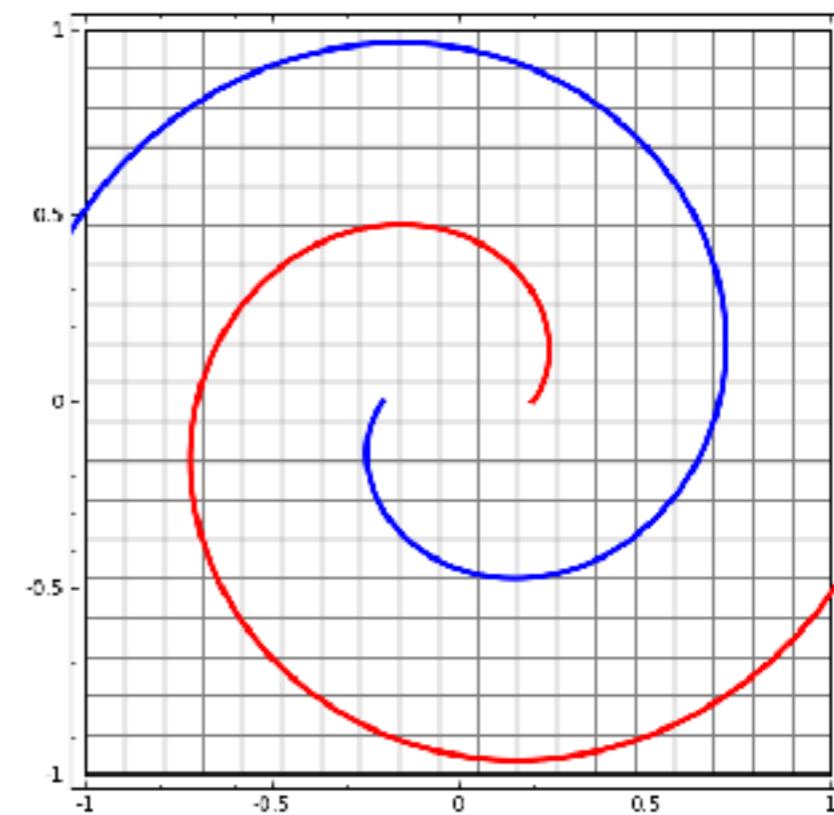
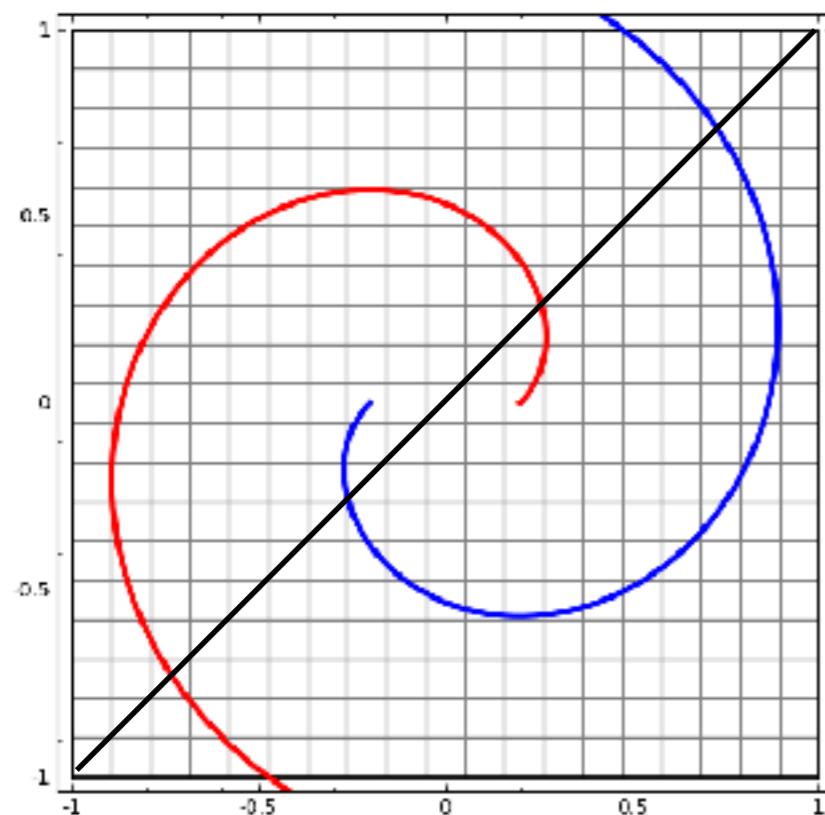
<https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/>

# 「利用」 与えられた事例以外へ転移可能な汎用表現の獲得



線形分離可能な表現への変換を学習

このタスクは実践的には依然むずかしく  
間違えると元より酷くなりうる..



<https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/>

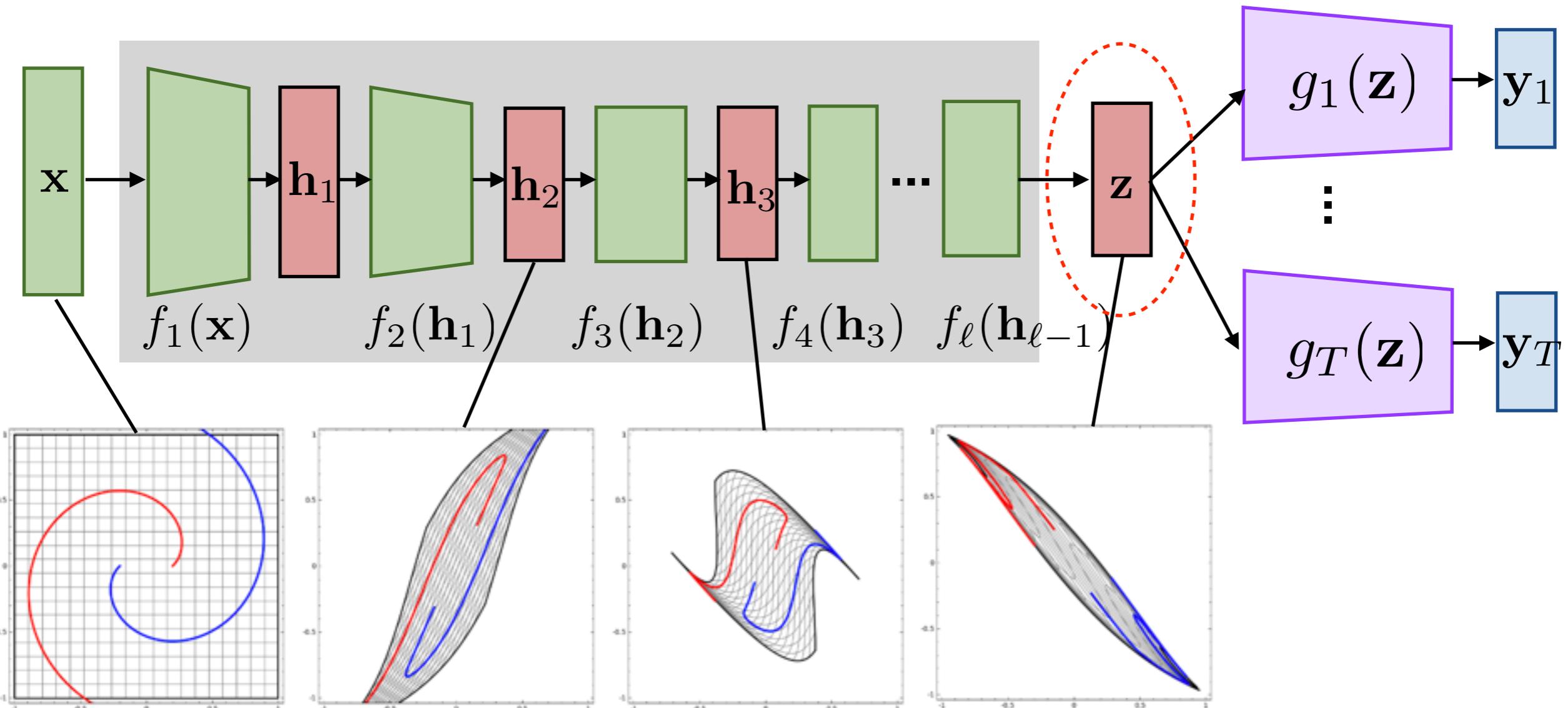
# データからの表現学習

いくらでも複雑な作り込みが自由に可能

→ 過適合のリスク・学習時間の増大

有効な(潜在的)特徴表現 $z$ への変換を学習するbackbone

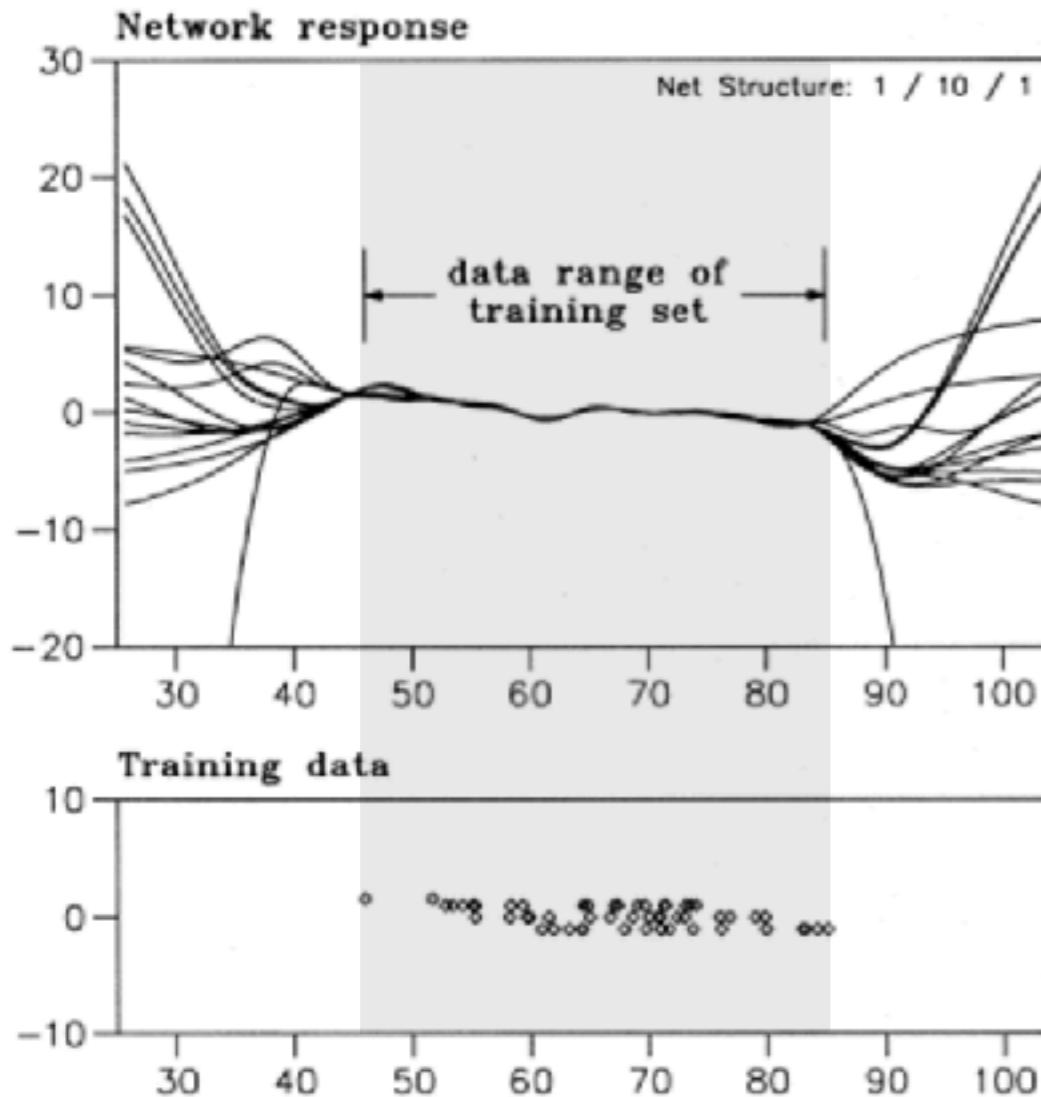
Task-specific head



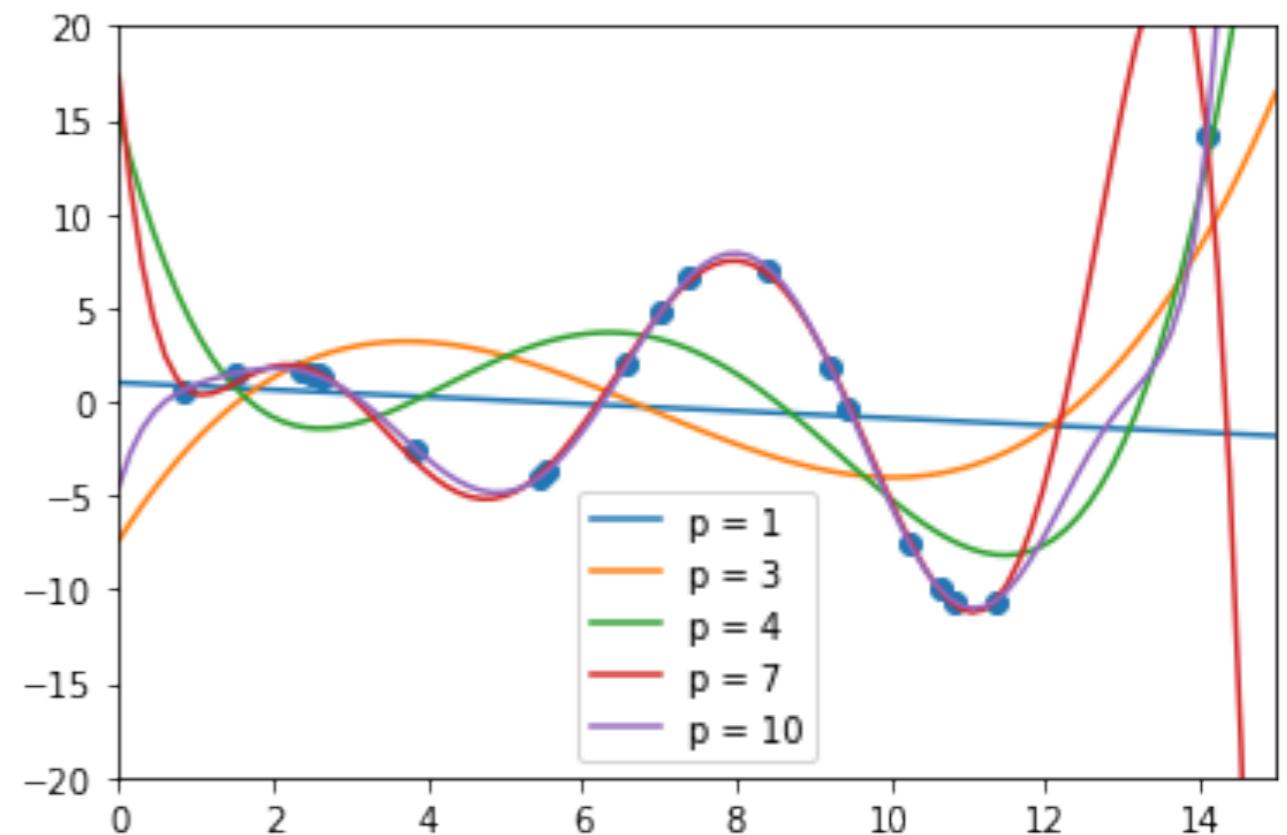
- 潜在変数空間での"内挿"は入力空間での直感と異なり得る (特に非常に高次元な場合)
- 目前の1タスクだけに特化しないための潜在空間の"良さ"をドメイン知識で正則化?

# 過適合と内挿 vs 外挿問題

パラメタ数(モデルの自由度・モデルの複雑度)をやみくもに大きくしても過適合のリスクが増大するだけ！(考えられる変数全部入り回帰の揶揄 = "Kitchen sink"回帰)



Highly Inaccurate Model Predictions from Extrapolation (Lohninger 1999)



$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x_p^p$$

# 高次元内挿にまつわる謎 overparametrization

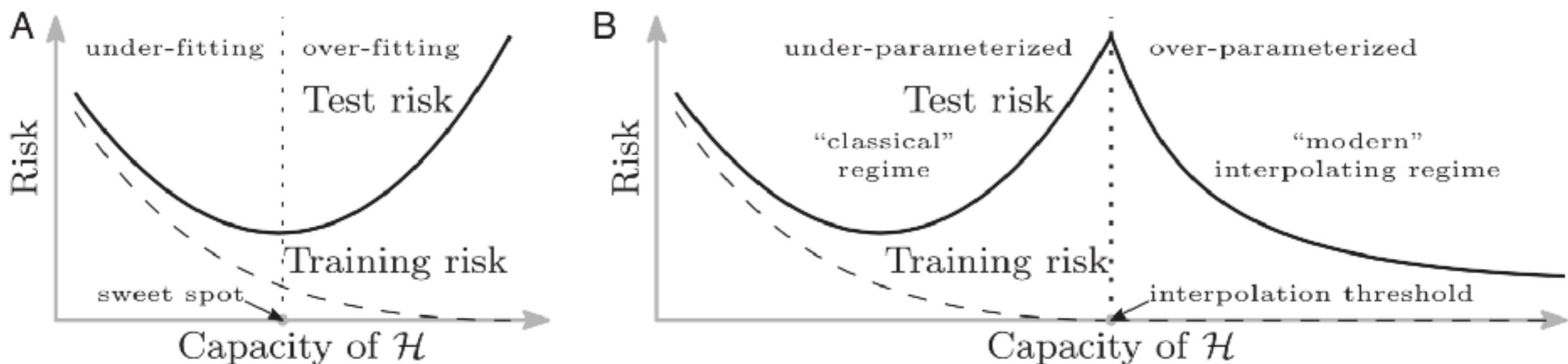
PNAS | August 6, 2019 | vol. 116 | no. 32 | 15849–15854

## Reconciling modern machine-learning practice and the classical bias–variance trade-off

Mikhail Belkin<sup>a,b,1</sup>, Daniel Hsu<sup>c</sup>, Siyuan Ma<sup>a</sup>, and Soumik Mandal<sup>a</sup>

<sup>a</sup>Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210; <sup>b</sup>Department of Statistics, The Ohio State University, Columbus, OH 43210; and <sup>c</sup>Computer Science Department and Data Science Institute, Columbia University, New York, NY 10027

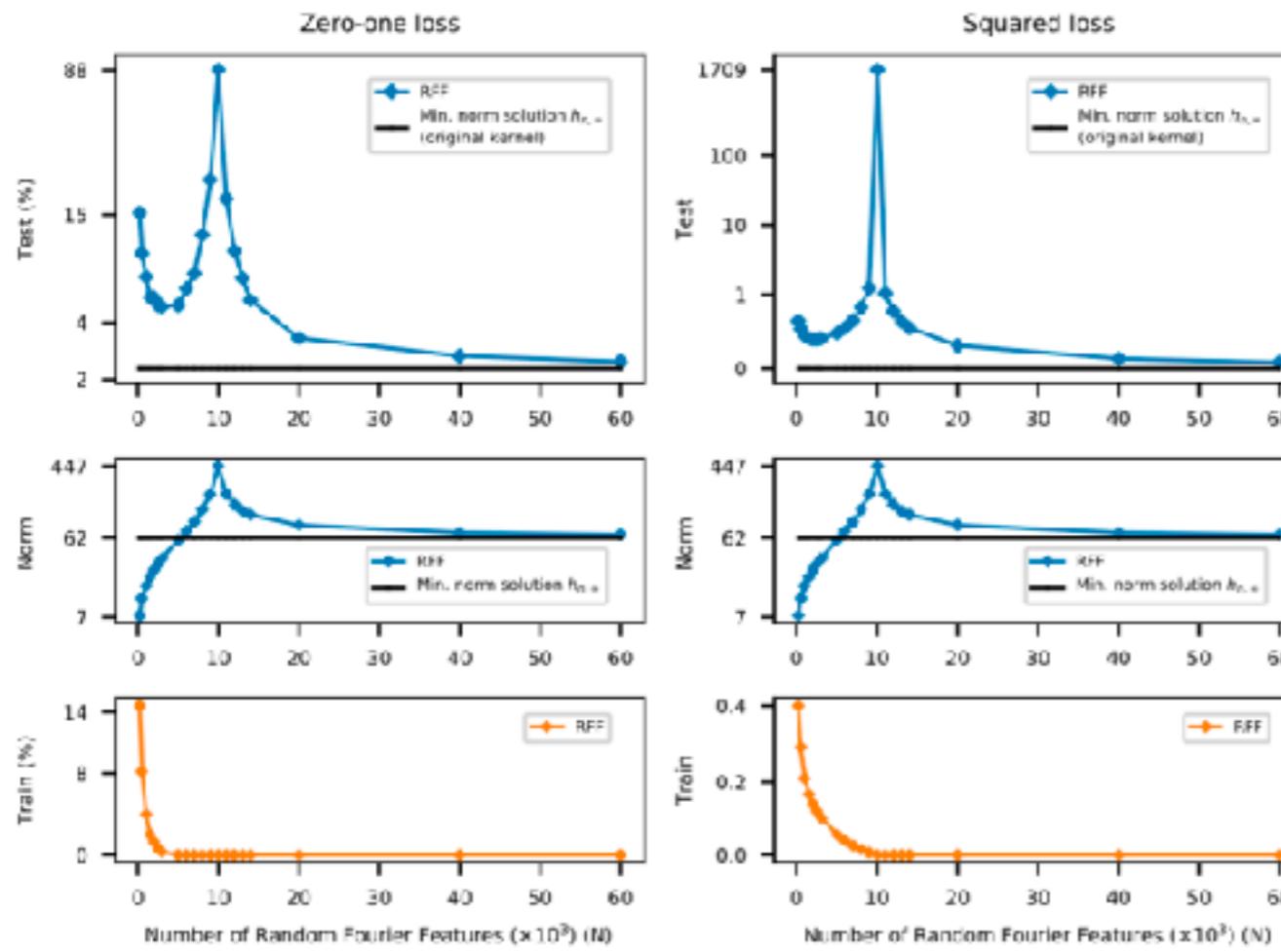
Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved July 2, 2019 (received for review February 21, 2019)



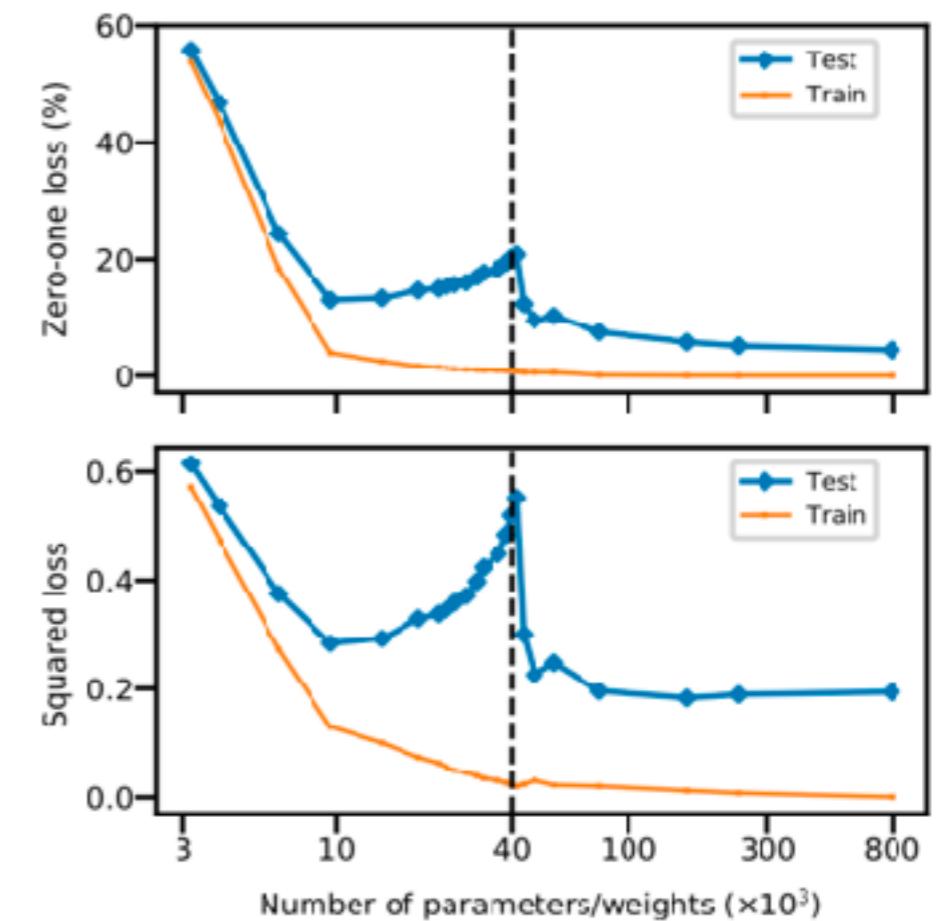
**Fig. 1.** Curves for training risk (dashed line) and test risk (solid line). **(A)** The classical U-shaped risk curve arising from the bias–variance trade-off. **(B)** The double-descent risk curve, which incorporates the U-shaped risk curve (i.e., the “classical” regime) together with the observed behavior from using high-capacity function classes (i.e., the “modern” interpolating regime), separated by the interpolation threshold. The predictors to the right of the interpolation threshold have zero training risk.

# "double-descent risk curve"は再現可能で議論は続く...

## Random Fourier Features (RFF) on MNIST



## FC-NN on MNIST



Also see  
AISTATS 2019

Does data interpolation contradict statistical optimality?

# 良性の過適合？とにかく訓練データを内挿すれば良い!?

PNAS (2020)

## Benign overfitting in linear regression

Peter L. Bartlett<sup>a,b,1</sup>, Philip M. Long<sup>c</sup>, Gábor Lugosi<sup>d,e,f</sup>, and Alexander Tsigler<sup>d</sup>

<sup>a</sup>Department of Statistics, University of California, Berkeley, CA 94720-3860; <sup>b</sup>Computer Science Division, University of California, Berkeley, CA 94720-3860; <sup>c</sup>Google Brain, Mountain View, CA 94043; <sup>d</sup>Economics and Business, Pompeu Fabra University, 08005 Barcelona, Spain; <sup>e</sup>Institució Catalana d'Estudis Avançats, Passeig, Lluís Companys 25, 08010 Barcelona, Spain; and <sup>f</sup>Barcelona Graduate School of Economics, 08005 Barcelona, Spain

Edited by Richard Baraniuk, Rice University, Houston, TX, and accepted by Editorial Board Member David L. Donoho March 4, 2020 (received June 2, 2019)

The phenomenon of benign overfitting is one of the key mysteries uncovered by deep learning methodology: deep neural networks seem to predict well, even with a perfect fit to noisy training data. Motivated by this phenomenon, we consider when a perfect fit to training data in linear regression is compatible with accurate prediction. We give a characterization of linear regression problems for which the minimum norm interpolating prediction rule has near-optimal prediction accuracy. The characterization is in terms of two notions of the effective rank of the data covariance. It shows that overparameterization is essential for benign overfitting in this setting: the number of directions in parameter space that are unimportant for prediction must significantly exceed the sample size. By studying examples of data covariance properties that this characterization shows are required for benign overfitting, we find an important role for finite-dimensional data: the accuracy of the minimum norm interpolating prediction rule approaches the best possible accuracy for much smaller values of the sample size.

enough that a perfect fit is guaranteed. We consider infinite-dimensional space (a separable Hilbert space) and results apply to a finite-dimensional subspace as a. There is an ideal value of the parameters,  $\theta^*$ , for the linear prediction rule that minimizes the expected loss. We ask when it is possible to fit the data exactly with the prediction accuracy of  $\theta^*$ . Since there are more parameters than the sample size in order for the solution might be underdetermined, and there might be many interpolating solutions. We consider a natural choice: choose the parameter vector  $\hat{\theta}$  with the smallest norm among all vectors that gives perfect predictions on the sample. (This corresponds to using the pseudoinverse of the normal equations; see below.) We ask when it is possible to overfit in this way—and embed all of the noise into the parameter estimate  $\hat{\theta}$ —without harming accuracy.

Ann. Statist. (2020)

## JUST INTERPOLATE: KERNEL “RIDGELESS” REGRESSION CAN GENERALIZE

BY TENGYUAN LIANG<sup>1</sup> AND ALEXANDER RAKHLINE<sup>2</sup>

<sup>1</sup>Econometrics and Statistics, Booth School of Business, University of Chicago, tengyuan.liang@chicagobooth.edu

<sup>2</sup>Center for Statistics & IDSS, Massachusetts Institute of Technology, rakhlin@mit.edu

In the absence of explicit regularization, Kernel “Ridgeless” Regression with nonlinear kernels has the potential to fit the training data perfectly. It has been observed empirically, however, that such interpolated solutions can still generalize well on test data. We isolate a phenomenon of implicit regularization for minimum-norm interpolated solutions which is due to a combination of high dimensionality of the input data, curvature of the kernel function and favorable geometric properties of the data such as an eigenvalue decay of the empirical covariance and kernel matrices. In addition to deriving a data-dependent upper bound on the out-of-sample error, we present experimental evidence suggesting that the phenomenon occurs in the MNIST dataset.

NeurIPS (2018)

## Overshooting or perfect fitting? Risk bounds for classification and regression rules that interpolate

Mikhail Belkin  
The Ohio State University

Daniel Hsu  
Columbia University

Partha P. Mitra  
Cold Spring Harbor Laboratory

### Abstract

Many modern machine learning models are trained to achieve zero or near-zero training error in order to obtain near-optimal (but non-zero) test error. This phenomenon of strong generalization performance for “overfitted” / interpolated classifiers appears to be ubiquitous in high-dimensional data, having been observed in deep networks, kernel machines, boosting and random forests. Their performance

arXiv (2019)

## Surprises in High-Dimensional Ridgeless Least Squares Interpolation

Trevor Hastie

Andrea Montanari\*

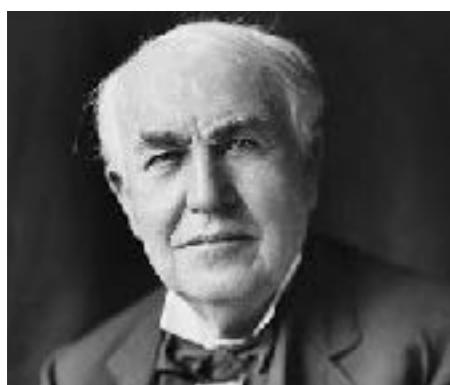
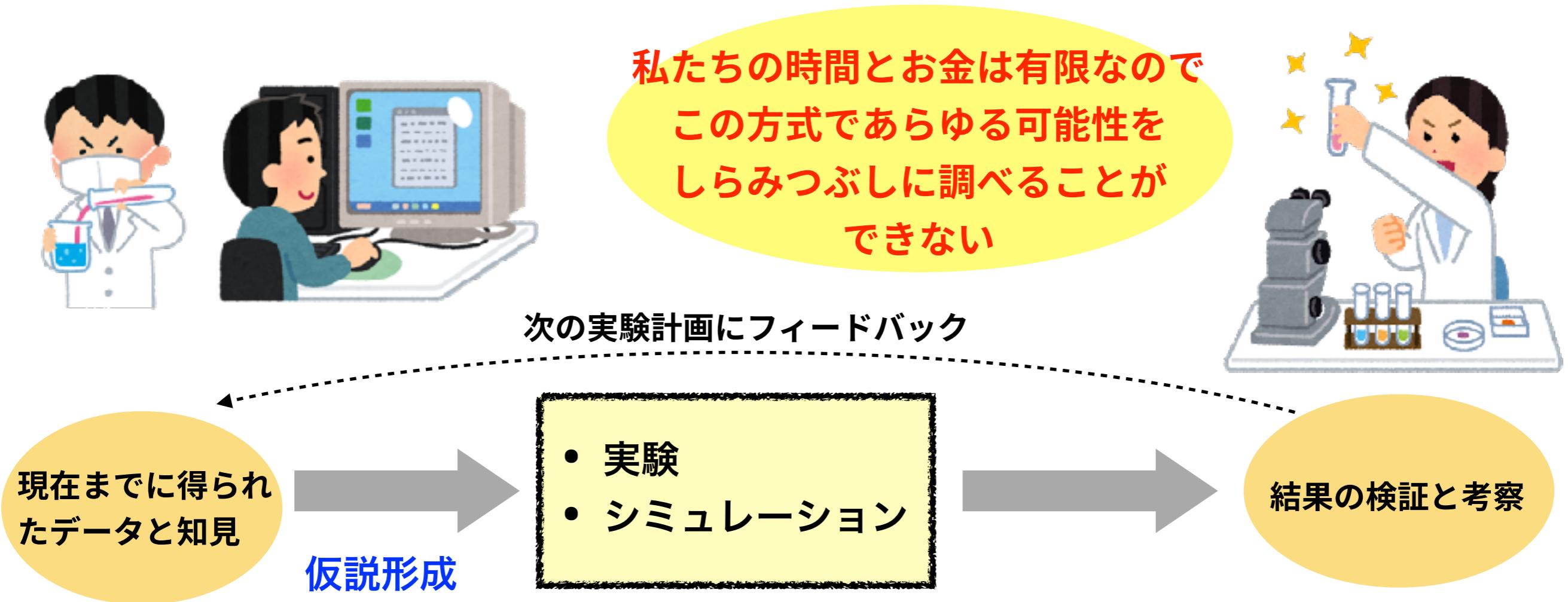
Saharon Rosset

Ryan J. Tibshirani\*

### Abstract

Interpolators—estimators that achieve zero training error—have attracted growing attention in machine learning, mainly because state-of-the-art neural networks appear to be models of this type. In this paper, we study minimum  $\ell_2$  norm (“ridgeless”) interpolation in high-dimensional least squares regression. We consider two different models for the feature distribution: a linear model, where the feature vectors  $x_i \in \mathbb{R}^p$  are obtained by applying a linear transform to a vector of i.i.d. entries,  $x_i = \Sigma^{1/2} z_i$  (with  $z_i \in \mathbb{R}^p$ ); and a nonlinear model, where the feature vectors are obtained by passing the input through a random one-layer neural network,  $x_i = \varphi(W z_i)$  (with  $z_i \in \mathbb{R}^d$ ,  $W \in \mathbb{R}^{p \times d}$  a matrix of i.i.d. entries, and  $\varphi$  an activation function acting componentwise on  $W z_i$ ). We recover—in a precise quantitative way—several phenomena that have been observed in large-scale neural networks and kernel machines, including the “double descent” behavior of the prediction risk, and the potential benefits of overparametrization.

# 「探索」 "Edisonian empiricism"を乗り越える



Thomas  
Edison

- Genius is 1% inspiration and 99% perspiration.
- There is no substitute for hard work.
- I have not failed. I've just found 10,000 ways that won't work.
- :



よくよく考えるとすごくブラックなことしか言ってない



# 有機合成におけるインフォマティクス(実験計画法)の活用

MS Sigman, KC Harper, EN Bess, A Milo

**The development of multidimensional analysis tools  
for asymmetric catalysis and beyond.**

*Accounts of Chemical Research* 2016 49 (6), 1292-1301.

1. Harper & Sigman, *PNAS* 2011
2. Harper & Sigman, *Science* 2011
  - Werner, Mei, Burckle & Sigman, *Science* 2012
3. Milo, Bess & Sigman, *Nature* 2014
  - Harper, Vilardi & Sigman, *JACS* 2013
  - Bess, Bischoff, Sigman, *PNAS* 2014
4. Niemeyer, Milo, Hickey & Sigman, *Nat Chem* 2016;  
Milo, Neel, Toste & Sigman, *Science* 2015



**Matthew S Sigman**

Dept. Chemistry  
The University of Utah

ARTICLE

18 JULY 2019 | VOL 571 | NATURE

<https://doi.org/10.1038/s41586-019-1384-z>

## Holistic prediction of enantioselectivity in asymmetric catalysis

Jolene P. Reid<sup>1</sup> & Matthew S. Sigman<sup>1\*</sup>

化学反応開発用の汎用記述子 +  
実験計画法 (完全実施要因計画 +  
応答曲面法)



ごく最近AI/MLによって来た!?

# 探索とオートメーション(自動化)

NATURE REVIEWS | DRUG DISCOVERY  
VOLUME 17 | FEBRUARY 2018 | 97

INNOVATION

## Automating drug discovery

Gisbert Schneider



Figure 2 | Automated drug discovery facilities. a | Millions of compound samples are stored in compact high-capacity facilities and handled by robots. b | Robot systems perform both high-throughput and medium-throughput screening of up to ten thousand samples per day to determine the activity against the biological target of interest. Multiple arms and flexible workstations enable fully automated liquid dispensing, compound

preparation and testing. These storage and screening systems have become cornerstones of contemporary drug discovery. c | A prototype of a novel miniaturized design–synthesize–test–analyse facility for rapid automated drug discovery at AstraZeneca is shown. Images a and b courtesy of Jan Kriegl, Boehringer Ingelheim Pharma; image c courtesy of Michael Kossenjans, AstraZeneca.

## Toyota teams with China's CATL and BYD to power electric ambitions

Automaker diversifies battery source and moves up electrification goal by 5 years

YUKIHIRO OMOTO, Nikkei staff writer

JUNE 07, 2019 02:00 JST • UPDATED ON JUNE 07, 2019 14:39 JST



製品ラインの(ほぼ)無人化は高い効率性だけではなく高い再現性も担保できる



# 探索とオートメーション(自動化)

NATURE REVIEWS | DRUG DISCOVERY  
VOLUME 17 | FEBRUARY 2018 | 97

INNOVATION

## Automating drug discovery

Gisbert Schneider



Figure 2 | Automated drug discovery facilities. a | Millions of compound samples are stored in compact high-capacity facilities and handled by robots. b | Robot systems perform both high-throughput and medium-throughput screening of up to ten thousand samples per day to determine the activity against the biological target of interest. Multiple arms and flexible workstations enable fully automated liquid dispensing, compound

preparation and testing. These storage and screening systems have become cornerstones of contemporary drug discovery. c | A prototype of a novel miniaturized design–synthesize–test–analyse facility for rapid automated drug discovery at AstraZeneca is shown. Images a and b courtesy of Jan Kriegl, Boehringer Ingelheim Pharma; image c courtesy of Michael Kossenjans, AstraZeneca.

## Toyota teams with China's CATL and BYD to power electric ambitions

Automaker diversifies battery source and moves up electrification goal by 5 years

YUKIHIRO OMOTO, Nikkei staff writer

JUNE 07, 2019 02:00 JST • UPDATED ON JUNE 07, 2019 14:39 JST



製品ラインの(ほぼ)無人化は高い効率性だけではなく高い再現性も担保できる



# 自然科学における探索的研究の(半)自動化?

"The authors estimate that a human scientist would have taken 1,000 times longer to produce similar results."

Article

Nature | Vol 583 | 9 July 2020 | 237

nature

## A mobile robotic chemist

<https://doi.org/10.1038/s41586-020-2442-2>

Received: 1 November 2019

Accepted: 25 March 2020

Published online: 8 July 2020

 Check for updates



# 自然科学における探索的研究の(半)自動化?

"The authors estimate that a human scientist would have taken 1,000 times longer to produce similar results."

Article

Nature | Vol 583 | 9 July 2020 | 237

nature

## A mobile robotic chemist

<https://doi.org/10.1038/s41586-020-2442-2>

Received: 1 November 2019

Accepted: 25 March 2020

Published online: 8 July 2020

 Check for updates



# この種の自動化も機械学習との境界領域へ

## AMAZON PICKING CHALLENGE

AMAZON ROBOT RESEARCH PROJECT



FUTURE OF ROBOTS

ROBOTICS

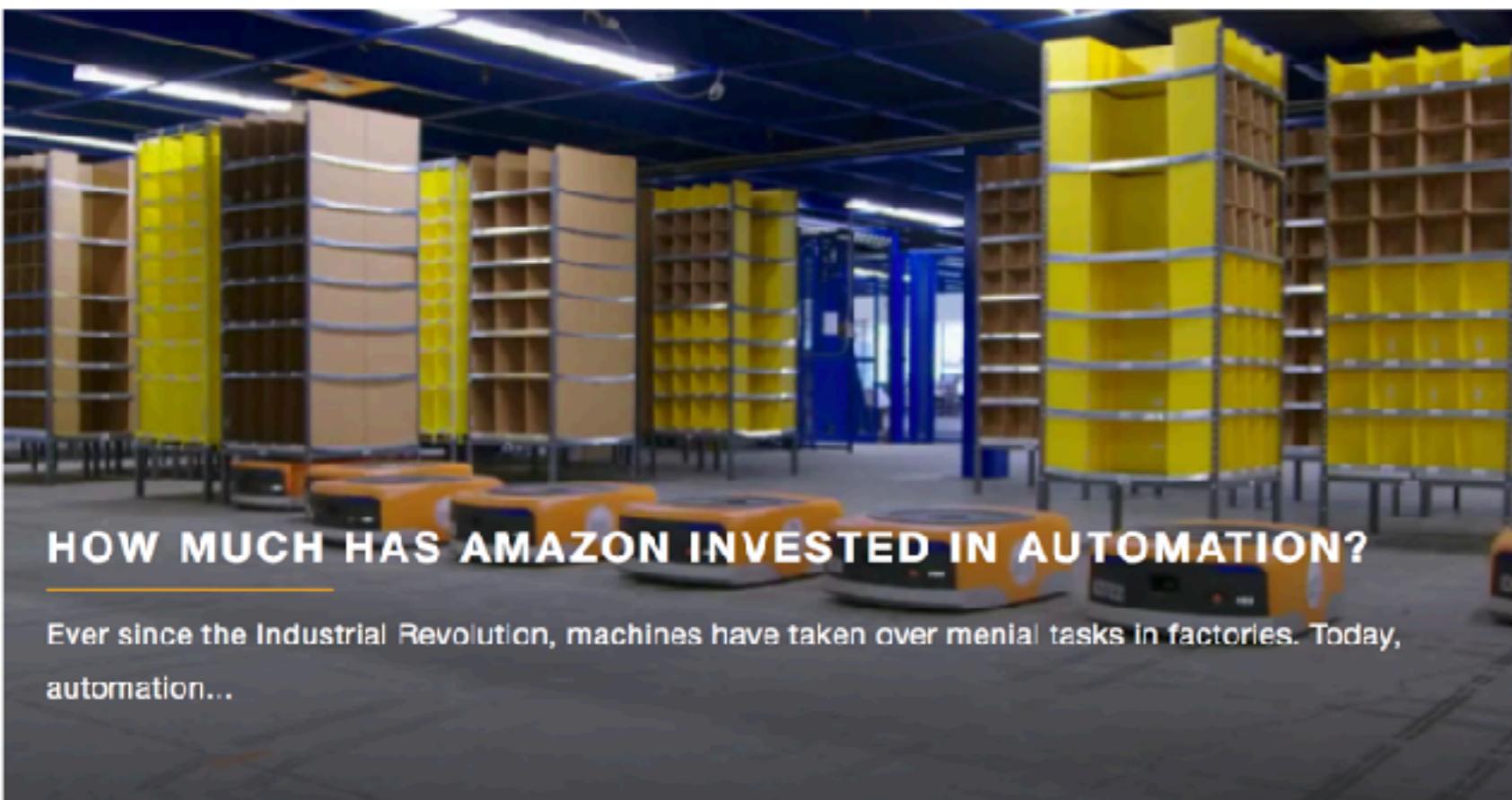
ABOUT US

MODERN WAR

FUTURE OF AI

CONTACT US

Search



### HOW MUCH HAS AMAZON INVESTED IN AUTOMATION?

Ever since the Industrial Revolution, machines have taken over menial tasks in factories. Today, automation...

ABOUT US

## AMAZON PICKING CHALLENGE

AMAZON ROBOT RESEARCH PROJECT

As we all know, Amazon is infamous when it comes to automation and scaling large tasks at ease. The Amazon Robotics/Picking Challenge was a competition from early 2017 that allowed individuals to create their own small robotic systems to help with mass production and management.

LATEST POSTS

After entering and coming second place, we

# "Autonomous Discovery in the Chemical Sciences"

Angew. Chem. Int. Ed. 2020, 59, 2–38

Machine Learning

How to cite:  
International Edition: doi.org/10.1002/anie.201909987  
German Edition: doi.org/10.1002/ange.201909987

## Autonomous Discovery in the Chemical Sciences Part I: Progress

Connor W. Coley,\* Natalie S. Eyke, and Klavs F. Jensen\*

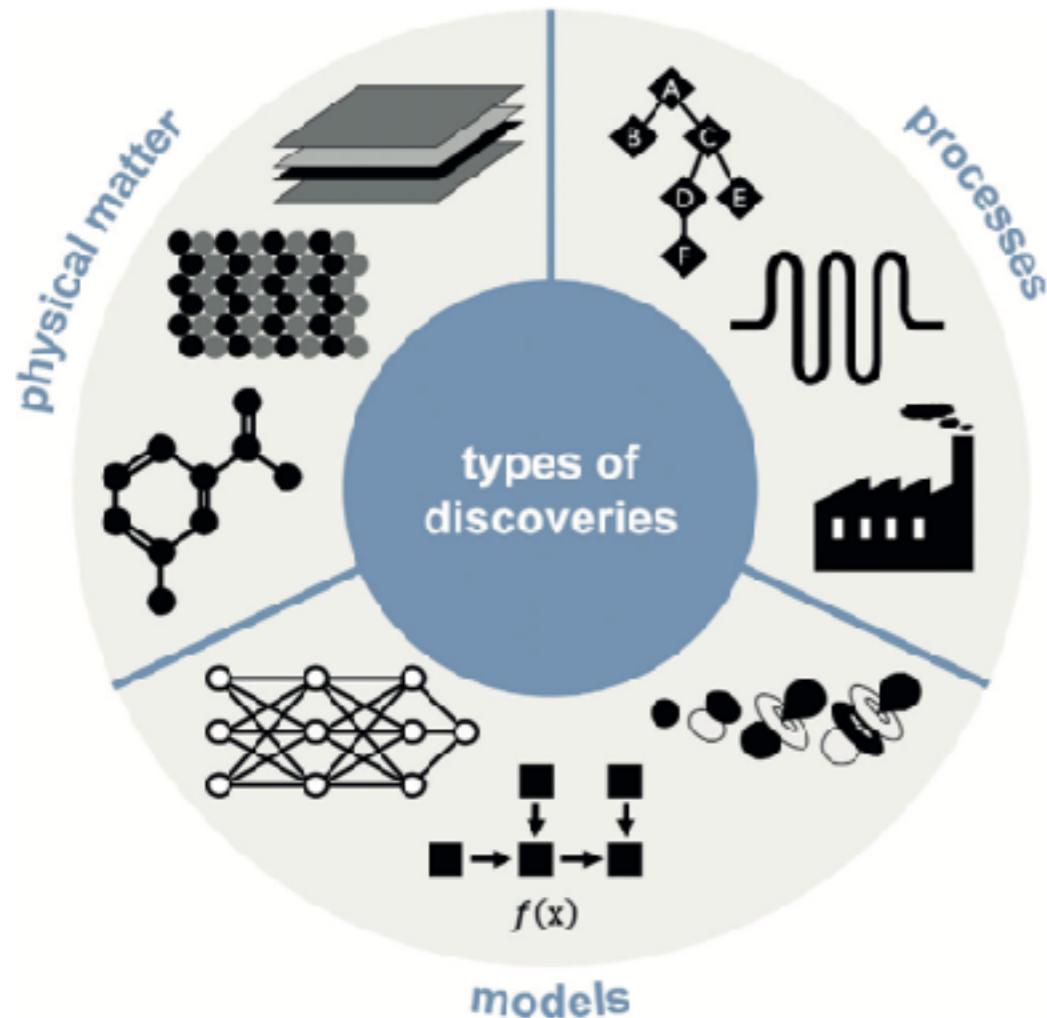
Angew. Chem. Int. Ed. 2020, 59, 2–25

Computer Chemistry

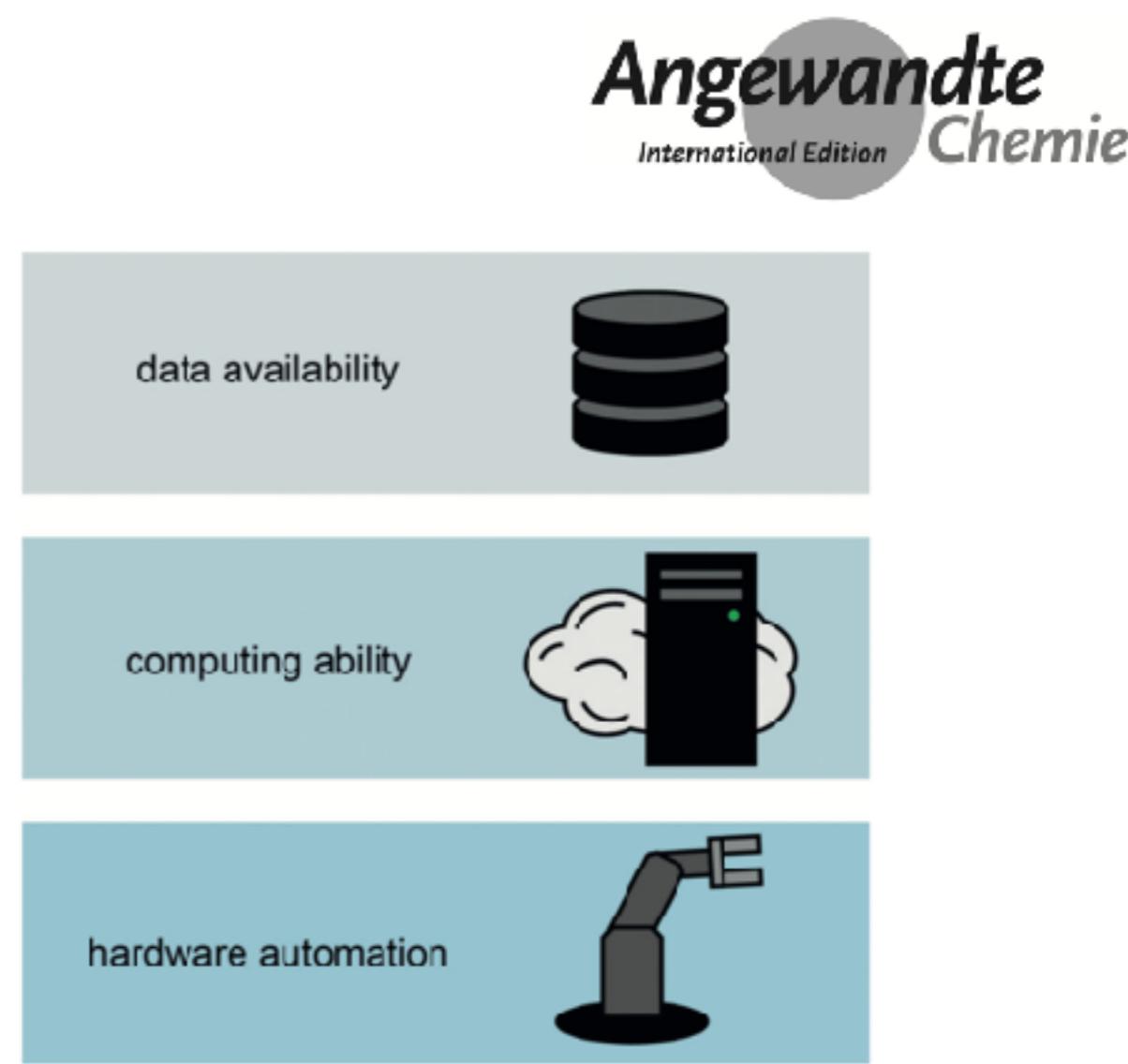
How to cite:  
International Edition: doi.org/10.1002/anie.201909989  
German Edition: doi.org/10.1002/ange.201909989

## Autonomous Discovery in the Chemical Sciences Part II: Outlook

Connor W. Coley,\* Natalie S. Eyke, and Klavs F. Jensen\*



**Figure 1.** The three broad categories of discovery described in this Review: physical matter, processes, and models.



**Figure 4.** The factors that have enabled autonomous discovery fall into one of three main categories.

**Angewandte**  
International Edition  
**Chemie**

# 探索とAutoML: 量子化学計算や口ボのselfplayで化学探索!?

## モデルベース強化学習



AlphaGo  
(Nature, Jan 2016)

### ARTICLE

#### Mastering the game of Go without human knowledge

David Silver<sup>1,2,3\*</sup>, Julian Schrittwieser<sup>1,2</sup>, Karen Simonyan<sup>1</sup>, Thomas Hubert<sup>1</sup>, Arthur Guez<sup>1</sup>, Thomas Lillicrap<sup>1</sup>, Matheus Lai<sup>1</sup>, Marc Lanctot<sup>1</sup>, David Silver<sup>1</sup>, Timothy Lillicrap<sup>1</sup>, Demis Hassabis<sup>1</sup>, DeepMind, London, UK; University College London, London, UK

A long-standing goal of artificial intelligence is for algorithms that learn, without any preexisting knowledge, to outperform domain experts. AlphaGo, a program that plays Go, has achieved this goal. This new version of AlphaGo, known as AlphaGo Zero, can learn to play Go from scratch in just three days. The algorithm uses entirely self-supervised learning from random games, and by self-improvement learns to beat itself. It has no human or expert knowledge input, and its performance is based solely on reinforcement learning, without human data, game rules, or domain knowledge. Go is a highly complex game, but the algorithm's performance is not limited to Go. Starting from scratch, it can learn to play chess and shogi, two other games that require superhuman performance. AlphaGo Zero is different to prior approaches in that it does not require a neural network to predict its own moves or actions, and that the training of AlphaGo's policy is not guided by human feedback. Instead, the algorithm uses a general-purpose reinforcement learning algorithm that can learn to play any two-player zero-sum game.

Many previous research articles in robotics have been evaluating supervised learning approaches to learn to play games, or to learn to solve domains by imitating domain experts<sup>1–4</sup>. However, agents that are not often reprogrammed or updated to change their behavior to reflect changes in the environment are not able to learn to play well against opponents who are not experts<sup>5–7</sup>. In contrast, the AlphaGo Zero program recently achieved superhuman performance in the game of Go by reinforcement learning from self-play<sup>8</sup>. In this paper, we generalize this approach into a single AlphaZero algorithm that can achieve super-human performance in many challenging games. Starting from random play and given no domain knowledge except the game rules, AlphaZero convincingly defeated a world champion program in the games of chess and shogi (Japanese chess), as well as Go.

AlphaGo Zero  
(Nature, Oct 2017)

○○ Zeroは人間がplayした棋譜データは全く不要!

Silver et al., *Science* **362**, 1140–1144 (2018)

7 December 2018 Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model

Julian Schrittwieser,<sup>1\*</sup> Ioannis Antonoglou,<sup>1,2\*</sup> Thomas Hubert,<sup>1</sup> Karen Simonyan,<sup>1</sup> Laurent Sifre,<sup>1</sup> Simon Schmitz,<sup>1</sup> Arthur Guez,<sup>1</sup> Edward Lockhart,<sup>1</sup> Demis Hassabis,<sup>1</sup> Thore Graepel,<sup>1,2</sup> Timothy Lillicrap,<sup>1</sup> David Silver<sup>1,2</sup>

<sup>1</sup>DeepMind, 6 Pancras Square, London N1C 4AG,  
<sup>2</sup>University College London, Gower Street, London WC1E 6BT.  
\*These authors contributed equally to this work.

### Abstract

Creating agents with planning capabilities has long been one of the main challenges in the pursuit of artificial intelligence. Tree-based planning methods have enjoyed huge success in challenging domains, such as chess and Go, where a perfect simulator is available. However, in real-world problems the dynamics governing the environment are often complex and unknown. In this work, we present the AlphaZero algorithm which, by combining a tree-based search with a learned model, achieves superhuman performance in a range of challenging and visually complex domains, without any knowledge of their underlying dynamics. AlphaZero is a model that, when applied iteratively, predicts the quantities most closely relevant to planning: the reward, the action-selection policy, and the value function. When evaluated on 27 different Atari games – the original video game environment for testing AI techniques – in which model-based planning approaches have historically struggled, our new algorithm achieves new state-of-the-art. When extended to Go, chess and shogi, without any knowledge of the game rules, AlphaZero matches the superhuman performance of the AlphaGoZero algorithm that was supplied with the game rules.

AlphaZero  
(Science, Dec 2018)

## AutoML

- Algorithm Configuration
- Hyperparameter Optimization (HPO)
- Neural Architecture Search (NAS)
- Meta Learning / Learning to Learn



## AutoML



Amazon  
SageMaker

# 今日の話

## 1. 機械学習+化学の面白さとチャレンジ

- 😊 CVとNLPの間の独特設定 + 第一原理がある + 実検証できる
- 🤔 表現学習 + 小サンプル + 強いデータバイアス

## 2. 機械学習の関連トピックの(面白さの)紹介

### 1. 探索と訓練分布外の予測

新たなデータを取る探索戦略 (実験計画・最適化・進化計算)  
外挿・内挿問題、良性の過適合

### 2. 学習の構成性と転移性

注意機構と転移学習・構成的学習

小サンプル学習 (転移学習、メタ学習、計量学習)

### 3. 演繹と帰納の融合

モデルベースの最適化・強化学習、第一原理計算との融合

# 注意機構と転移学習

Transformer(による転移学習)はLSTM/RNN/CNN(+注意)ベースだったNLPを革新した...

[papers.nips.cc](#) › paper › 7181-attenti... ▾ PDF このページを訳す

## Attention is All you Need - NIPS Proceedings

The dominant sequence transduction models are based on complex recurrent neural networks that include an encoder and a decoder. The best performing models connect the encoder and decoder through an attention.

A Vaswani 著 - 2017 - **被引用数: 11206** - 関連記事

挑戦的なタイトルを美しく回収 (NeurIPS 2017)

画像もTransformerでいけるのか...!?

## Image GPT

OpenAI

We find that, just as a large transformer model trained on language can generate coherent text, the same exact model trained on pixel sequences can generate coherent image completions and samples. By establishing a correlation between sample quality and image classification accuracy, we show that our best generative model also contains features competitive with top convolutional nets in the unsupervised setting.

[CODE](#) [ICML 2020 PAPER \(V1\)](#) [PAPER \(V2\)](#)

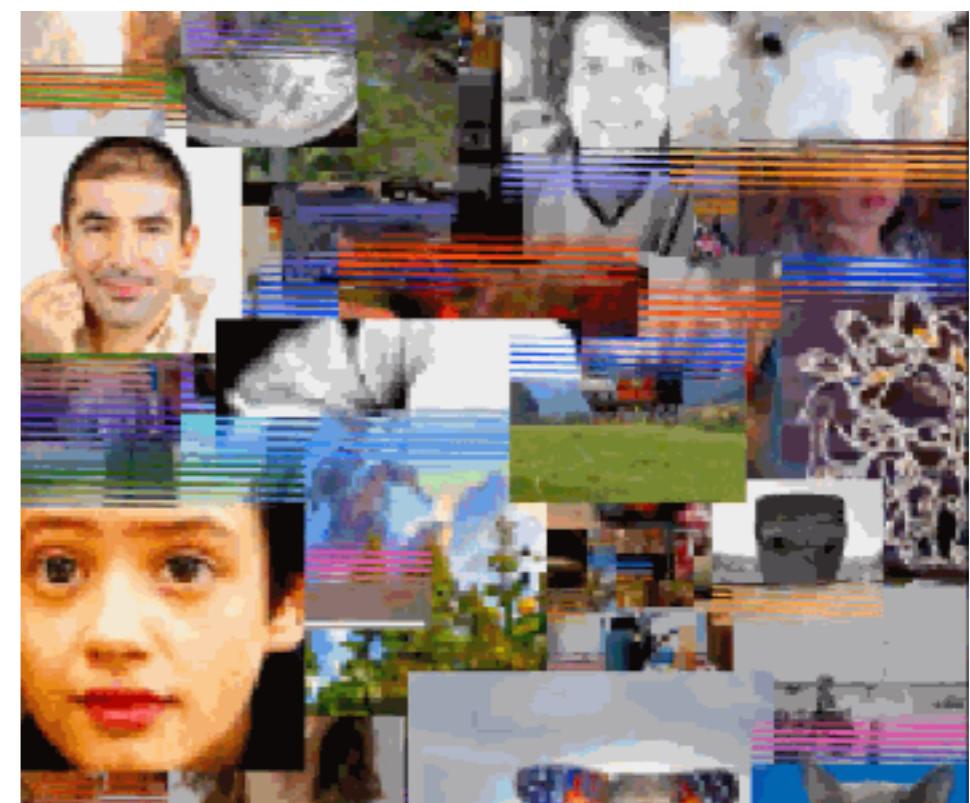


## 第2部：技術動向

# トランスフォーマー時代到来 翻訳技術から汎用言語系AIに

2016年のニューフル機械翻訳(NMT)の実用化は、翻訳業界に衝撃を与え、ボケトークのような自動翻訳端末の市場拡大につながるなど、社会に大きなインパクトを与えた。ただし、翻訳技術や自然言語処理技術(NLP)分野では、その後も革命級のブレークスルーが相次いでいる。翻訳を含む言語系の人工知能(AI)が従来の常識を次々と塗り替え、ありえないベースで発展している。

日経エレクトロニクス 2019年9月号



# 注意機構と転移学習

Transformer(による転移学習)はLSTM/RNN/CNN(+注意)ベースだったNLPを革新した...

[papers.nips.cc](#) › paper › 7181-attenti... ▾ PDF このページを訳す

## Attention is All you Need - NIPS Proceedings

The dominant sequence transduction models are based on complex recurrent neural networks that include an encoder and a decoder. The best performing models connect the encoder and decoder through an attention.

A Vaswani 著 - 2017 - [被引用数: 11206](#) - 関連記事

挑戦的なタイトルを美しく回収 (NeurIPS 2017)

画像もTransformerでいけるのか...!?

## Image GPT

OpenAI

We find that, just as a large transformer model trained on language can generate coherent text, the same exact model trained on pixel sequences can generate coherent image completions and samples. By establishing a correlation between sample quality and image classification accuracy, we show that our best generative model also contains features competitive with top convolutional nets in the unsupervised setting.

[CODE](#) [ICML 2020 PAPER \(V1\)](#) [PAPER \(V2\)](#)

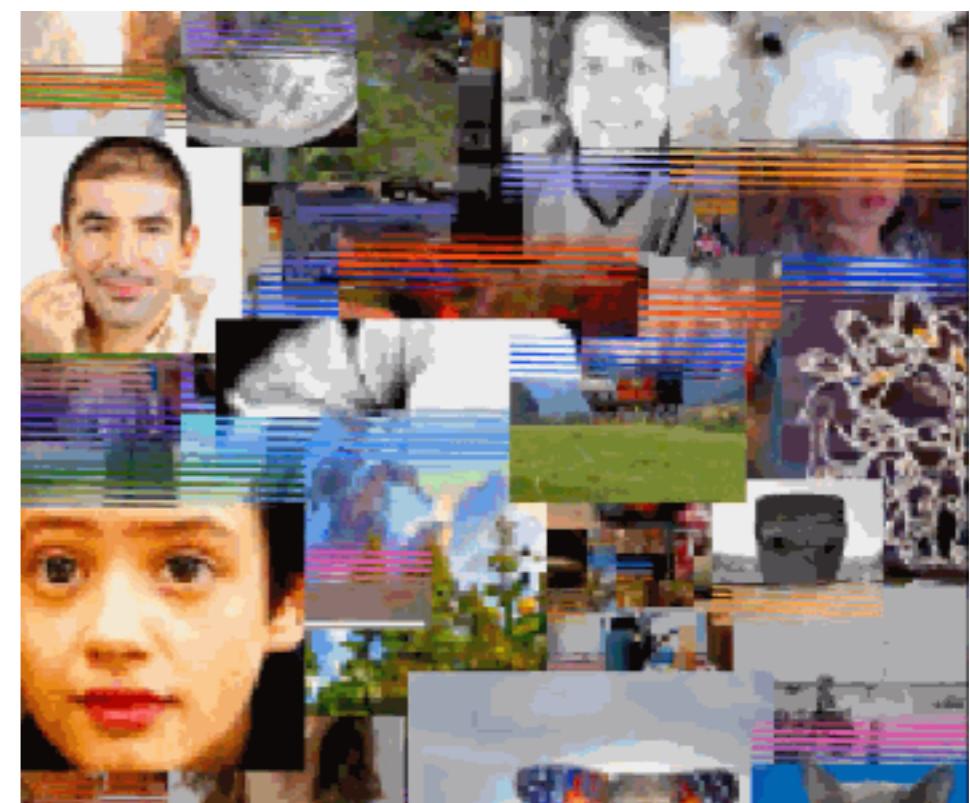
AI翻訳が人間超え、言葉の壁崩壊へ

## 第2部：技術動向

# トランスフォーマー時代到来 翻訳技術から汎用言語系AIに

2016年のニューフル機械翻訳(NMT)の実用化は、翻訳業界に衝撃を与え、ボケトークのような自動翻訳端末の市場拡大につながるなど、社会に大きなインパクトを与えた。ただし、翻訳技術や自然言語処理技術(NLP)分野では、その後も革命級のブレークスルーが相次いでいる。翻訳を含む言語系の人工知能(AI)が従来の常識を次々と塗り替え、ありえないベースで発展している。

日経エレクトロニクス 2019年9月号



# 注意機構 (Attention)

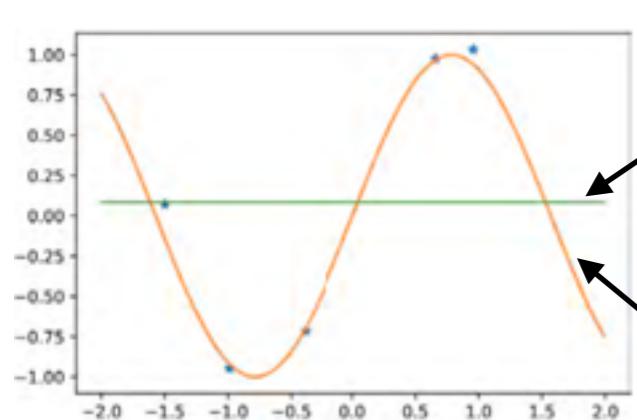
## A Tutorial on Attention in Deep Learning (ICML2019)

Alex Smola · Aston Zhang

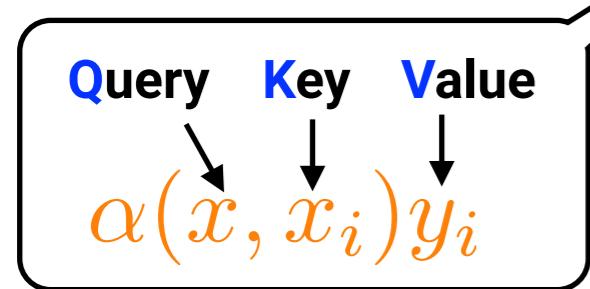
<https://icml.cc/Conferences/2019/ScheduleMultitrack?event=4343>

<https://d2l.ai>

The world's dumbest estimator



$$\hat{y} = \frac{1}{m} \sum_{i=1}^m y_i$$
$$\hat{y} = \sum_{i=1}^m \alpha(x, x_i) y_i$$



Better idea :  
Watson-Nadaraya Estimator  
(Watson, Nadaraya, 1964)

- • 重み関数の学習?  
• 重み付き Pooling?

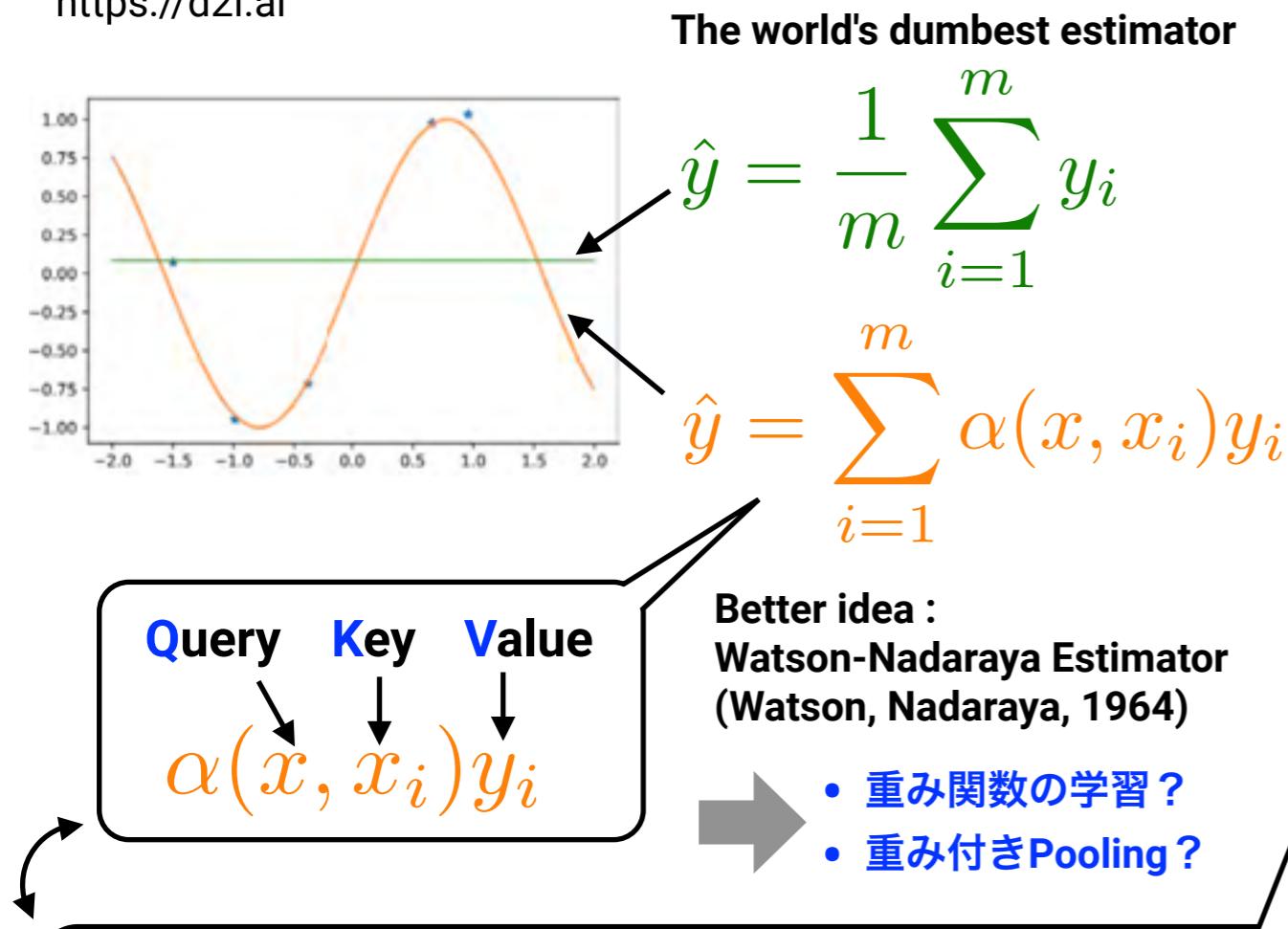
# 注意機構 (Attention)

## A Tutorial on Attention in Deep Learning (ICML2019)

Alex Smola · Aston Zhang

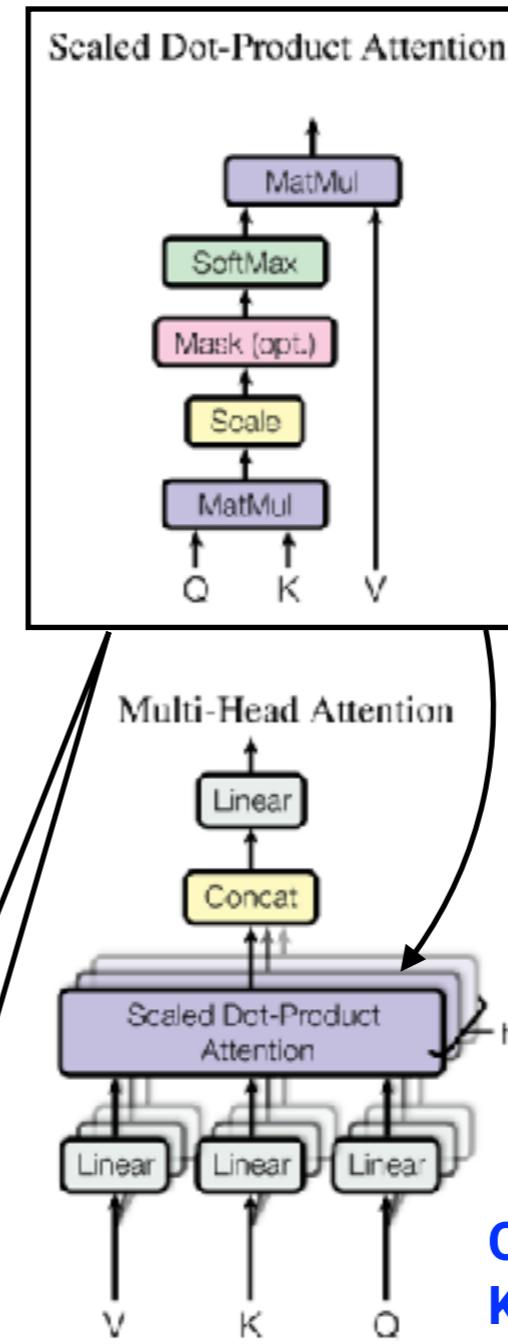
<https://icml.cc/Conferences/2019/ScheduleMultitrack?event=4343>

<https://d2l.ai>



### Transformerで使うScaled dot-product attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

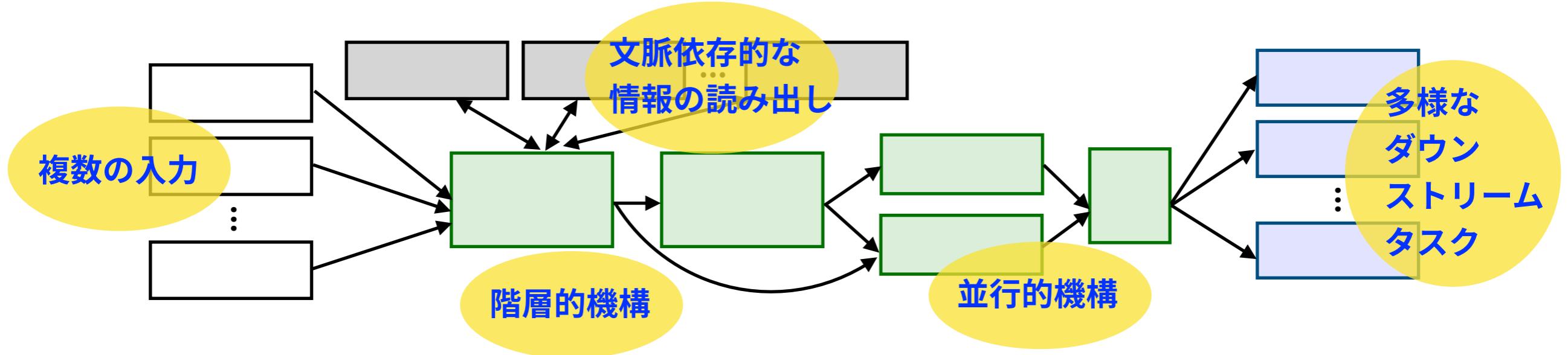


# 注意機構と構成的学习

## Compositionality (構成性)

高度なタスクの処理を構成要素・モジュールとその組合せ合成により実現する

- 例) • 単語をうまく組み合わせることで様々な意味の文章を無数に作り出す  
• 様々なモジュールをうまく組み合わせることで脳の高次機能を柔軟に作り出す

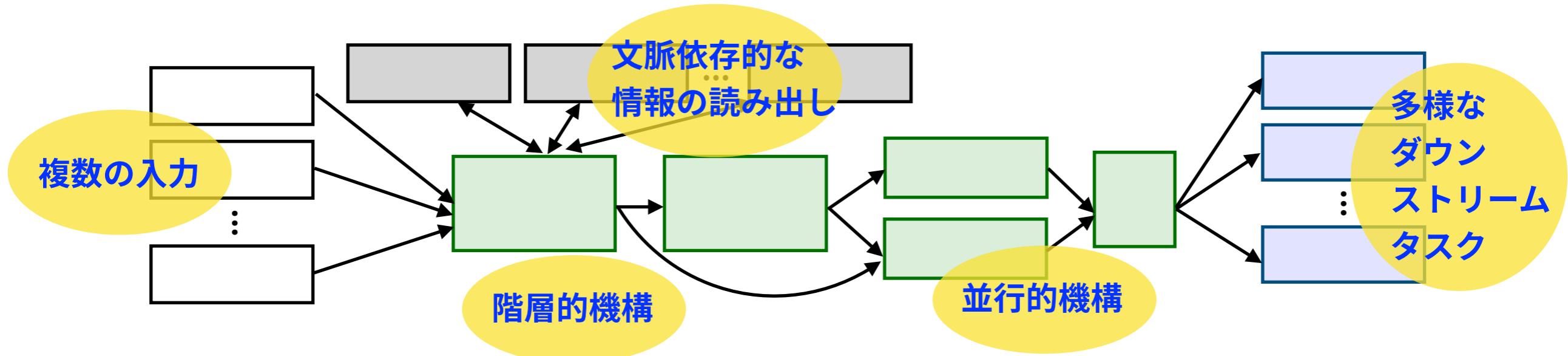


# 注意機構と構成的学习

## Compositionality (構成性)

高度なタスクの処理を構成要素・モジュールとその組合せ合成により実現する

- 例) • 単語をうまく組み合わせることで様々な意味の文章を無数に作り出す  
• 様々なモジュールをうまく組み合わせることで脳の高次機能を柔軟に作り出す



$$\begin{array}{c} \text{Query} \quad \text{Key} \quad \text{Value} \\ \downarrow \quad \downarrow \quad \downarrow \\ \alpha(x, x_i) y_i \end{array}$$

注意機構は覚えておいた情報をcontextに応じてとりだす  
辞書オブジェクト(query, key, value)とみなせる

→ 高度なタスクに必要

- 文脈に関して厚いこと (Context-thick)  
多様な種類の文脈において適切な動作を行う能力
- 重要な情報にいかにして注意(Attention)を向けるか

# 巨大な転移学習の有効性

## "小データ"の機械学習

- 本当にそのデータしかない場合
  - ドメイン知識や第一原理ハイブリッドなどで既知知見・背景知識によってモデル探索空間を作り込む必要がある
- 当該分野で何らかの大規模な参照データが利用できる場合
  - 技術的工夫でzero-shotやfew-shotな転移が可能なケースがある
- グラフ学習や強化学習で有効な転移学習が可能かは最大の関心の一つ？

Artificial intelligence / Machine learning

## OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.

by Will Douglas Heaven

MIT  
Technology  
Review

July 20, 2020

66,877 views | Jul 19, 2020, 06:56pm EDT

## GPT-3 Is Amazing—And Overhyped



Rob Toews Contributor



I write about the big picture of artificial intelligence.

Forbes



# 機械学習屋の関心

- Pre-trainingからSelf-training, Self-supervisedな表現学習へ
- 大規模転移によるfewshot/zeroshot学習  
(大規模テキスト BERT, GPT-3; 大規模画像 BiT, SimCLRv2)

Model Name	$n_{\text{params}}$	$n_{\text{layers}}$	$d_{\text{model}}$	$n_{\text{heads}}$	$d_{\text{head}}$	Batch Size	Learning Rate	パラメタ数
GPT-3 175B or “GPT-3”	175.0B	96	12288	96	128	3.2M	$0.6 \times 10^{-4}$	1750億

arXiv.org > cs > arXiv:2005.14165

Search... Help | Advanced

Computer Science > Computation and Language

(Submitted on 26 May 2020 (v1), last revised 22 Jul 2020 (this version, v4))

## Language Models are Few-Shot Learners

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei

arXiv.org > cs > arXiv:1912.11370

Search... Help | Advanced

Computer Science > Computer Vision and Pattern Recognition

(Submitted on 24 Dec 2019 (v1), last revised 5 May 2020 (this version, v3))

## Big Transfer (BiT): General Visual Representation Learning

Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Joan Puigcerver, Jessica Yung, Sylvain Gelly, Neil Houlsby

arXiv.org > cs > arXiv:2006.10029

Computer Science > Machine Learning

(Submitted on 17 Jun 2020)

## Big Self-Supervised Models are Strong Semi-Supervised Learners

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, Geoffrey Hinton

arXiv.org > cs > arXiv:2006.06882

Computer Science > Computer Vision and Pattern Recognition

(Submitted on 11 Jun 2020)

## Rethinking Pre-training and Self-training

Barret Zoph, Golnaz Ghiasi, Tsung-Yi Lin, Yin Cui, Hanxiao Liu, Ekin D. Cubuk, Quoc V. Le

arXiv.org > cs > arXiv:2006.07159

Computer Science > Computer Vision and Pattern Recognition

(Submitted on 12 Jun 2020)

## Are we done with ImageNet?

Lucas Beyer, Olivier J. Hénaff, Alexander Kolesnikov, Xiaohua Zhai, Alfon van den Oord

arXiv.org > cs > arXiv:1810.04805

Computer Science > Computation and Language

(Submitted on 11 Oct 2018 (v1), last revised 24 May 2019 (this version, v2))

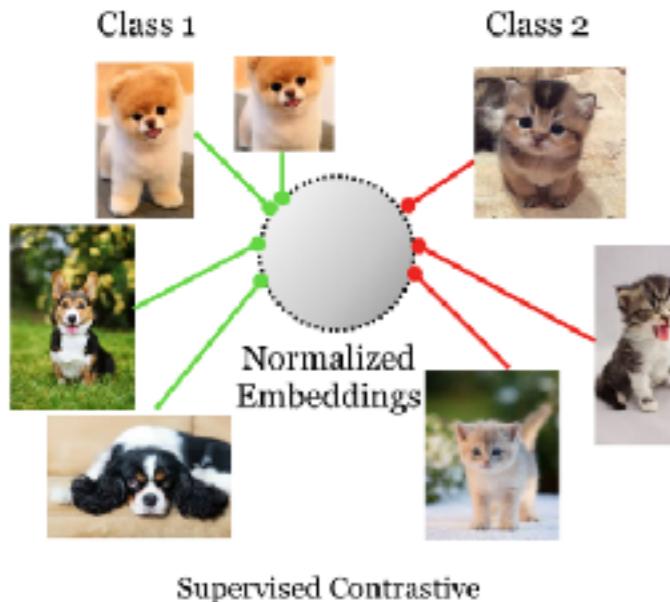
## BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

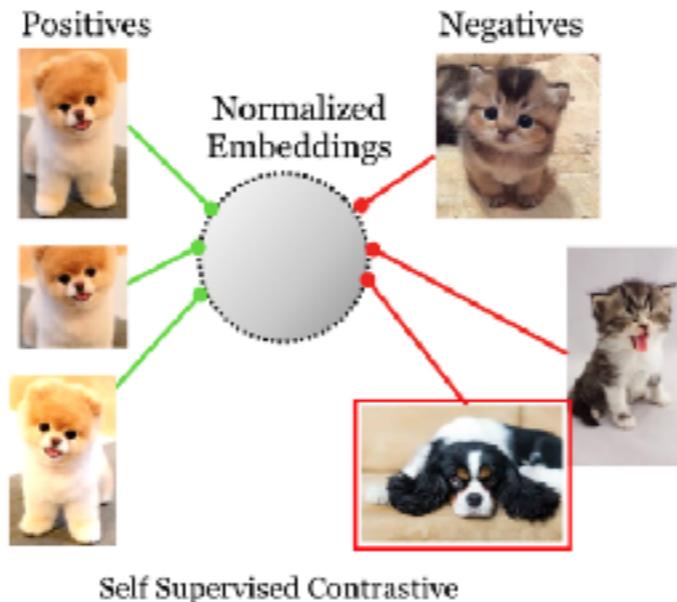
# Self-supervised learning (self training)

直接の目的ではないが自前で多数データが作れるpretextタスクで学習したものをして直接の目的であるダウンストリームタスクに転移 (word2vec, BERT, GPTなど多数)

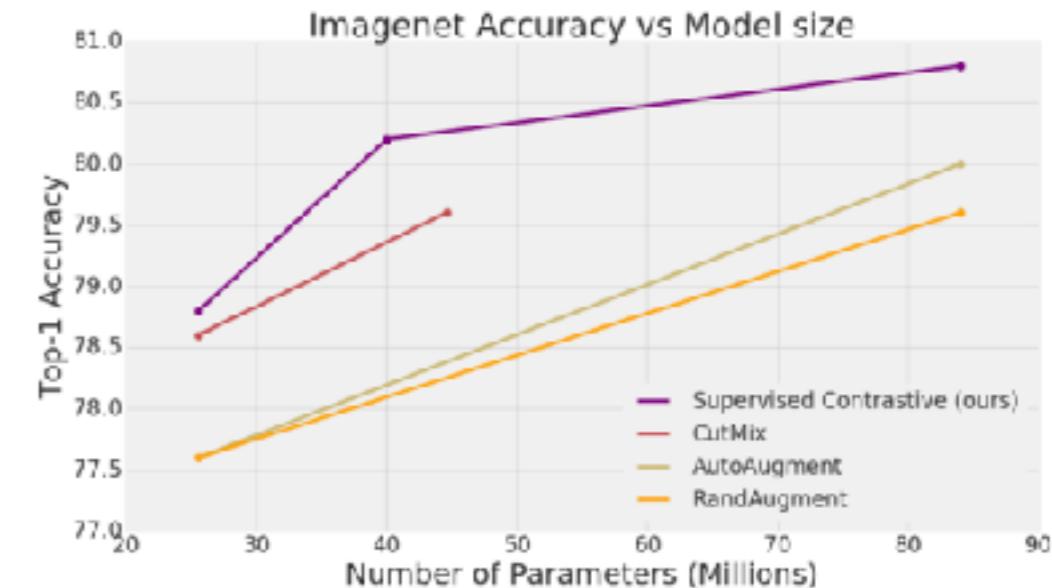
SupContrast (2020)



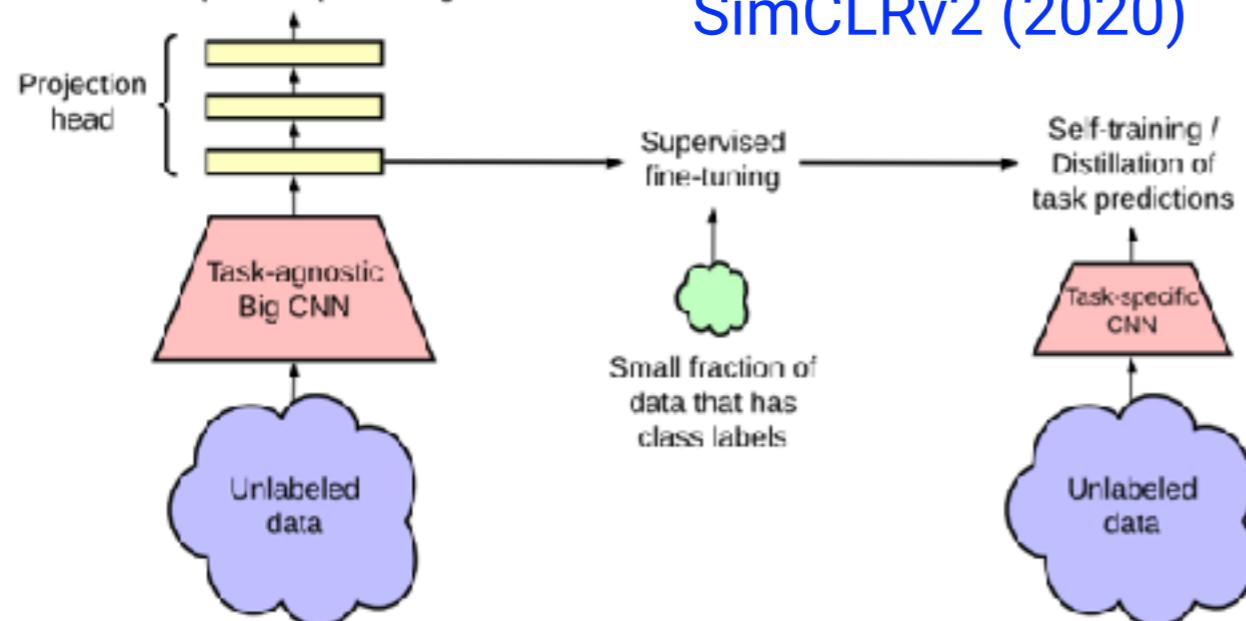
SimCLR (ICML2020)



arXiv:2004.11362 (2020)

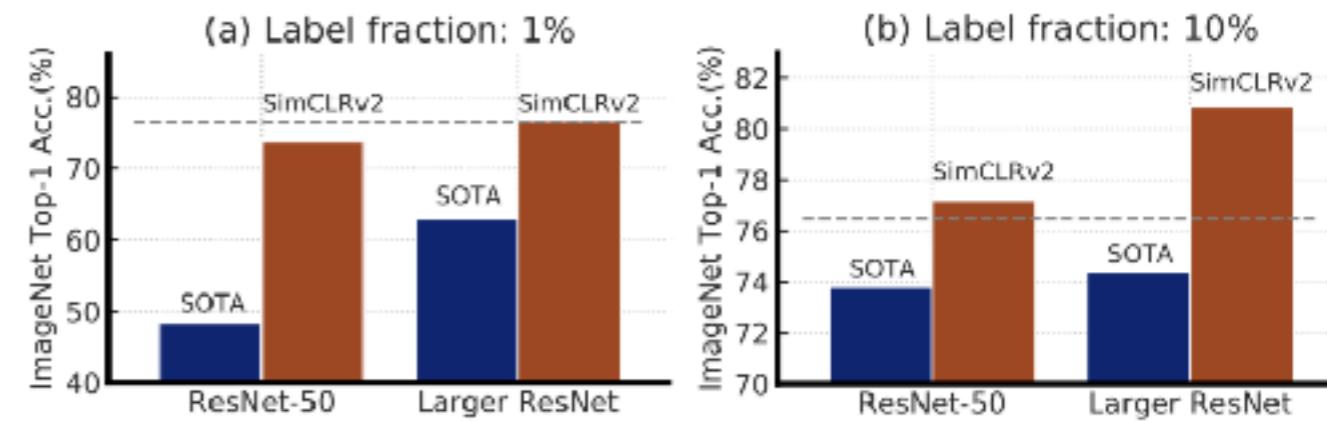


Unsupervised pretraining



SimCLRv2 (2020)

arXiv:2006.10029 (2020)



# 小サンプルで表現が学習できるのか？

Tenenbaum et al, **How to Grow a Mind: Statistics, Structure, and Abstraction**. *Science*, 331:6022, 1279-1285 (2011)

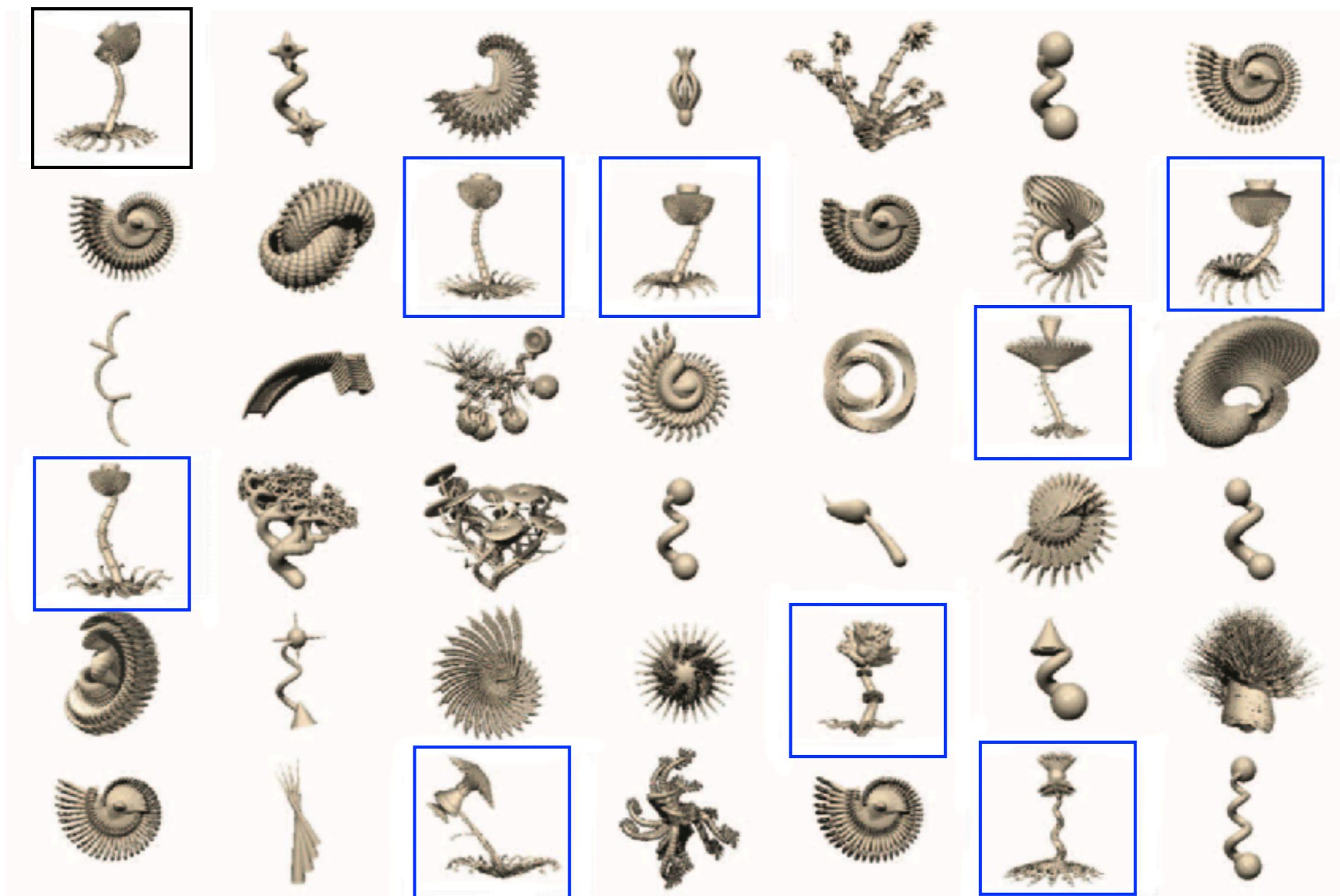
これと  
同じのは  
どれ？



# 小サンプルで表現が学習できるのか？

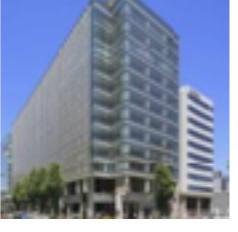
Tenenbaum et al, **How to Grow a Mind: Statistics, Structure, and Abstraction**. *Science*, 331:6022, 1279-1285 (2011)

これと  
同じのは  
どれ？



# Fewshot learning (n-shot k-wayのタスク)

## 2-shot 3-way fewshot learning

	support set						query set	
training set								
	1	1	2	2	3	3	?	?
⋮								
query set								
test set	1	1	2	2	3	3	?	?
⋮								
query set								
test set	1	1	2	2	3	3	?	?

# メタ学習

## MAML (ICML2017)

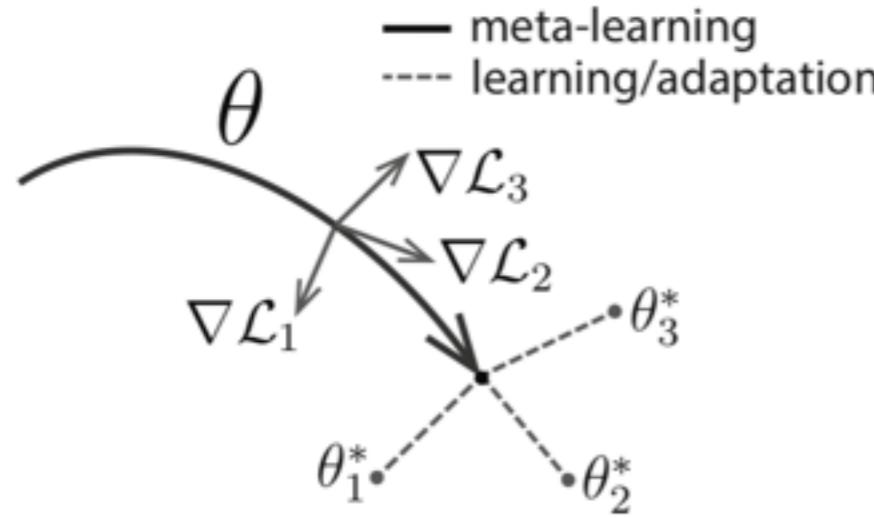


Figure 1. Diagram of our model-agnostic meta-learning algorithm (MAML), which optimizes for a representation  $\theta$  that can quickly adapt to new tasks.

---

### Algorithm 1 Model-Agnostic Meta-Learning

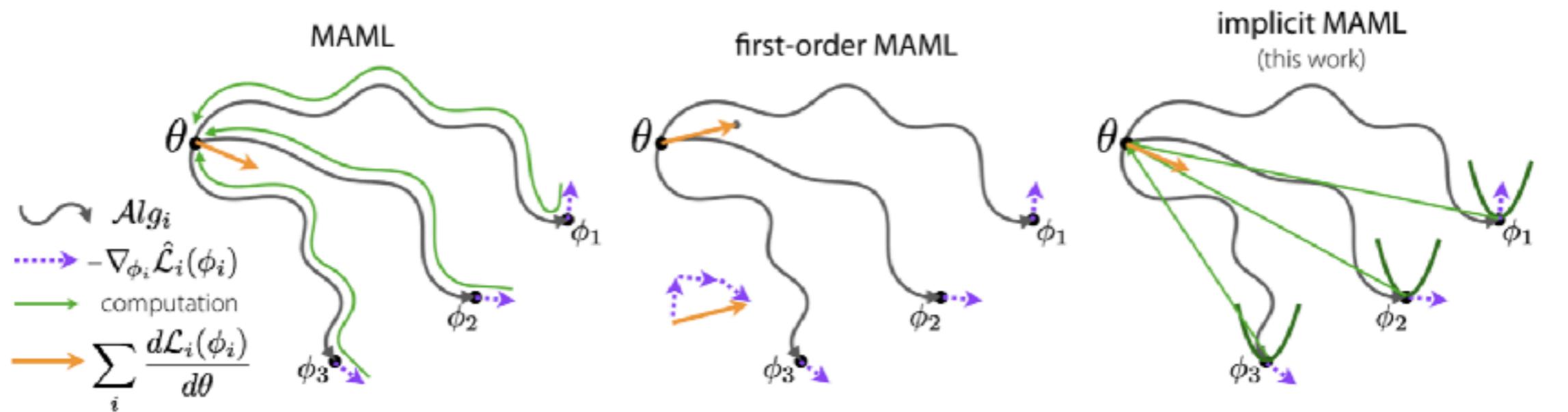
---

**Require:**  $p(\mathcal{T})$ : distribution over tasks

**Require:**  $\alpha, \beta$ : step size hyperparameters

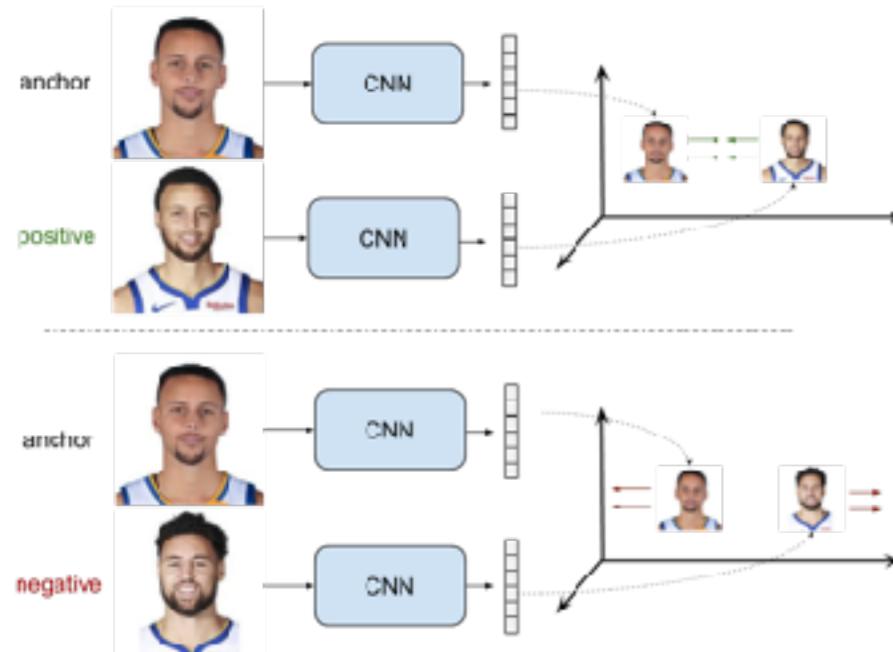
- 1: randomly initialize  $\theta$
  - 2: **while** not done **do**
  - 3:     Sample batch of tasks  $\mathcal{T}_i \sim p(\mathcal{T})$
  - 4:     **for all**  $\mathcal{T}_i$  **do**
  - 5:         Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$  with respect to  $K$  examples
  - 6:         Compute adapted parameters with gradient descent:  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$  ← (Red arrow)
  - 7:     **end for**
  - 8:     Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$  ← (Red arrow)
  - 9: **end while**
- 

## iMAML (NeurIPS 2019)

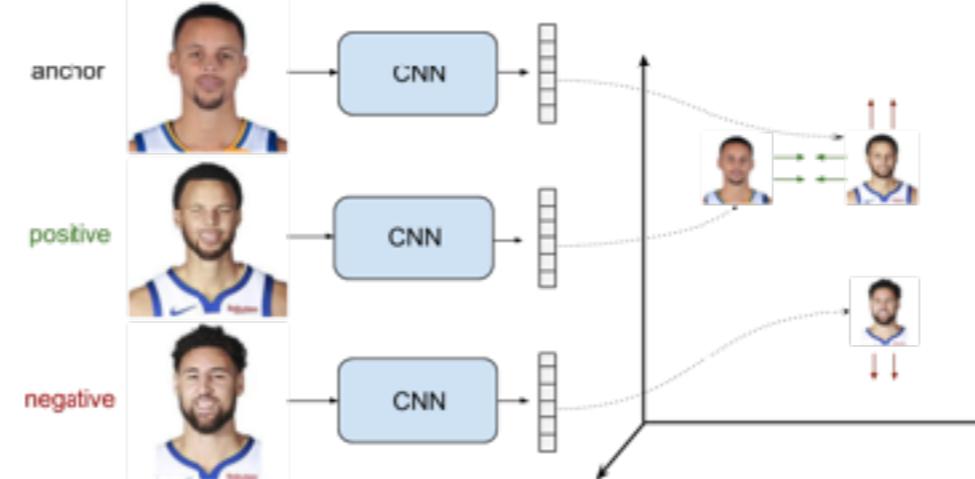


# 計量学習

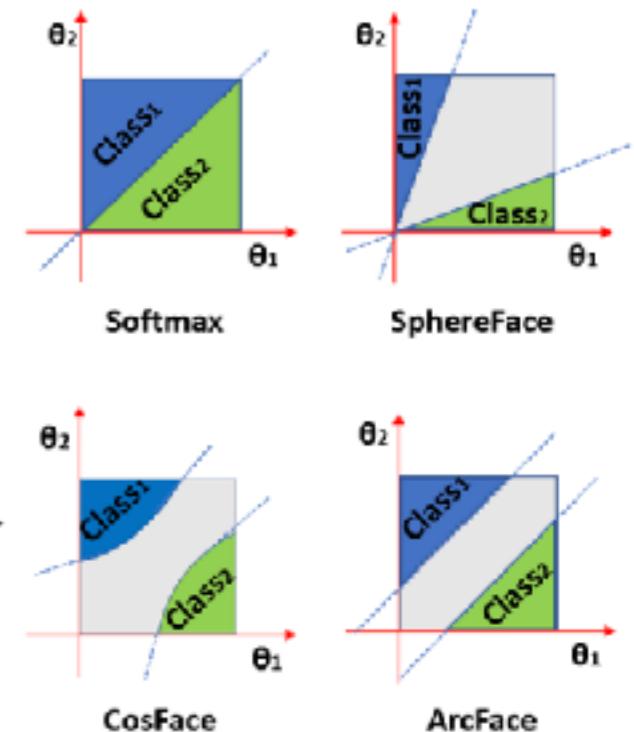
## Siamese Network (contrastive loss)



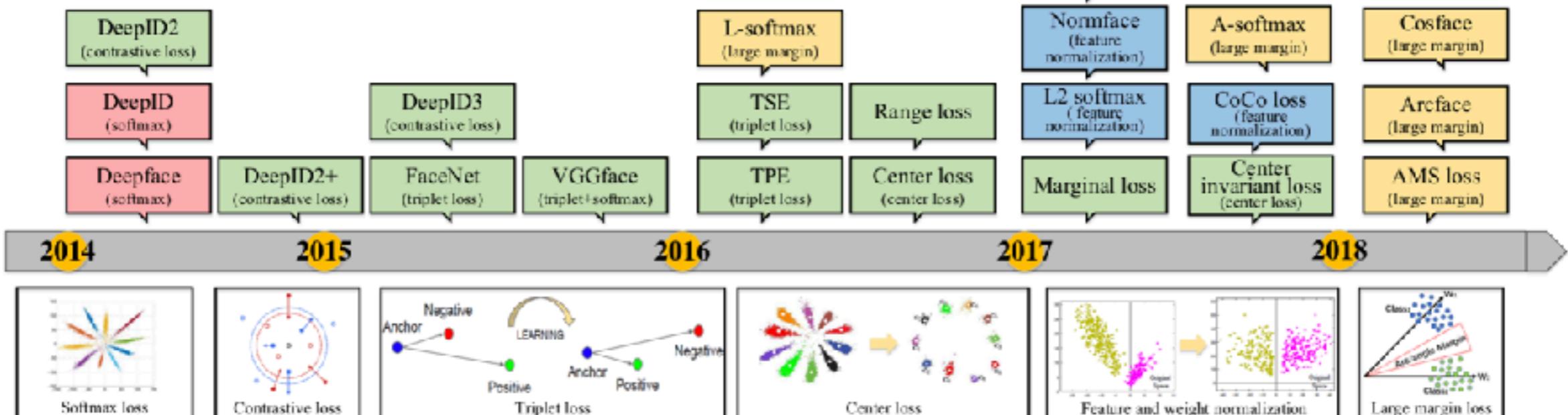
## Triplet Network (triplet loss)



## Angular margin loss



[https://gombru.github.io/2019/04/03/ranking\\_loss/](https://gombru.github.io/2019/04/03/ranking_loss/)



# 今日の話

## 1. 機械学習+化学の面白さとチャレンジ

- 😊 CVとNLPの間の独特設定 + 第一原理がある + 実検証できる
- 🤔 表現学習 + 小サンプル + 強いデータバイアス

## 2. 機械学習の関連トピックの(面白さの)紹介

### 1. 探索と訓練分布外の予測

新たなデータを取る探索戦略 (実験計画・最適化・進化戦略)  
外挿・内挿問題、良性の過適合

### 2. 学習の構成性と転移性

注意機構と転移学習・構成的学習

小サンプル学習 (転移学習、メタ学習、計量学習)

### 3. 演繹と帰納の融合

モデルベースの最適化・強化学習、第一原理計算との融合

# 陽な原理や知識の利用とデータ駆動の融合へ向けて

## 量子化学計算・分子動力学など計算化学(第一原理によるシミュレーション)の発展と限界

- 探索空間が組合せ的に巨大: 理論的に可能な経路の探索も組合せ爆発を起こす
- 計算時間・リソースが大きく計算できる系が限られる: 現実の系では何か妥協が必要
- 現実の化学反応の複雑さと不確定さ: 理論計算に入らない多様な要因が影響
- 現状の理論モデルの単純な仮定や不完全さ: 実験での反応と理論との大きな隔たり

こうした限界や制約をデータに基づくアプローチで打開できるか?

## Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data

Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee,  
Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar

# 陽な原理や知識の利用とデータ駆動の融合へ向けて

Ann. Rev. Phys. Chem. 71:361–90 (2020)

Nat. Rev. Chem. 4: 347–358 (2020)



## *Annual Review of Physical Chemistry* Machine Learning for Molecular Simulation

Frank Noé,<sup>1,2,3</sup> Alexandre Tkatchenko,<sup>4</sup>  
Klaus-Robert Müller,<sup>5,6,7</sup> and Cecilia Clementi<sup>1,3,8</sup>

PNAS (2020)

NATURE REVIEWS | CHEMISTRY

PERSPECTIVES

## Exploring chemical compound space with quantum-based machine learning

O. Anatole von Lilienfeld, Klaus-Robert Müller and Alexandre Tkatchenko

**Abstract** | Rational design of compounds with specific properties requires understanding and fast evaluation of molecular properties throughout chemical compound space — the huge set of all potentially stable molecules. Recent advances in combining quantum-mechanical calculations with machine learning

## The frontier of simulation-based inference

Kyle Cranmer<sup>a,b,1</sup> , Johann Brehmer<sup>a,b</sup> , and Gilles Louppe<sup>c</sup>

<sup>a</sup>Center for Cosmology and Particle Physics, New York University, New York, NY 10003; <sup>b</sup>Center for Data Science, New York University, New York, NY 10011;  
and <sup>c</sup>Montefiore Institute, University of Liège, B-4000 Liège, Belgium

Edited by Jitendra Malik, University of California, Berkeley, CA, and approved April 10, 2020 (received for review November 4, 2019)

**Many domains of science have developed complex simulations to describe phenomena of interest. While these simulations provide high-fidelity models, they are poorly suited for inference and lead to challenging inverse problems. We review the rapidly developing field of simulation-based inference and identify the forces giving additional momentum to the field. Finally, we describe how the frontier is expanding so that a broad audience can appreciate the profound influence these developments may have on science.**

the simulator—is being recognized as a key idea to improve the sample efficiency of various inference methods. A third direction of research has stopped treating the simulator as a black box and focused on integrations that allow the inference engine to tap into the internal details of the simulator directly.

Amidst this ongoing revolution, the landscape of simulation-based inference is changing rapidly. In this review we aim to provide the reader with a high-level overview of the basic ideas

# 陽な原理や知識の利用とデータ駆動の融合へ向けて

## Neural Abstract Machines & Program Induction

<https://uclnlp.github.io/nampi/>

Machine intelligence capable of **learning complex procedural behavior, inducing (latent) programs, and reasoning with these programs** is a key to solving artificial intelligence. Recently, there have been a lot of success stories in the deep learning community related to learning neural networks capable of **using trainable memory abstractions**.

- Differentiable Neural Computers / Neural Turing Machines (Graves+ 2014)
- Memory Networks (Weston+ 2014)
- Pointer Networks (Vinyals+ 2015)
- Neural Stacks (Grefenstette+ 2015, Joulin+ 2015)
- Hierarchical Attentive Memory (Andrychowicz+ 2016)
- Neural Program Interpreters (Reed+ 2016)
- Neural Programmer (Neelakantan+ 2016)
- DeepCoder (Balog+ 2016)
- :

手続き的・記号的操作も学習できるプログラムとして扱えるようになってきた



Computer-Aided Synthetic Planning

International Edition: DOI: 10.1002/anie.201506101  
German Edition: DOI: 10.1002/ange.201506101

## Computer-Assisted Synthetic Planning: The End of the Beginning

Sara Szymkuć, Ewa P. Gajewska, Tomasz Klucznik, Karol Molga, Piotr Dittwald, Michał Startek, Michał Bajczyk, and Bartosz A. Grzybowski\*

Angew. Chem. Int. Ed. 2016, 55, 5904–5937



AI-Assisted Synthesis Very Important Paper

International Edition: DOI: 10.1002/anie.201912083  
German Edition: DOI: 10.1002/ange.201912083

## Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning

Tomasz Badowski, Ewa P. Gajewska, Karol Molga, and Bartosz A. Grzybowski\*

Angew. Chem. Int. Ed. 2019, 58, 1–7

# 現状の「人工知能技術」が抱える明らかな限界

- Deep learning techniques thus far have proven to be data hungry, shallow, brittle, and limited in their ability to generalize (Marcus, 2018)
- Current machine learning techniques are data-hungry and brittle—they can only make sense of patterns they've seen before. (Chollet, 2020)
- A growing body of evidence shows that state-of-the-art models learn to exploit spurious statistical patterns in datasets... instead of learning meaning in the flexible and generalizable way that humans do. (Nie et al., 2019)
- Current machine learning methods seem weak when they are required to generalize beyond the training distribution, which is what is often needed in practice. (Bengio et al., 2019)

ただし適切な目的へ応用すれば極めて有効な技術だと実証してきた！



Feb 19, 2020, 08:00am

## In Praise Of Boring AI (A.K.A. Machine Learning)

JC Schutterle Forbes Councils Member  
Forbes Technology Council COUNCIL POST | Paid Program

# 現状の「人工知能技術」が抱える明らかな限界

- Deep learning techniques thus far have proven to be data hungry, shallow, brittle, and limited in their ability to generalize (Marcus, 2018)
- Current machine learning techniques are data-hungry and brittle—they can only make sense of patterns they've seen before. (Chollet, 2020)
- A growing body of evidence shows that statistical methods can find spurious statistical patterns in datasets... in a much less flexible and generalizable way than humans (Bengio et al., 2019)
- Current machine learning methods seem weak at generalizing beyond the training distribution, in practice. (Bengio et al., 2019)

ただし適切な目的へ応用すれば極めて有効！



Feb 19, 2020, 08:00am

## In Praise Of Boring AI (And Machine Learning)

JC Schutterle Forbes Councils Member  
Forbes Technology Council COUNCIL POST | Paid Program



From AAAI-20 Oxford-Style Debate

# 合理論(計算化学)と経験論(機械学習)の融合に向けて

## Theory-driven 【合理論】

- 対象現象の複雑化
- シミュレーション技法も複雑化
- "経験的に決める"パラメタや初期値
- 汎関数、交換相関項の設計

## (人工知能分野)

- 知識ベースと論理推論(記号AI)の限界
- 厳密推論や探索の計算爆発(NP困難性)
- 大量データの知識化の問題
- 制約プログラミングや組合せ最適化

→ データ同化、模倣学習、論理合成、etc

→ 表現学習、モデルベース最適化・強化  
学習、メタ学習、生成モデル、etc

→ 新たな方法論へ？

## Data-driven 【経験論】

- 小サンプル・低カウントの問題
- 帰納バイアスのモデルエンコード
- 外挿の低信頼性と探索
- Blackbox性・解釈性の問題

## (人工知能分野)

- Data-Driven手法(機械学習)と人間の論理的思考との大きなギャップ
- Dataがない領域の探索や「ひらめき」
- モデル適用範囲と信頼性・安全性

# 今日の話(まとめ)

## 1. 機械学習+化学の面白さとチャレンジ

- 😊 CVとNLPの間の独特設定 + 第一原理がある + 実検証できる
- 🤔 表現学習 + 小サンプル + 強いデータバイアス

## 2. 機械学習の関連トピックの(面白さの)紹介

### 1. 探索と訓練分布外の予測

新たなデータを取る探索戦略 (実験計画・最適化・進化計算)  
外挿・内挿問題、良性の過適合

### 2. 学習の構成性と転移性

注意機構と転移学習・構成的学習

小サンプル学習 (転移学習、メタ学習、計量学習)

### 3. 演繹と帰納の融合

モデルベースの最適化・強化学習、第一原理計算との融合