

自己紹介

機械学習・機械発見とデータ中心的自然科学

2021年12月22日

瀧川 一学

ichigaku.takigawa@riken.jp

理化学研究所 革新知能統合研究センター@京阪奈ATR
北海道大学 化学反応創成研究拠点 (ICReDD)





たきがわ いちがく
瀧川 一学

<https://itakigawa.github.io>

機械学習を研究している技術屋デス！

- うどん県高松市生まれ
- 1995～2004 北海道大 (工学研究科)
2004 博士(工学) "劣決定信号源分離の解の理論分析"
- 2005～2011 京都大 (化学研究所/薬学研究科)
バイオインフォマティクスセンター 助教
- 2012～2018 北海道大 (情報科学研究科)
大規模知識処理研究室 准教授
2015～2018 JSTさきがけ (材料インフォマティクス)
- 2019～ 北海道大学 化学反応創成研究拠点(ICReDD)
2019～ 理化学研究所 革新知能統合研究センター(AIP)

普段は京大iPS細胞研との連携ラボ@京阪奈ATRに勤務
(iPS細胞連携医学的リスク回避チーム)



たきがわ いちがく
瀧川 一学

<https://itakigawa.github.io>

と同時に機械学習のユーザでもあります！

- うどん県高松市生まれ
 - 1995～2004 北海道大 (工学研究科)
2004 博士(工学) "劣決定信号源分離の解の理論分析"
 - 2005～2011 京都大 (化学研究所/薬学研究科)
バイオインフォマティクスセンター 助教
 - 2012～2018 北海道大 (情報科学研究科)
大規模知識処理研究室 准教授
2015～2018 JSTさきがけ (材料インフォマティクス)
 - 2019～ 北海道大学 化学反応創成研究拠点(ICReDD)
2019～ 理化学研究所 革新知能統合研究センター(AIP)
- 普段は京大iPS細胞研との連携ラボ@京阪奈ATRに勤務
(iPS細胞連携医学的リスク回避チーム)



たきがわ いちがく
瀧川 一学

<https://itakigawa.github.io>

と同時に機械学習のユーザでもあります！

- うどん県高松市生まれ
 - 1995～2004 北海道大 (工学研究科)
2004 博士(工学) "劣決定信号源分離の解の理論分析"
 - 2005～2011 京都大 (化学研究所/薬学研究科)
バイオインフォマティクスセンター 助教
 - 2012～2018 北海道大 (情報科学研究科)
大規模知識処理研究室 准教授
2015～2018 JSTさきがけ (材料インフォマティクス)
 - 2019～ 北海道大学 化学反応創成研究拠点(ICReDD)
2019～ 理化学研究所 革新知能統合研究センター(AIP)
- 普段は京大iPS細胞研との連携ラボ@京阪奈ATRに勤務
(iPS細胞連携医学的リスク回避チーム)

技術研究上の関心：離散構造を伴う機械学習

離散構造

=集合と論理、関係、組合せ、系列、木、グラフ、代数系、言語、etc

←→ 機械学習

連続的な確率統計

対象に離散構造

- 分子構造(1D~4D)
- ゲノム/エピゲノム
- 糖鎖
- ペプチド/タンパク質
- 原子の周期配置
- 論理や命題
- アイテムセット
- 言語/テキスト

関係に離散構造

- 化学反応経路ネットワーク
(代謝経路/KEGGなど)
- 遺伝子ネットワーク
- 分子相互作用ネットワーク
(PPI、DTI、DDI、etc)
- 因子組合せ
- 論理制約
- 変数間の交互作用

モデルに離散構造

- 決定木アンサンブル
(木や森やDAG)
- 計算グラフ・自動微分系
(深層学習)
- 階層構造
- 分解構造
- 構造制約
- 制約パターン列挙

決定木アンサンブルと自動微分フレームワーク(深層学習)

2021年度 L-B2: 機械学習とデータサイエンスの現代的手法

決定木に基づくアンサンブル学習

1. 決定木とは？

- (1) 決定木の歴史と基本
- (2) 発展：モデル木と多変量木

2. 決定森(決定木アンサンブル)モデル

- (1) Random Forest, Extra Trees
- (2) 勾配ブースティング木
(GBDT, XGBoost, LightGBM)

3. 決定木モデルによるデータ分析

- (1) 変数重要度 (MDIとPFI), PDP, SHAP
- (2) 予測値(回帰)の信頼区間推定

機械学習と自動微分

2021年1月22日(金)：最適化法 第15回

4講時 14:55 - 16:25

たきがわ いちがく
瀧川一学

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

理化学研究所 革新知能統合研究センター

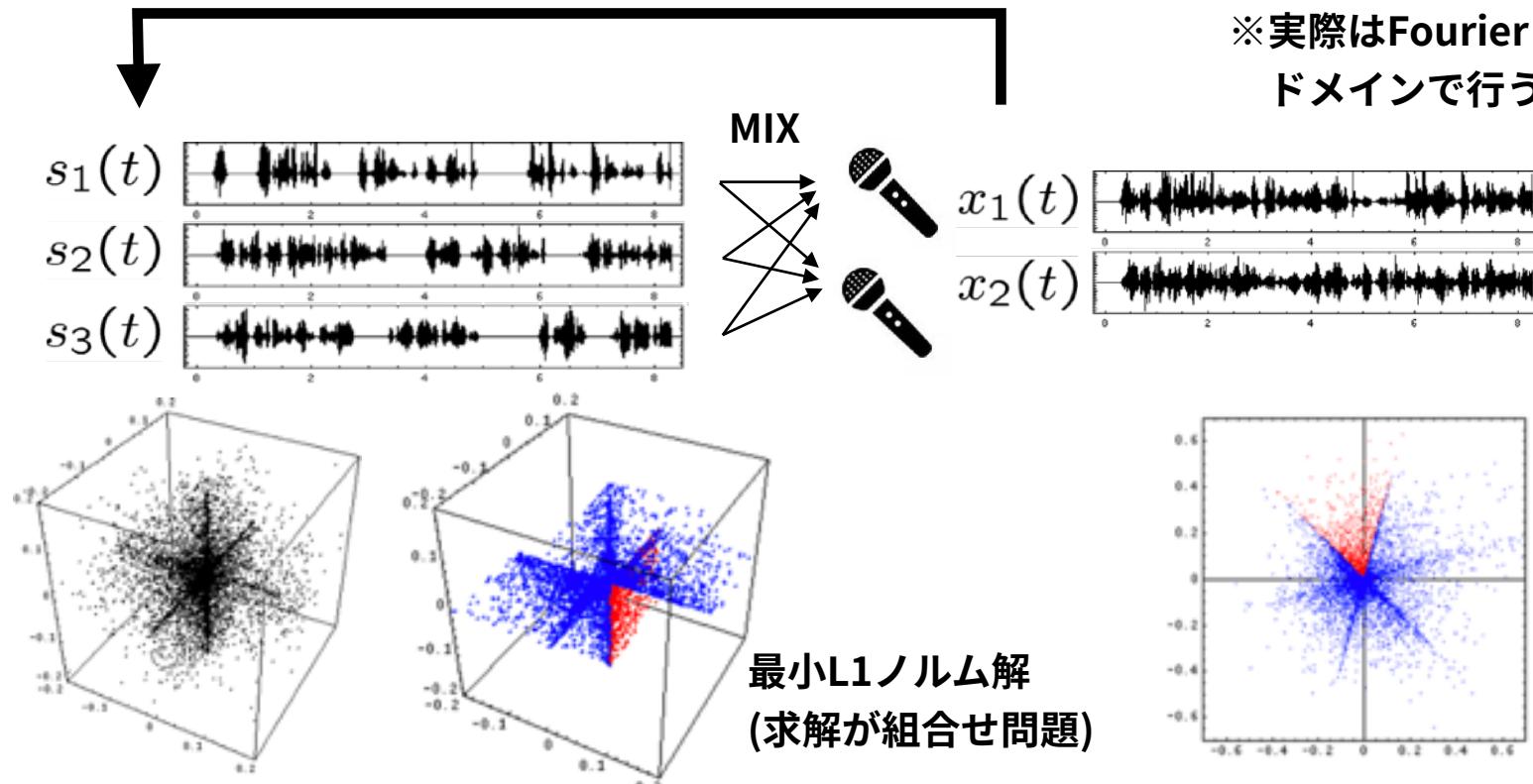
統計数理研究所 リーディングDAT(Data Analytics Talents)講座
L-B2 機械学習とデータサイエンスの現代的手法

同志社大学「最適化法」第15回(ゲストトーク)

<https://www.slideshare.net/itakigawa/ss-241690171>

学位を取った研究テーマ

劣決定信号源分離の解の性質の理論分析と効率的求解法



Theorem 3: If $n \geq 2$, the expectation of $\|\mathbf{s}\|_1 - \|\hat{\mathbf{s}}\|_1$ is bounded above as

$$\mathbb{E} \{ \|\mathbf{s}\|_1 - \|\hat{\mathbf{s}}\|_1 \} \leq 2n\beta \frac{w \log(-nw)}{1+w}, \quad w = W_{-1} \left(-\frac{1}{en} \right)$$

where function W is the Lambert W function defined by $W(x)e^{W(x)} = x$, and the subscript -1 indicates the second real branch of $W(x)$, which indicates $w < -1$.

Theorem 5:

$$\mathbb{E} \{ \|\mathbf{s} - \mathbf{s}^+\|_1 \} \leq \gamma \beta n$$

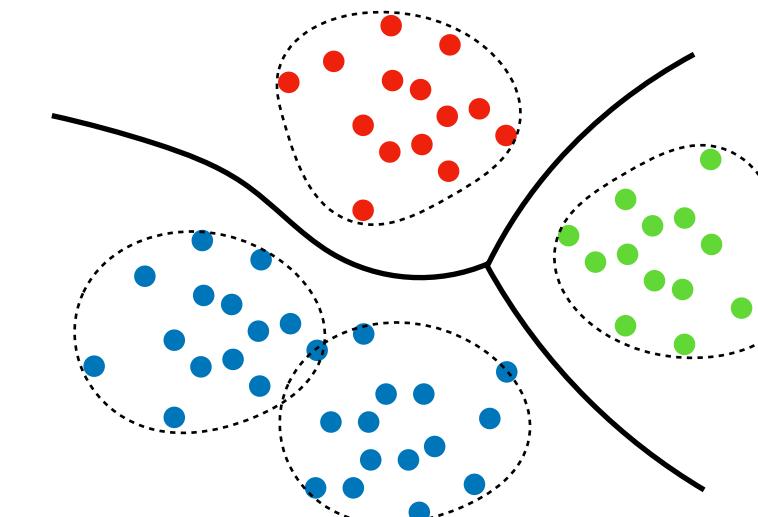
and the ℓ_1 -error of the Moore–Penrose solution has bounds

$$\mathbb{P} \{ \|\mathbf{s} - \mathbf{s}^+\|_1 \geq \xi \} \leq \frac{1}{\xi} \gamma \beta n$$

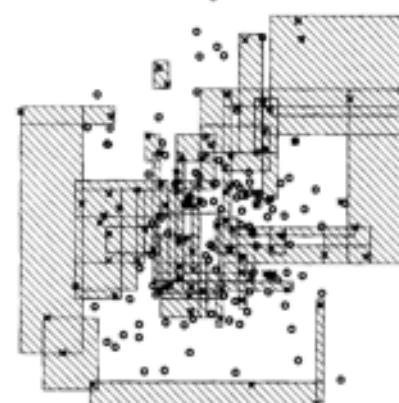
$$\mathbb{P} \{ \|\mathbf{s} - \mathbf{s}^+\|_1 \geq \xi \} \leq \inf_{0 < \theta < \frac{1}{\gamma \beta}} \{ \exp(-\theta \xi) (1 - \theta \gamma \beta)^{-n} \}$$

where $\gamma = \|\mathbf{I} - \mathbf{A}^+ \mathbf{A}\|_1$, and β is a parameter of the Laplace distribution (3).

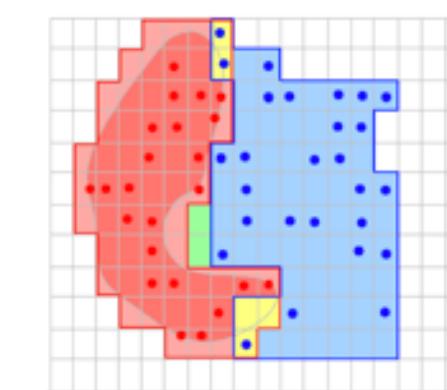
凸集合による部分クラス被覆による機械学習



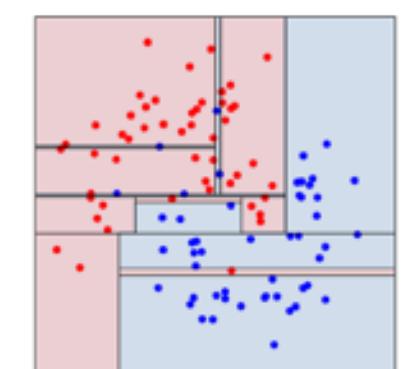
→ 「点集合 X から一意に定まる凸集合 $S(X)$ が他のクラスの点を含まないような極大な部分集合 X の列挙」として解く



Subclass method
by Mineichi Kudo



Logical analysis of data (LAD)
by Peter L. Hammer



Decision Trees / CART
by BFOS

京大化学研究所バイオインフォマティクスセンター



[English | Japanese]

センターの概要

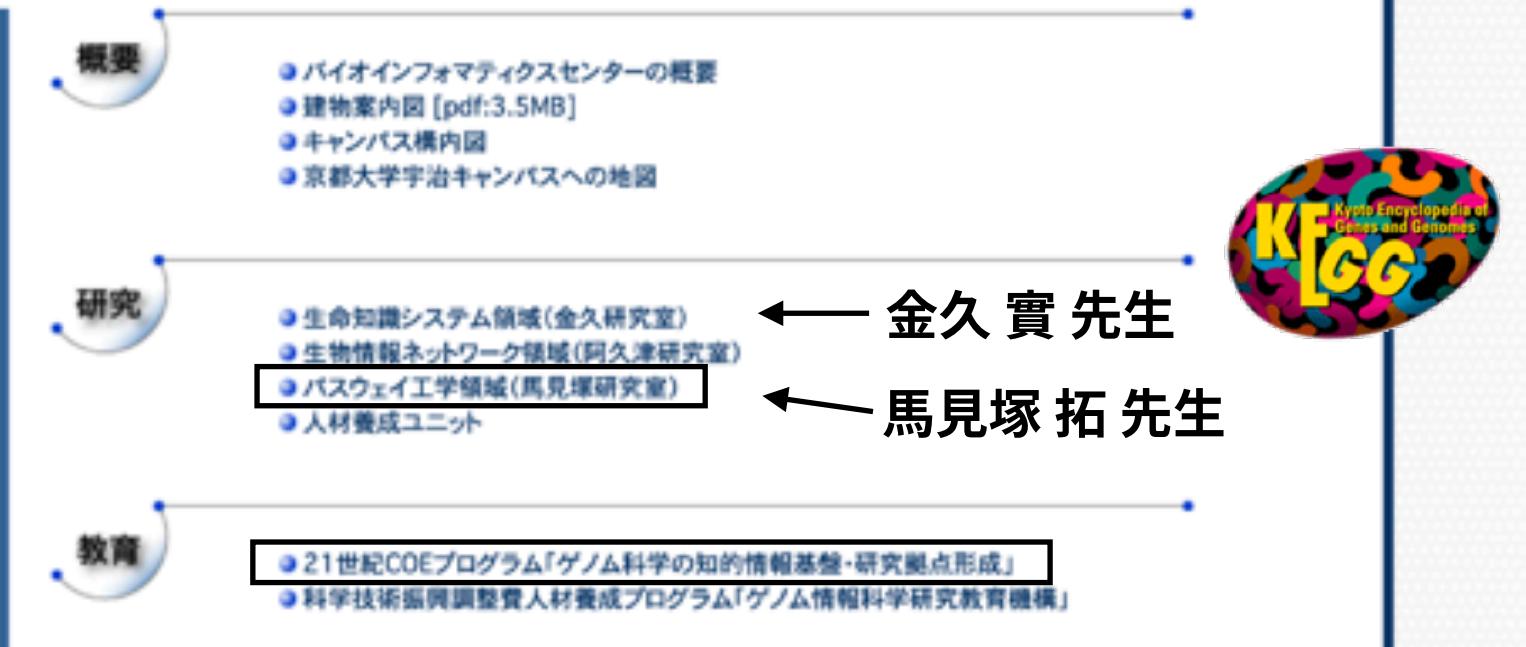
研究

- 生命知識システム領域（金久研究室）
- 生物情報ネットワーク領域（阿久津研究室）
- パスウェイ工学領域（宮野研究室）
- プロテオームインフォマティクス領域（馬見塚研究室）－日本SGI寄附研究部門
- ゲノムインフォマティクス領域（藤研究室）－人材養成ユニット

教育

- ゲノム情報科学研究教育機構（人材養成プログラム）

宮野 悟 先生
馬見塚先生の
博士論文指導教員



科研費特定領域領域番号294, 略称:発見科学 (1998-2000)



研究課題/領域番号	10143106	サマリー
研究種目	特定領域研究(A)	
配分区分	補助金	
研究機関	九州大学	
研究代表者	有川 節夫 九州大学、大学院・システム情報科学研究院、教授 (40037221)	← 有川節夫 先生
研究分担者	丸岡 恵 東北大学、大学院・情報科学研究科、教授 (50005427) 佐藤 泰介 東京工業大学、大学院・情報理工学研究科、教授 (90272690) 佐藤 雅彦 京都大学、大学院・情報学研究科、教授 (20027387) 金田 康正 東京大学、情報基盤センター、教授 (90115551) 宮野 悟 東京大学、医科学研究所、教授 (50128104)	← 宮野悟 先生
研究期間(年度)	1998 – 2000	
研究課題ステータス	完了 (2001年度)	
配分額 *注記	66,400千円 (直接経費: 66,400千円) 2001年度: 3,000千円 (直接経費: 3,000千円) 2000年度: 15,300千円 (直接経費: 15,300千円) 1999年度: 19,100千円 (直接経費: 19,100千円) 1998年度: 29,000千円 (直接経費: 29,000千円)	
キーワード	発見科学 / 知識科学 / データマイニング / データベース / 科学的発見の論理 / アブダクション / 機械学習 / ネットワークエージェント / 知識発見	

機械学習から機械発見へ(1996)

有川節夫, 機械学習から機械発見へ. 人工知能学会誌, 11巻(1996)6号

https://doi.org/10.11517/jjsai.11.6_865

1.はじめに

—創造工学を機械化できないか—

1968年から1972年の頃であったと思う。北川敏男先生や国沢清典先生、森口繁一先生達の企画で日本科学技術連盟主催のセミナー(講習会)が定期的に開催されていた。その一つに「創造工学」があった。KJ法や等価変換論といった代表的な創造工学の手法の提唱者自身による講演があり、私も、北川先生の好意で出席させてもらった。その当時計算理論の研究をしていたので、このような講演が非常に新鮮に感じられ、また、そうした創造工学の手法が、非常に主観的で精神論的なものに感じられた。

もう少し客観的に機械的にこうした手法を実現できないものか、どのような研究はきっと非常に重要になり盛んになるはずだ、というような生意気なことを北川先生に話したように記憶している。

10.ゲノム情報学

1987年の暮に2年半に及ぶドイツ滞在を終えた宮野悟氏が助教授として着任した。その間に彼は、P完全性と並列アルゴリズムに関する大作を完成させていた。人工知能における論理と推論や学習をテーマにして研究教育をしていた私に対して、彼は、「計算量を考慮していないAIの研究は、価値がない。」といっていた。私は、「コンピュータにできるかどうかわからない彼方にある問題をどうにかしてコンピュータにのっけることが、AIの重要な仕事だ。二本差しが怖くて田楽が食えるか。計算量が怖くてAIがやれるか。」などと対応していた。しかし、彼がいうように、計算量の問題は、実際にシステムを構築するときには避けて通れない問題であり、また、PAC学習やMAT学習は、計算量を中心に据えたパラダイムでもある。篠原歩君が入学したのを機に宮野氏も計算論的学習理論に関心を持つようになり、先に述べた教示の理論を2人で展開し始めた。

1991年から文部省科学研究費重点領域研究「ゲノム情報」がスタートした。それに私の研究室も参加してほしい旨の誘いが複数の人からあった。この分野が情報科学としても重要であることは認識していたので、この誘いに乗りたかった。しかし、私自身は、大須賀先生を代表とする別の重点領域研究「知識科学」の班長をしていて、直接参加することができなかつた。そこで、宮野氏にお願いした。快く引き受けてくれたが、我々は、遺伝子とかアミノ酸とかタンパク質といつても、何の知識もなかった。そこで、分子生物学の専門家である久原哲先生(現九州大学教授)に加わってもらい、その方面的勉強会を持ちながら、先生の出された問題を解く形で研究を始めた。

ゲノム解析に伴う大量知識情報処理の研究(1991-1995)

ゲノム解析に伴う大量知識情報処理の研究

研究課題

研究課題/領域番号	04261101	サマリー
研究種目	重点領域研究	
配分区分	補助金	
研究機関	京都大学	
研究代表者	金久 實 京都大学, 化学研究所, 教授 (70183275)	
研究分担者	西岡 幸明 京都大学, 化学研究所, 助教授 (80026559) 後藤 修 埼玉県立がんセンター, 主任研究員 (40142111) 米澤 明彦 東京大学, 理学系研究科, 教授 (00133116) 美宅 成樹 東京農工大学, 工学部, 教授 (10107542) 高木 利久 東京大学, 医科学研究所, 教授 (30110636) 小笠原 直毅 豊良先端科学技術大, バイオサイエンス科, 教授 (10110553) 宮野 信 九州大学, 理学部, 教授 (50128104) 久原 哲 九州大学, 農学研究科, 助教授 (00153320) 今井 浩 東京大学, 理学部, 助教授 (80183010)	
研究期間(年度)	1991 - 1995	
研究課題ステータス	完了(1995年度)	

機械学習から機械発見へ(1996)

有川節夫, 機械学習から機械発見へ. 人工知能学会誌, 11巻(1996)6号

https://doi.org/10.11517/jjsai.11.6_865

12. 機械発見の論理

ゲノム情報学では、分子生物学のある問題を対象にして、実験データからの学習アルゴリズムを使った科学的知識発見を行った。また、有村君らの研究では、データベースにおける知識の発見、あるいはデータマイニングの一つの手法が示された。これらの経験などから、機械発見の論理あるいは科学発見の計算論理とでも呼ぶべきものを確立し、どのような種類の機械発見が可能であるかについて、正確に議論すべきであると考え始めた。1992年の暮のことである。

がって、Popperの科学発見の論理は、(彼の意図に反して)帰納推論の基礎を与えていることになる。そして、その帰納推論は、機械学習の理論的な基礎を与えている。それでは、機械発見の論理あるいは科学発見の計算論理とは何であろうか？

ここでいう機械発見とは、与えられたデータからある種の科学的発見をコンピュータにさせることである。したがって、機械学習は機械発見の基本的な技術となる。機械学習においては、我々はまず、コンピュータが理論、すなわち仮説を探すべき空間を設定する。その空間は当然大きいものであってほしいし、また一方で学習を効率的にするためにには、小さいものであってほしい。データが、通常の帰納推論におけるように、その仮説空間内のある未知の仮説に従って与えられる限り、そのコンピュータはいつかは仮説を同定できるので、何ら問題は生じない。しかし、機械発見においては、このことを仮定できない。

13. おわりに—数値データの扱い—

ところで、機械発見が対象の一つとする科学的発見では、与えられるデータは遺伝子やアミノ酸配列のような文字列の場合はまれであり、多くは数値データである。したがって、数値データを対象にした機械学習の基礎を確立する必要がある。そのためには、実数をどう表現するか、という問題にまず直面する。幸いなことに数値データには誤差が含まれているので、そのことを積極的に活用することによって、便利で有用な表現系が得られる。帰納的(計算可能)実数に基づいた解析学の計算理論や計算量理論もすでに十分展開されている。また、区間解析の手法も確立している。現在、こうした枠組みを使い、数値データからある意味で精度の保証された(微分)方程式を機械学習・発見するシステムの研究に取り組んでいるところである。

領域の成果出版物：発見科学とデータマイニング(2001)



森下 真一・宮野 悟 編

ISBN 978-4-320-12018-1

判型 A4変型

ページ数 318ページ

発行年月 2001年06月

価格 4,840円（税込）

序章 日本の発見科学プロジェクト（有川節夫）

第I部 推論による知識発見

- 第1章 「発見」の科学哲学—歴史的素描（野家啓一）
- 第2章 統計的記号処理言語PRISM（亀谷由隆・佐藤泰介）
- 第3章 予測モデルからのルール抽出—数式から言語へ（月本 洋・森田千絵）
- 第4章 帰納論理プログラミングと証明補完（山本章博・有村博紀・平田耕一）
- 第5章 KeyGraph—キーワード抽出ツールから発見ツールへの展開（砂山 渡・大澤幸生・谷内田正彦）
- 第6章 IntelligentPadの合成と再利用—帰納推論の立場から（原口 誠・平田 淳）

第II部 計算学習理論に基づく知識発見

- 第7章 能動学習と発見科学（安倍直樹・馬見塚 拓）
- 第8章 くり返しゲームとしての学習アルゴリズム（丸岡 章・瀧本英二）
- 第9章 コンピュータサイエンスのための単純かつ効率的なサンプリング技法（渡辺 治）
- 第10章 学習アルゴリズムの評価（上原邦昭）
- 第11章 幾何クラスタリングの情報計算幾何構造（今井 浩）
- 第12章 Support Vector Machineによる分類（高須淳宏）

領域の成果出版物：発見科学とデータマイニング(2001)



森下 真一・宮野 悟 編

ISBN 978-4-320-12018-1

判型 A4変型

ページ数 318ページ

発行年月 2001年06月

価格 4,840円（税込）

第III部 機械学習とデータマイニングに基づく知識発見

- 第13章 コンピュータ支援による科学的知識の発見 (Pat Langley著／宮野 悟・丸山 修訳)
- 第14章 学習か、マイニングか、モデリングか？—古生態学からの事例研究 (Heikki Mannila et al.著／森下真一訳)
- 第15章 分枝限定法を用いた並列グラフ探索による最適結合ルールの発見 (中谷明弘・森下真一)
- 第16章 知識発見と自己組織型の統計モデル (北川源四郎・樋口知之)
- 第17章 顧客の購買履歴からのデータマイニング (矢田勝俊・加藤直樹・羽室行信)
- 第18章 発見システムとヒューマンエキスパートのインテグレーション (丸山 修・宮野 悟)

第IV部 大規模数値データからの知識発見

- 第19章 太陽地球系物理学への知識発見の応用 (家森俊彦・上野玄太・能勢正仁・町田 忍・荒木 徹・亀井豊永・竹田雅彦)
- 第20章 ブラインドセパレーションとウェーブレットによる隠蔽画像の発見 (新島耕一)
- 第21章 計算機による科学的法則・モデルの発見方法の展開 (鷺尾 隆・元田 浩)
- 第22章 多変量データからの多項式型法則の発見 (中野良平・斎藤和巳)
- 第23章 音声データベースからの音声知識の発見 (鈴木基之・牧野正三)
- 第24章 仮想化された人体からのナビゲーションに基づく知識発見の支援ツール (斎藤豊文・鳥脇純一郎)

第V部 ネットワーク環境における知識発見

- 第25章 ミームメディアを用いた知財流通と科学技術データの可視化 (田中 譲)
- 第26章 ズーミング技術を用いた対話的情報検索インターフェース (豊田正史・柴山悦哉)
- 第27章 リンク情報からの知識網構成 (廣川佐千男・池田大輔・田口剛史)
- 第28章 インターネットでの企業間情報共有に向けたマルチエージェントシステム (毛利隆夫・高田裕志)

余談：二分決定グラフとデータマイニング

KAKEN

研究課題をさがす

研究者をさがす

KAKENの使い方

日本語

二分決定グラフを知識表現に用いたデータマイニングシステムの開発

研究課題

研究課題/領域番号

09558032

サマリー

研究種目

基礎研究(B)

配分区分

補助金

応募区分

属間研究

研究分野

知識情報学

研究機関

東京大学

研究代表者

宮野 悟 東京大学、医科学研究所、教授 (50128104)

← 宮野 悟 先生

研究分担者

植原 歩 九州大学、システム情報科学研究科、助教授 (00226151)

丸山 修 東京大学、医科学研究所、助手 (20262519)

阿久津 遼也 東京大学、医科学研究所、助教授 (90261859)

下薗 真一 九州工業大学、情報工学部、助教授 (70243988)

正代 隆義 九州大学、システム情報科学研究科、助教授 (50226304)

内田 智之 広島市立大学、情報科学部、助教授 (70264934)

研究期間(年度)

1997 - 1999

余談：二分決定グラフとデータマイニング

キーワード

二分決定グラフ / 学習 / 知識発見 / パターンマッチングアルゴリズム / データマイニング / 計算量 / 情報量 / ゲノム / データマインイング / ゲノム情報

研究概要

発見をすることは科学における最も基本的な活動であり,その発見は新しい彗星の発見であったり,ケプラーの法則のように観測データから導き出された物理法則であったりする.人間の知的活動として重要な位置にあるこうした知識発見は,最近注目されているデータマイニングシステムのように,人間に代わって計算機上のシステムが行うことが強く期待されている.

こうした状況に対応するために,本研究は,二分決定グラフ(BDD)を知識表現に用い,データからの高次の知識発見を可能とするシステムの開発を目的として研究を開始した.二分決定グラフは,論理回路設計の分野で重要な概念であり,論理関数の合成,最適化などでその有効性が發揮されているものである.また,こうしたシステムの開発を通して,発見科学の原理や理論の構築の実証的足がかりをつくることも本研究のもう一つの重要な目的である.

本研究では,まず決定グラフを知識表現に用いることにあるが,本研究を展開する中で,計算量に関する研究から,それを直接に取り扱うことには困難が伴うことが様々の観点から判明した.そのため,従来の決定木としてまず仮説を生成し,その後,決定ダイアグラムを編集するという方式をとった.また,仮説生成方式の研究としては,様々の学習方式を検討し,木構造,グラフ構造,ネットワーク構造等を学習するための理論的研究とシステム開発を行った.そしてこうした研究を総括した知識発見システムとしてHypothesisCreatorを開発し,本研究では主に仮説生成エンジンとして,数千万属性に対応できる並列決定木構成方式を実装した.最後に,仮説編集システムDecision Diagram Editorを開発することにより,Human Expertの介入を積極的に支援して,決定木から決定ダイアグラムを作り出すためのシステムを完成させた).



たきがわ いちがく
瀧川 一学

<https://itakigawa.github.io>

と同時に機械学習のユーザでもあります！

- うどん県高松市生まれ
- 1995～2004 北海道大 (工学研究科)
2004 博士(工学) "劣決定信号源分離の解の理論分析"
- 2005～2011 京都大 (化学研究所/薬学研究科)
バイオインフォマティクスセンター 助教
- 2012～2018 北海道大 (情報科学研究科)
大規模知識処理研究室 准教授
- 2015～2018 JSTさきがけ (材料インフォマティクス)
- 2019～ 北海道大学 化学反応創成研究拠点(ICReDD)

2019～ 理化学研究所 革新知能統合研究センター(AIP)

普段は京大iPS細胞研との連携ラボ@京阪奈ATRに勤務
(iPS細胞連携医学的リスク回避チーム)

大規模知識処理研究室(旧：知識メディア研究室)

北海道大学大学院工学研究科 電子情報工学専攻
計算機情報通信工学講座 計算機アーキテクチャ工学分野

<http://ca.meme.hokudai.ac.jp/profile-j.html> Go JUL FEB APR
◀ 04 ▶
1997 1999 2000 About this capture
22 captures
26 Jun 1997 – 20 Apr 2005
[English | Japanese]

スタッフ	博士課程	修士課程	学部
→ 田中 譲 教授	伊藤公人 内海久夫 岡田義広 助手 赤石 美奈 助手 石川栄一 技官 黒岩智子 秘書	アリフ・ユスフ 大東誠 堀井克敏 村本典子 中川光紀 黎亞和 平野亮太 田中一生 村上悌治	玉置伸匡 岩井靖志 関泰規 藤間淳 相笠健一 岡田亮 新保健郎 杉測剛史
→ 山本章博 助教授		東高志 伊藤正彦 田中一志 田中慈郎 堀田均	
Profile			

瀧川は北大学生当時違う専攻の研究室↓だったため、ほぼ縁なし

北海道大学大学院工学研究科 システム情報工学専攻
計算機情報通信工学講座 計算機アーキテクチャ工学分野

[OB・OG、その他の関係者の方々](#)

[Index]

北海道大学大学院情報科学研究科 コンピュータサイエンス専攻
知識ソフトウェア科学講座 知識メディア研究室

http://www.ist.hokudai.ac.jp/intro/depart.php?depart_ID=821 Go MAY JUN AUG
◀ 26 ▶
2003 2004 2005 About this capture

Graduate School of Information Science and Technology Hokkaido University ホーム | サイトマップ

専攻・講座の紹介

コンピュータサイエンス専攻
知識ソフトウェア科学講座

キーワード
基盤的ソフトウェア、計算機アーキテクチャ、知識メディア、知識ベース、知識発見、巨大ネットワークからの情報獲得、ウェブセキュリティ、アルゴリズムの設計と解析、アルゴリズムの計算量、超分散自律計算モデル

スタッフ

氏名	役職	部屋	電話番号
田中 譲	教授	情615	7252
原口誠	教授	情602	7106
有村 博紀	教授	情706(3)	7678
Thomas Zeugmann	教授	情819	7684
吉田 哲也	助教授	情616	7253
吉岡 真治	助教授	情601	7107
喜田 拓也	助教授	情706(2)	7679
渡 真一	助教授	情821	7682
伊藤 公人	助手	情614	7262
大久保 好章	助手	情604	7161

大規模知識処理研究室(旧：知識メディア研究室)

北海道大学大学院情報科学研究科 コンピュータサイエンス専攻
知識ソフトウェア科学講座 **知識メディア研究室**



スタッフ

知識メディア研究室

教授 [湊真一](#) Shin-ichi MINATO minato (ist) 内6469
准教授 [滝川一学](#) Ichigaku TAKIGAWA takigawa (ist) 内6470

ERATO

特任准教授 [安田 宜仁](#) Norihiro YASUDA yasuda (ist) 内7656
博士研究員 [竹内 聖悟](#) Shogo TAKEUCHI takeuchi (erato) 内7656
技術補佐員 [渡辺 幸恵](#) Yukie WATANABE yukie-wat (erato) 内7656

共同プロジェクト拠点

特任教授 田中 譲	Yuzuru TANAKA	tanaka (meme)	内7250
特任助教 猪村 元	Hajime IMURA	hajime (meme)	内7262
特任助教 ミカエル ニキヤンデル クワハラ	Mikael NICANDER KUWAHARA mkuwahara (meme)	内7261	
特任助教 ヨーナス シューベルグ	Jonas SJOBERGH	js (meme)	内7259
学術研究員 遠藤 隆浩	Takahiro ENDO	endo (meme)	内7262
学術研究員 高橋 知宏	Tomohiro TAKAHASHI	takahashi (meme)	内7262
非常勤研究員 ヤニス ヨルガリス	Ioannis GEORGALIS	jgeorgal (meme)	内7250
事務補助員 犬上 薫	Kaoru INUGAMI	k.inugami (meme)	内7250

北海道大学大学院情報科学研究科 情報理工学専攻
知識ソフトウェア科学講座 **大規模知識処理研究室**



大規模知識処理研究室

教授 [湊真一](#) Shin-ichi MINATO minato (ist) 内6469
准教授 [滝川一学](#) Ichigaku TAKIGAWA takigawa (ist) 内6470
事務補佐員 [相馬 佐智子](#) Sachiko SOMA soma (art) 内6471

基盤(S) 離散構造処理系プロジェクト

特任助教 [石島 正和](#) Masakazu ISHIHATA ishihata.masakazu (ist) 内7259
博士研究員 [Hei CHAN](#) Hei CHAN 内7259
技術補佐員 [渡辺 幸恵](#) Yukie WATANABE yukie-wat (erato) 内7259

共同プロジェクト拠点

特任助教 猪村 元	Hajime IMURA	hajime (meme)	内7262
特任助教 ミカエル ニキヤンデル クワハラ	Mikael NICANDER KUWAHARA mkuwahara (meme)	内7261	
特任助教 ヨーナス シューベルグ	Jonas SJOBERGH	js (meme)	内7259
学術研究員 遠藤 隆浩	Takahiro ENDO	endo (meme)	内7262
事務補助員 犬上 薫	Kaoru INUGAMI	k.inugami (meme)	内7250

大規模知識処理研究室(旧：知識メディア研究室)

大規模知識処理研究室

Large-Scale Knowledge Processing Lab



[トップページ](#) [ニュース](#) [研究室について](#) [リンク](#) [言語:](#) ●

スタッフ

- 教授: 堀山 貴史
- 准教授: 脊戸 和寿
- ポスドク研究員: 三重野 琢也
- 技術補助員: 渡辺 & 相馬

学生

- M1: 5名
- B4: 3名

過去のスタッフ

- 教授: 漆 真一
- 准教授: 濵川 一学

知識ソフトウェア科学分野

研究室	氏名	かな	役職
大規模知識処理	堀山 貴史	ほりやまたかし	教授
知識ベース	吉岡 真治	よしおか まさはる	教授
情報知識ネットワーク	有村 博紀	ありむら ひろき	教授
アルゴリズム	中村 篤祥	なかむら あつよし	教授
大規模知識処理	脊戸 和寿	せと かずひさ	准教授
知識ベース	大久保 好章	おおくぼ よしあき	助教
アルゴリズム	林 達也	はやし たつや	特任助教

人工知能学会 人工知能基本問題研究会 (SIG-FPAI)

表1 人工知能基本問題研究会 歴代幹事

年 度	主 査	主幹事	幹 事
※人工知能基礎論研究会 (SIG-FAI) として発足			
1987-1989	有川節夫	後藤滋樹	佐藤泰介, 原口 誠
1990-1991	佐藤泰介	元田 浩	久野 巧, 横森 貴
1992-1993	西田豊明	櫻井彰人	西野哲朗, 三浦欽也
1994-1995	石田 亨	大沢英一	赤埴淳一, 麻生英樹
1996-1997	國藤 進	西野哲朗	山田誠二, 鶯尾 隆
1998-1999	元田 浩	榎原康文	櫻井彰人, 鶯尾 隆
2000-2001	櫻井彰人	松原 仁	大澤幸生, 榎原康文, 月本 洋
2002-2003	松原 仁	大澤幸生	有村博紀, 小野哲雄
※人工知能基本問題研究会 (SIG-FPAI) に名称変更			
2004	佐藤 健	鈴木 讓	市瀬龍太郎, 庄司裕子, 平田耕一, 吉岡真治
2005	佐藤 健	鈴木 让	赤石美奈, 市瀬龍太郎, 平田耕一, 吉岡真治
2006	有村博紀	佐藤 健	赤石美奈, 喜田拓也, 坂本比呂志, 村上知子, 山本章博
2007	有村博紀	山本章博	相原健郎, 伊藤公人, 喜田拓也, 坂本比呂志, 村上知子

2008	山本章博	平田耕一	相原健郎, 伊藤公人, 植野真臣, 久保山哲二
2009	山本章博	平田耕一	伊藤公人, 植野真臣, 久保山哲二
2010	平田耕一	久保山哲二	赤石美奈, 磯崎隆司, 伊藤公人, 鍋島英知
2011	平田耕一	久保山哲二	赤石美奈, 磯崎隆司, 中村篤祥, 鍋島英知
2012	久保山哲二	坂本比呂志	河原吉伸, 川前憲章, B. Chakraborty, 中村篤祥
2013	久保山哲二	坂本比呂志	大久保好章, 鍛治伸裕, 河原吉伸, 川前憲章, B. Chakraborty
2014	坂本比呂志	河原吉伸	大久保好章, 鍛治伸裕, 越村三幸, 田部井靖生
2015	坂本比呂志	河原吉伸	越村三幸, 瀧川一学, 田部井靖生
2016	河原吉伸	瀧川一学	石畠正和, 西郷浩人, 宋剛秀, 田部井靖生
2017	河原吉伸	瀧川一学	石畠正和, 西郷浩人, 宋剛秀, 戸田貴久
2018-2019	瀧川一学	大久保好章	井 智弘, 杉山磨人, 戸田貴久, 西野正彬

2020-2021 大久保好章 石畠正和
大滝 啓介, 小林 靖明,
後藤 啓介, 萩田 玲緒奈

人工知能学会 人工知能基本問題研究会 (SIG-FPAI)

学会誌「人工知能」 Vol. 34, No 5, 2019年9月 (オープンアクセス)

https://doi.org/10.11517/jjsai.34.5_603

人工知能基本問題研究会 (FPAI)

Special Interest Group on Fundamental Problems in Artificial Intelligence

瀧川 一学

Ichigaku Takigawa

理化学研究所革新知能統合研究センター

RIKEN Center for Advanced Intelligence Project.

ichigaku.takigawa@riken.jp, <https://itakigawa.github.io/>

"論理, 学習, 知識の表現と獲得, 並列計算モデル, 知的プログラミング, 自然言語理解, パターン理解などに関する人工知能としての基礎的研究 (有川, 1990)"

ところで、今月の人工知能学会誌の瀧川さんのSIG-FPAI記事、**ただの研究会の紹介記事のハズ**のなのに、えらい見識になっていて、他の研究会の記事との落差に**困惑しました**。みなさまも是非。

kashi_pong先生



たきがわ いちがく
瀧川 一学

<https://itakigawa.github.io>

と同時に機械学習のユーザでもあります！

- うどん県高松市生まれ
- 1995～2004 北海道大 (工学研究科)
2004 博士(工学) "劣決定信号源分離の解の理論分析"
- 2005～2011 京都大 (化学研究所/薬学研究科)
バイオインフォマティクスセンター 助教
- 2012～2018 北海道大 (情報科学研究科)
大規模知識処理研究室 准教授

2015～2018 JSTさきがけ (材料インフォマティクス)

- 2019～ 北海道大学 化学反応創成研究拠点(ICReDD)
2019～ 理化学研究所 革新知能統合研究センター(AIP)

普段は京大iPS細胞研との連携ラボ@京阪奈ATRに勤務
(iPS細胞連携医学的リスク回避チーム)

常行さきがけ(材料インフォ領域)と物材機構MI²I



[マテリアルズインフォ] 理論・実験・計算科学とデータ科学が連携・融合した先進的マテリアルズインフォマティクスのための基盤技術の構築

◀ トップに戻る

□ 領域活動についてはこちらをご覧ください。

| 戦略目標

「多様な天然炭素資源を活用する革新的触媒の創製」
「情報デバイスの超低消費電力化や多機能化の実現に向けた、素材技術・デバイス技術・ナノシステム最適化技術等の融合による革新的基盤技術の創製」
「分野を超えたビッグデータ利活用により新たな知識や洞察を得るために革新的な情報技術及びそれらを支える数理的手法の創出・高度化・体系化
「環境・エネルギー材料や電子材料、健康・医療用材料に革新をもたらす分子の自在設計「分子技術」の構築」

研究総括

常行 真司 (東京大学 大学院 理学系研究科 教授)

領域アドバイザ

青柳 岳司

産業技術総合研究所 機能材料コンピュテーション研究センター 総括研究主幹

→ 伊藤 聰

(公財) 計算科学振興財団 チーフコーディネータ

射場 英紀

トヨタ自動車(株) 先端材料技術部 チーフプロフェッショナルエンジニア

小谷 元子

東北大学 材料科学高等研究所 教授

佐藤 寛子

情報・システム研究機構 准教授／チューリッヒ大学 研究員

田中 功

京都大学 大学院 工学研究科 材料工学専攻 教授

→ 知京 豊裕

物質・材料研究機構統合型材料開発・情報基盤部門

情報統合型物質・材料研究拠点 副拠点長

津田 宏治

東京大学 大学院 新領域創成科学研究科 メディカル情報生命専攻 教授

→ 寺倉 清之

産業技術総合研究所 名誉リサーチャー

中井 浩巳

早稲田大学 理工学術院 教授

樋口 知之

中央大学 理工学部 教授

当時の統計数理研究所・所長

常行さきがけ(材料インフォ領域)と物材機構MI²I

北海道大学 名誉教授 田中 譲 (MIリサーチアドバイザー) ←

東京大学 教授 岡田 真人 (副プロジェクトリーダー)

← 岡田新学術(スペースモデリング)に公募班で参加

https://www.nims.go.jp/MII-I/about/index_g.html

データ科学グループ



グループリーダー：津田 宏治 (東京大学)

本グループでは、機械学習・人工知能のアルゴリズムを物質・材料データに応用し、新たな物質・材料の発見につなげることを目標とし、理論・実践の両面から研究を行っています。

[\(詳細ページへ\)](#)

物質・材料記述基盤グループ



グループリーダー：吉田 亮 (情報・システム研究機構 統計数理研究所)

本グループでは、マテリアルズインフォマティクスの分野で世界最大級となる包括的記述子ライブラリの開発を進めています。

[\(詳細ページへ\)](#)

トポロジカル解析グループ



グループリーダー：赤木 和人 (東北大学)

本グループでは、位相幾何学（トポロジー）と呼ばれる数学に立脚した新しい解析手法を用いることで乱雑さや複雑さに埋もれた構造的特徴を定量的に抽出して符号化し、Mi²Iプロジェクトが目指すデータ科学的な材料開発の対象を大きく広げることに挑戦しています。

[\(詳細ページへ\)](#)

データプラットフォームグループ



グループリーダー：徐 一斌 (物質・材料研究機構)

本グループは、NIMS物質・材料データベースMatNaviおよびMi²Iデータプラットフォームコンテンツの開発、さらに、システム構築と運用管理を担当しています。

[\(詳細ページへ\)](#)



たきがわ いちがく
瀧川 一学

<https://itakigawa.github.io>

と同時に機械学習のユーザでもあります！

- うどん県高松市生まれ
 - 1995～2004 北海道大 (工学研究科)
2004 博士(工学) "劣決定信号源分離の解の理論分析"
 - 2005～2011 京都大 (化学研究所/薬学研究科)
バイオインフォマティクスセンター 助教
 - 2012～2018 北海道大 (情報科学研究科)
大規模知識処理研究室 准教授
2015～2018 JSTさきがけ (材料インフォマティクス)
 - 2019～ 北海道大学 化学反応創成研究拠点(ICReDD)
2019～ 理化学研究所 革新知能統合研究センター(AIP)
- 普段は京大iPS細胞研との連携ラボ@京阪奈ATRに勤務
(iPS細胞連携医学的リスク回避チーム)

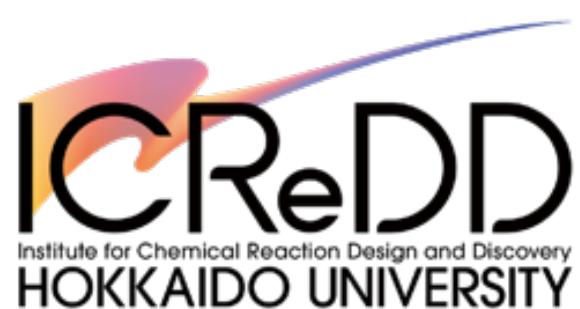
現職 2019年4月1日～ 2拠点のクロスアポイント

- 理化学研究所 革新知能統合研究センター (AIP) 研究員



- 離散構造・組合せ構造を伴う機械学習
- 幹細胞生物学のための機械学習 (細胞画像 + 深層学習)
- 新しいアルゴリズム・最適化の定式化と求解法
- 透過型電子顕微鏡+機械学習による動的観察

- 北海道大学 化学反応創成研究拠点 (WPI-IICReDD) 特任准教授



- 機械学習から機械発見へ
- 化学反応のデザインと発見のための機械学習
- 分子のグラフ表現の学習と生成
- 量子化学計算 + 機械学習の融合
- 機械発見：探索、実験計画、知識発見

理化学研究所革新知能統合研究中心



勤務地：京阪奈地区(京都府相楽郡精華町)

けいはんな地区

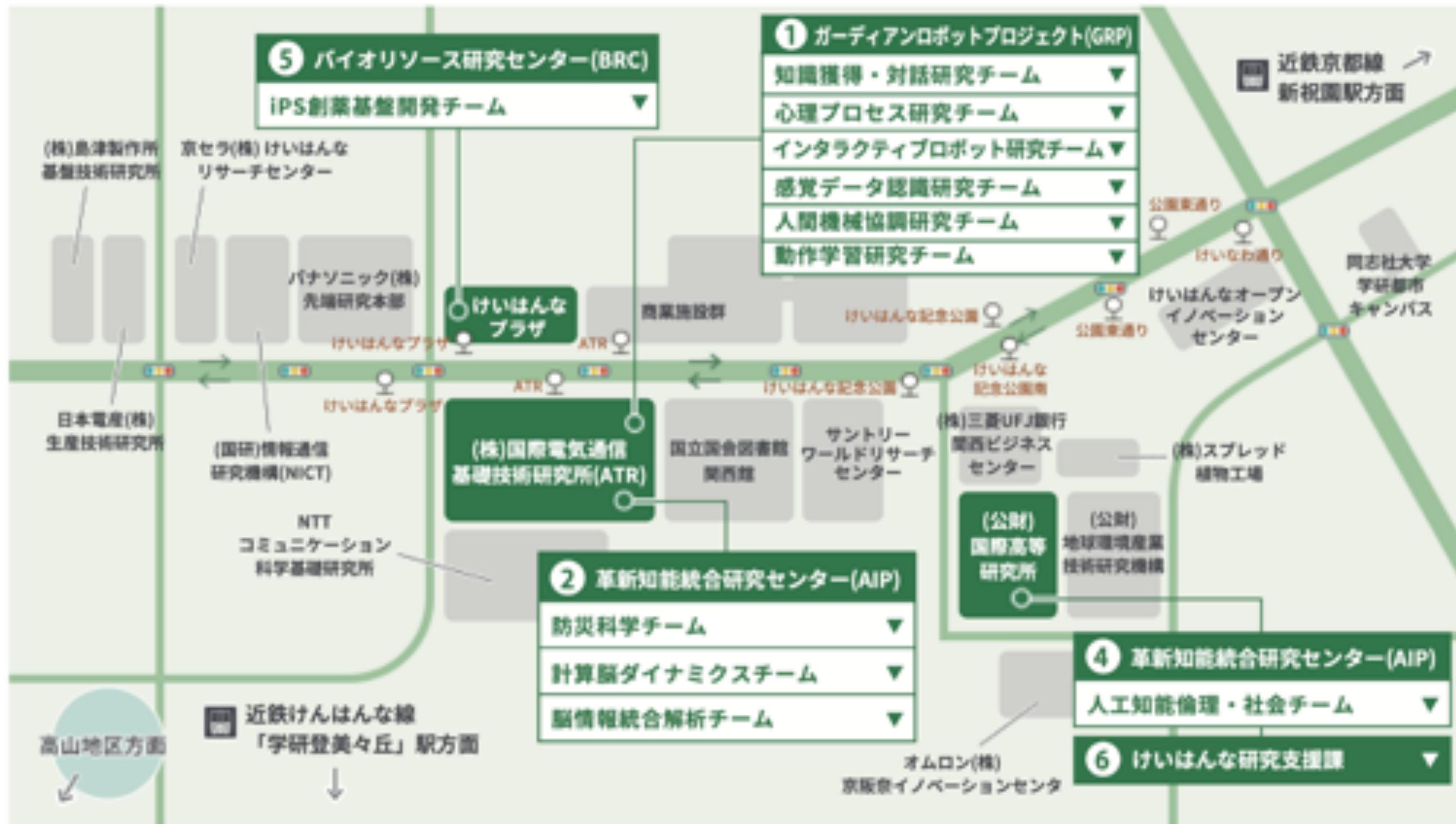
<https://www.kobe.riken.jp/about/map/keihanna/>



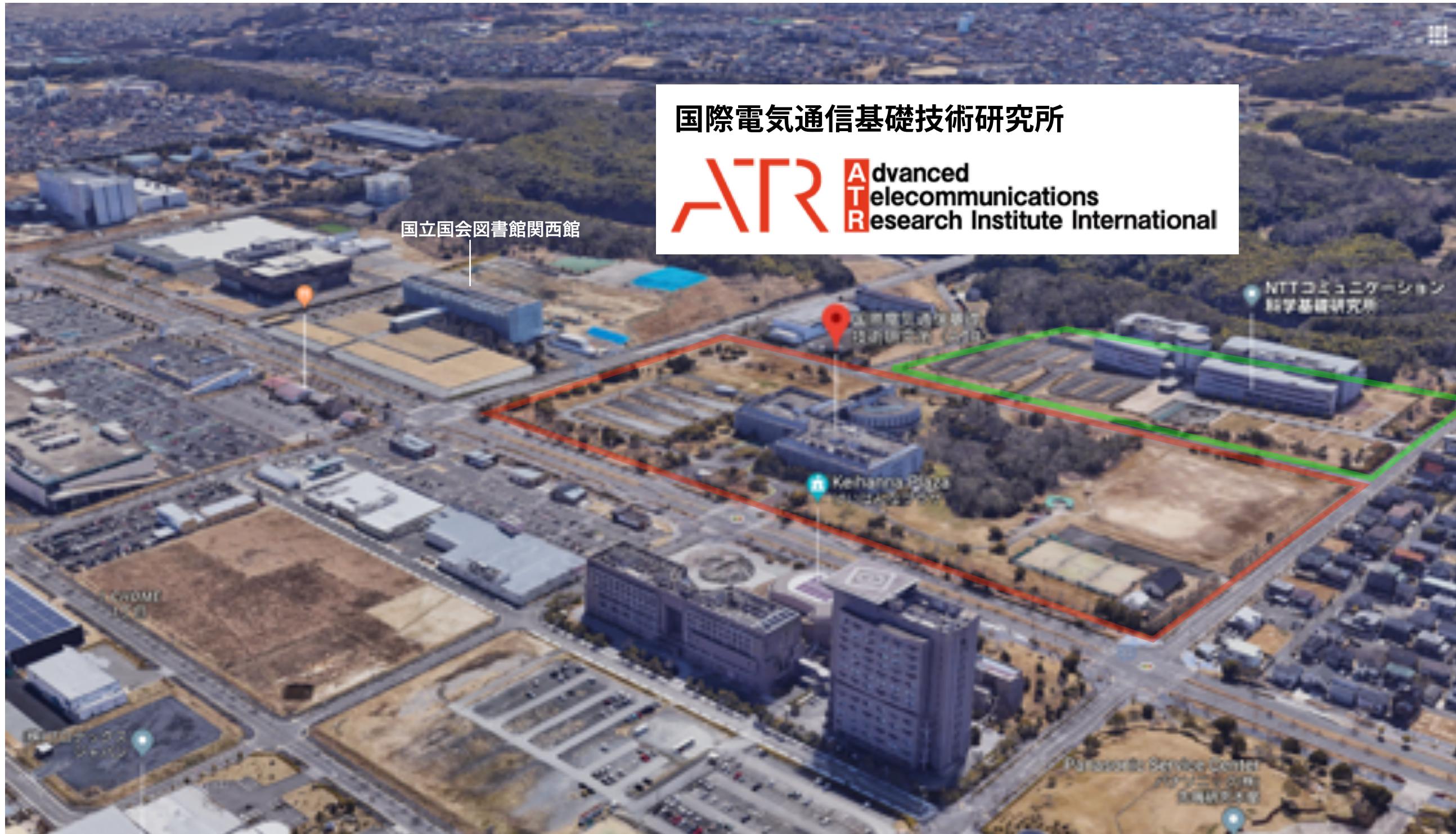
第三セクターの研究所



勤務地：京阪奈地区(京都府相楽郡精華町)



勤務地：京阪奈地区(京都府相楽郡精華町)



国際電気通信基礎技術研究所 (ATR)



- **理化学研究所**
革新知能統合研究センター (AIP)
ガーディアンロボットプロジェクト (GRP)
- **ATR脳情報通信総合研究所**
脳情報研究所
認知機構研究所
脳情報解析研究所

- **ATR深層インタラクション総合研究所**
インタラクション技術バンク
インタラクション科学研究所
石黒浩特別研究所, 萩田紀博特別研究所
- **ATR適応コミュニケーション研究所**
- **ATR波動工学研究所**
- **ATR佐藤匠徳特別研究所**

理研AIP@ATR



- 防災科学チーム (上田 修功)
- 脳情報統合解析チーム (川鍋 一晃)
- 計算脳ダイナミクスチーム (山下 宙人)
- iPS細胞連携医学的リスク回避チーム (上田 修功)

理研AIPと京大iPS細胞研の連携ラボ



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



京都大学 iPS細胞研究所
CiRA (サイラ)



チームリーダー 上田 修功	研究員 瀧川 一学
技師 永橋 文子	テクニカルスタッフ 松林 由季
テクニカルスタッフ 江浪 貴子	テクニカルスタッフ 井上 育代
客員主管研究員 井上 治久	客員研究員 中川 誠人
客員研究員 山本 拓也	

文科省世界トップレベル拠点形成プログラム(WPI)

コンセプト

WPIプログラムは、「世界最高レベルの研究水準」、「融合領域の創出」、「国際的な研究環境の実現」、「研究組織の改革」という4つのミッションの下、ブレインソーキュレーションの中にしっかりと位置づけられる、「目に見える研究拠点」の形成を目指しています。WPIセンターは、日本の研究機関のモデルになるとともに、科学技術に革新をもたらすことを期待されています。



特徴

研究者のクリティカル・マス

国際水準の研究環境

長期的な財政支援

厳格な評価システム

「目に見える研究拠点」の実現へ

https://www.mext.go.jp/a_menu/kagaku/toplevel/

予算規模：最大7億円/年×10年

文科省世界トップレベル拠点形成プログラム(WPI)

世界トップレベル研究拠点

宇宙/地球・生命/知性
の起源

KAVLI
iPMU
平成19年度採択

ELSI
EARTH, LIFE, SCIENCE INSTITUTE
平成24年度採択

IRCN
International Research Center for Neurointelligence
平成29年度採択

生物

WPI Osaka University
iFReC
平成19年度採択

IHS
INTERNATIONAL INSTITUTE FOR HIGH-SPEED
IMAGING SCIENCE
平成24年度採択

ASHBi
WPI Kyoto University
平成30年度採択

NanoLSI
MURAKAMI FUTURE SCIENCES
UNIVERSITY
平成29年度採択

材料/エネルギー

JAiMR
平成19年度採択

MANA
平成19年度採択

I²CNER
平成22年度採択

ICReDD
HOKKAIDO UNIVERSITY
平成30年度採択

データ・情報科学



ニューロインテリジェンス国際研究機構



データ・情報科学に重なる
WPI拠点は二拠点のみ

北大拠点は「化学反応のデザインと発見」

北大WPI拠点：化学反応創成研究拠点 (ICReDD)

化学反応創成研究拠点 (ICReDD) とは

化学反応創成研究拠点 (ICReDD／アイクレッド) では、計算科学、情報科学、実験科学の 3 分野を融合させることにより、新しい化学反応をより深く理解し効率的に開発することを目指しています。



北大WPI拠点：化学反応創成研究拠点 (ICReDD)

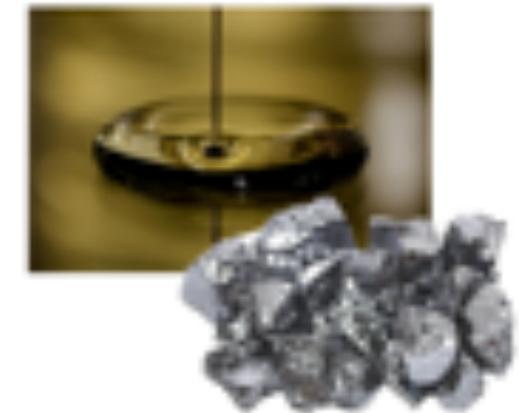


最小限の前おき：化学=物質AをBに変えたい！

原子がひとまとめになった「分子」とその化学反応は広範囲な分野の主人公

化学、医学・生理学、物理学、生物学、創薬、材料科学、環境科学、農学、食品、化粧品、…

身の回りのもの



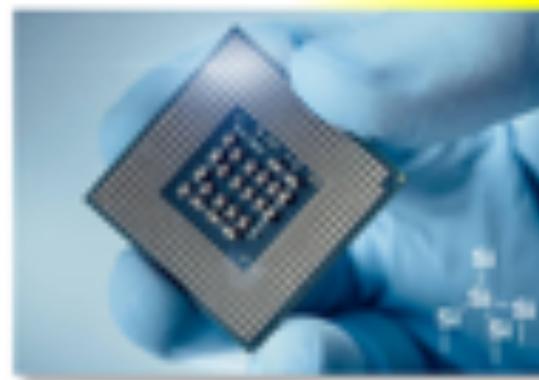
私たち自身(生命現象)

Chemical
Reactions



"化"学

エネルギー



分子は「組合せ的」側面を持つ

c&en

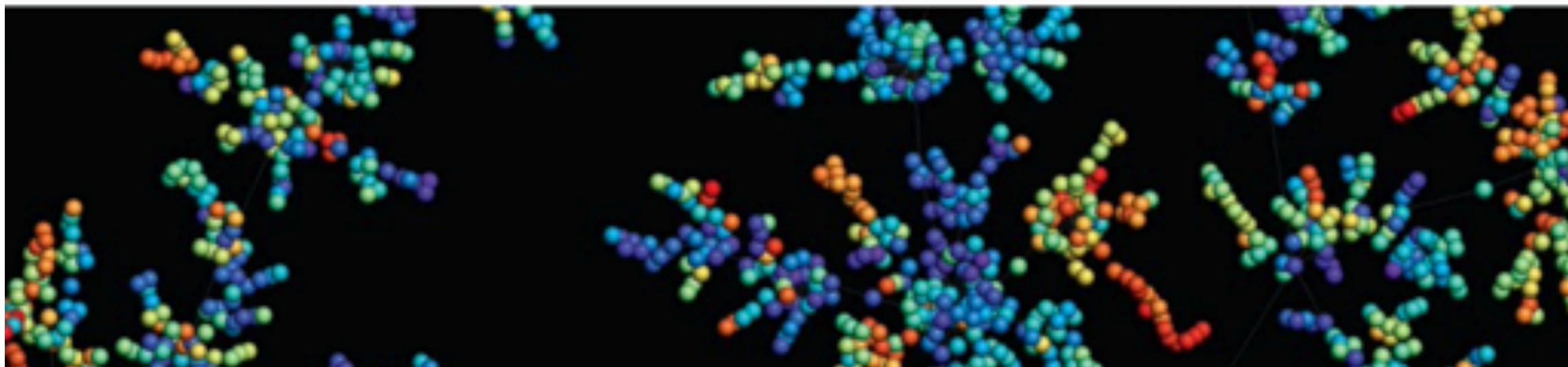
COMPUTATIONAL CHEMISTRY

Exploring chemical space: Can AI take us where no human has gone before?

Artificial intelligence is helping us find novel, useful molecules. For the field to really take off, though, these tools will need to be accessible to the wider chemistry community

by Sam Lemonick

April 6, 2020 | A version of this story appeared in **Volume 98, Issue 13**



BY THE NUMBERS

10^{180}

An upper estimate of the number of possible molecules

10^{80}

Estimated number of atoms in the universe

10^{60}

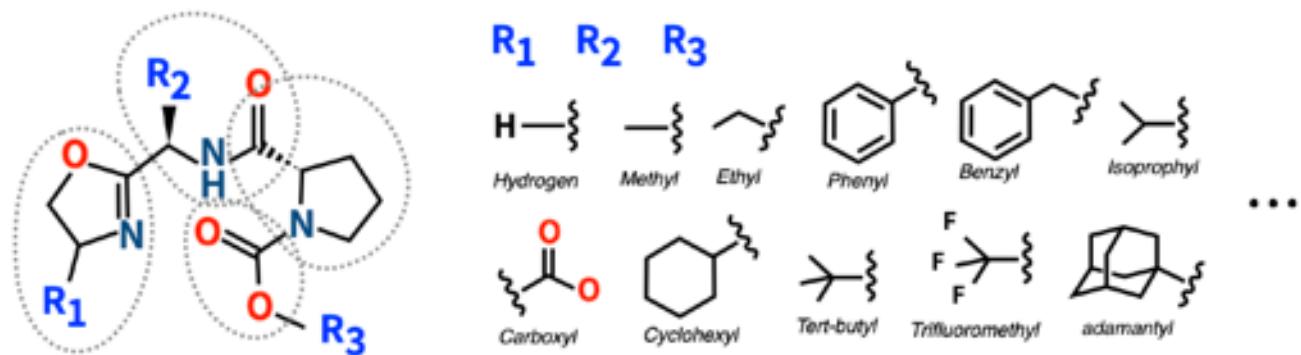
An estimate of the number of possible small organic molecules

10^8

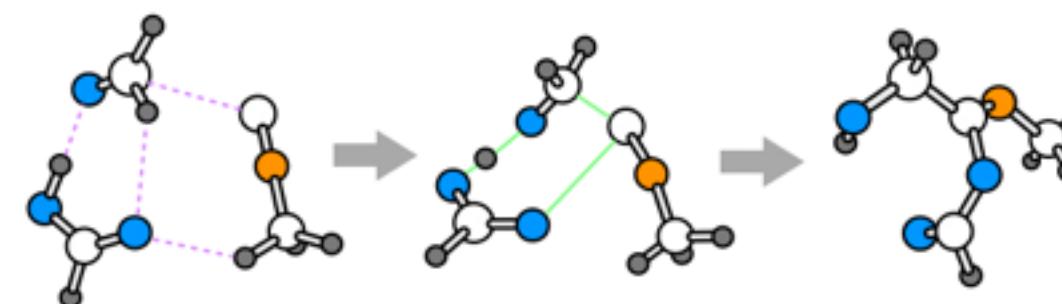
The number of organic and inorganic substances in the CAS database

機械学習で化学をやってみた！

分子は原子間の結合の組合せ（組合せ規則は量子力学）→ 技術上の関心は「組合せの機械学習」



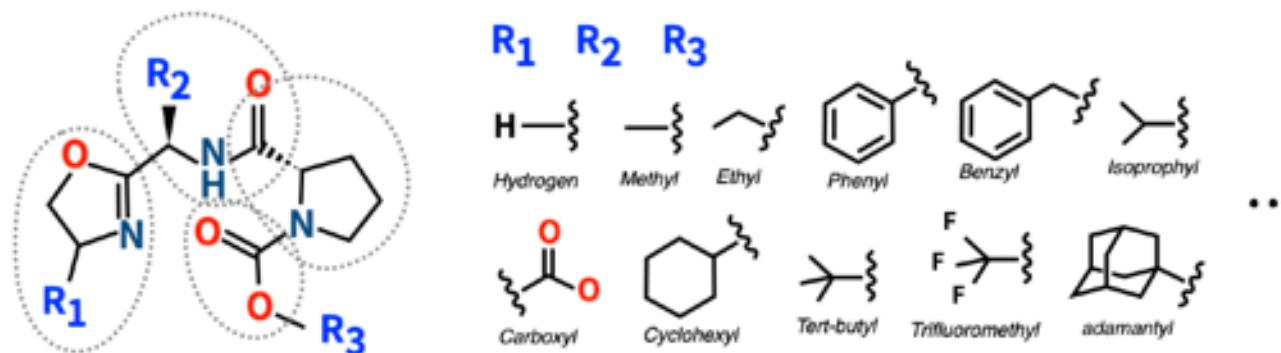
要素の組合せに構成性と階層性がある（言語に似ている？）



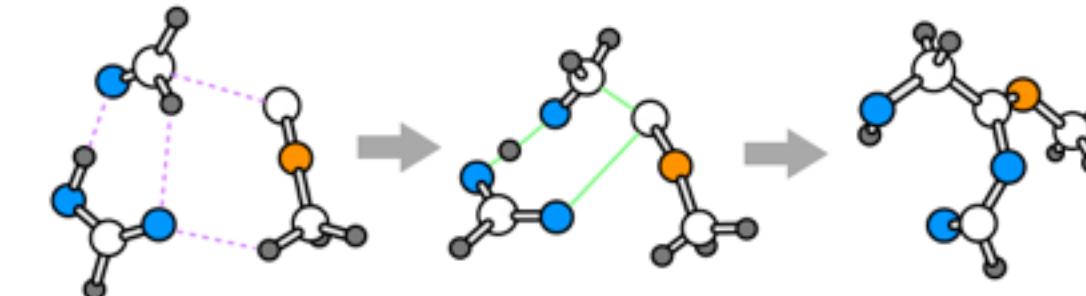
化学反応は結合の組み替え（文法に似ている？）

機械学習で化学をやってみた！

分子は原子間の結合の組合せ（組合せ規則は量子力学）→ 技術上の関心は「組合せの機械学習」



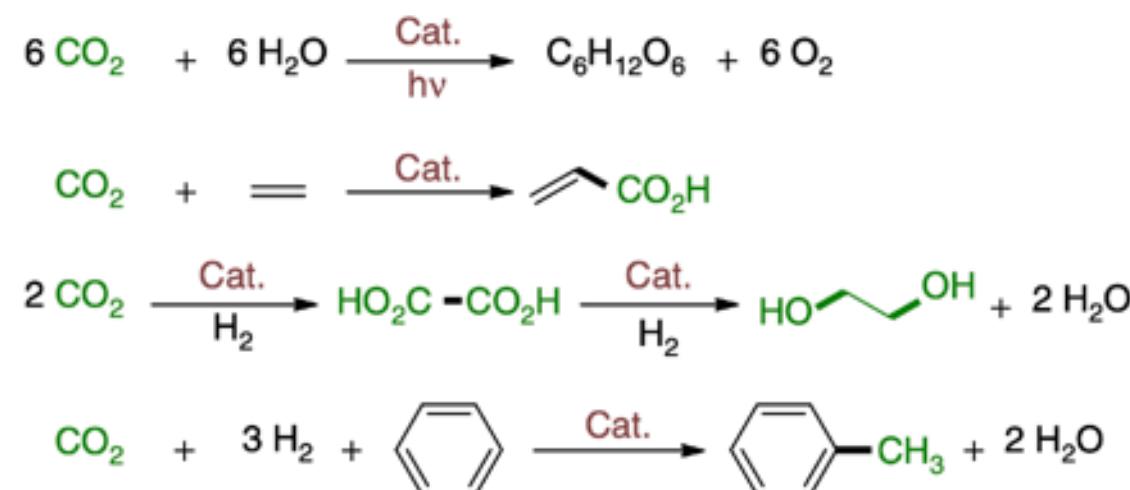
要素の組合せに構成性と階層性がある（言語に似ている？）



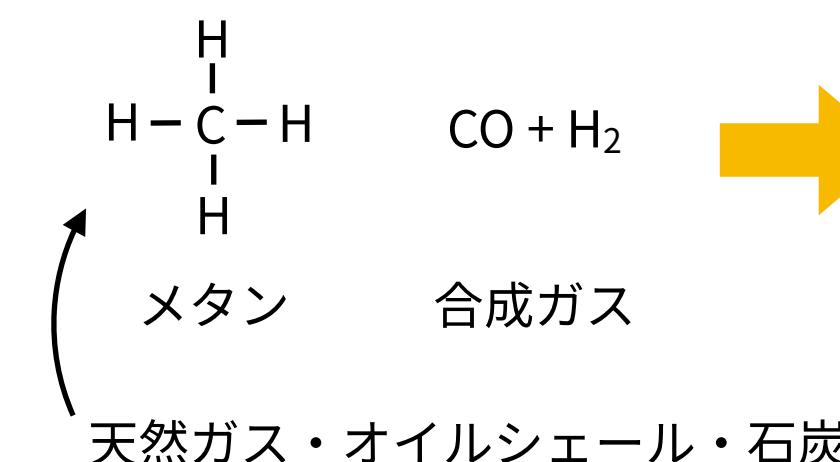
化学反応は結合の組み替え（文法に似ている？）

本当のゴールは「化学反応のデザインと発見」で機械学習はそのための道具の一つ

CO₂の資源利用：人工光合成



C1化学 原料はCが一つ



C2炭化水素類の合成

エチレン: C ₂ H ₄	エタン: C ₂ H ₆

天然ガス・オイルシェール・石炭・バイオマスなど石油以外から得られる

機械学習の力の光明面と暗黒面

Q. 機械学習・機械発見の技術って**自然現象の理解・発見**の役に立つのかな？

ライトサイド（光明面）

機械学習は「データを予測に変える」強力なテクノロジー！

- ✓ 分子の表現学習とGraph Neural Networks
- ✓ 帰納バイアスの設計とグレイボックス最適化

ダークサイド（暗黒面）

自然科学の実現象データで使うのはいろいろ激ムズ！！！

- ✓ 羅生門効果とUnderspecification
- ✓ 「予測ができること」は「理解」や「発見」ができるることを意味しない！

フォースと共にあらんことを *May the ML force be with you...*

京大 2021年度知能情報学専攻コロキウム

<https://kuist-colloquium.github.io>

動画(59:32) <https://vimeo.com/656871480>

機械学習・機械発見から見るデータ中心型化学の野望と憂鬱

瀧川 一学

理化学研究所 革新知能統合研究センター 研究員

日時：12月9日（木） 12時～13時

場所：オンライン

この発表では化学を例に機械学習屋から見た自然科学研究における「機械学習・機械発見技術の力」の光明面と暗黒面を概観する。化学はある分子A[例えば二酸化炭素]を別の分子B[例えばアルコール]に変える方法を研究する。私たちの身の回りの材料・医薬品・エネルギーなどに加え、あらゆる生命現象も化学反応の産物であり、その波及範囲は学術上も産業上も広大である。しかし現在の化学は依然として経験科学的な側面が強く、ご多分にもれず機械学習の利活用が現在盛んに研究されるようになった。化学は主役である分子や化学反応が「組合せ的対象」である点とその組合せを制約する「第一原理(量子力学)が分かっている」点により、情報科学者にとっても面白い問題の宝庫である。私たちを熱中させるに足る技術上の研究課題には事欠かない一方で、新しい反応～実際にAをBに変える方法～の実現は依然として超絶的に困難なままで「自在な化学反応の設計・発見」への道のりは遠い。機械学習×化学というトピックの何が面白いのか、なぜそんなに難しいのか、を様々な自然科学研究に携わってきた機械学習屋の視点から紹介する。



機械学習は「新しい(雑な)コンピュータプログラムの作り方」

プログラム

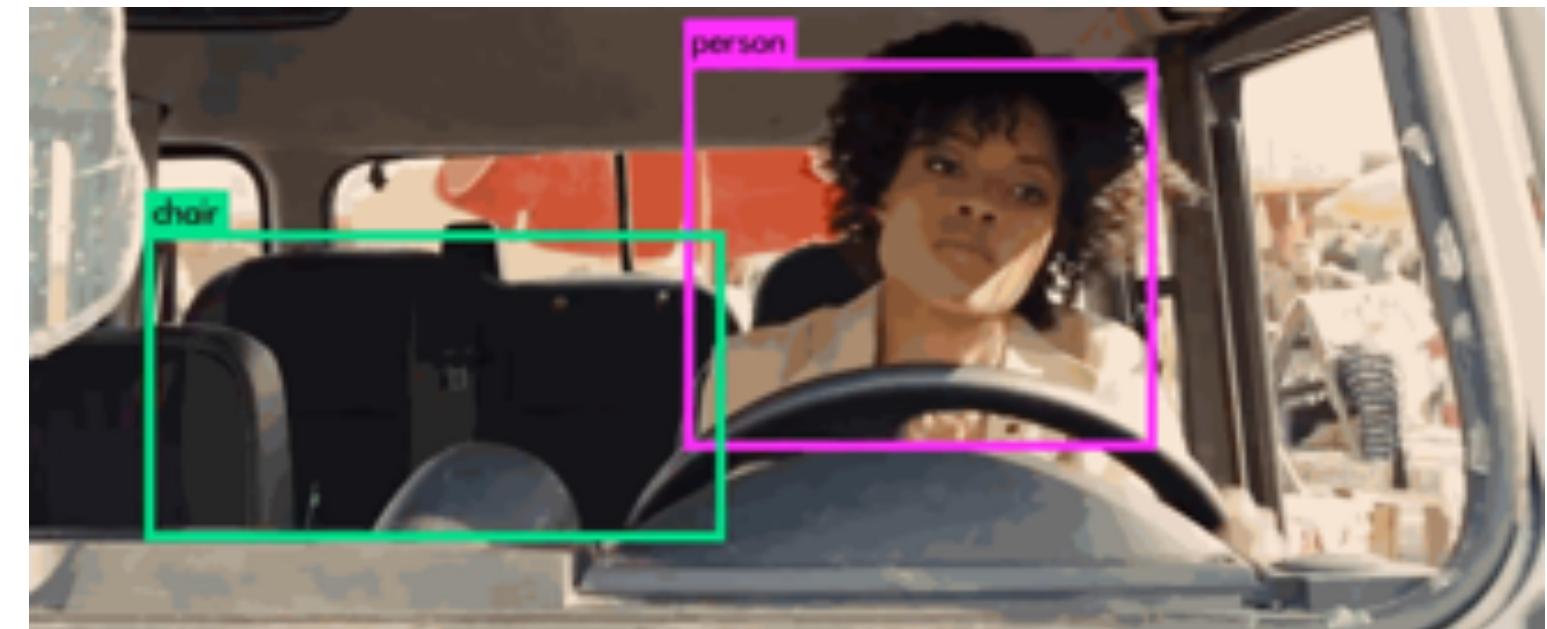
の入力と出力の関係がよく分からない場合でも、

たくさんの入出力の見本データによって間接的に見本を再現するプログラムを作り出す技術



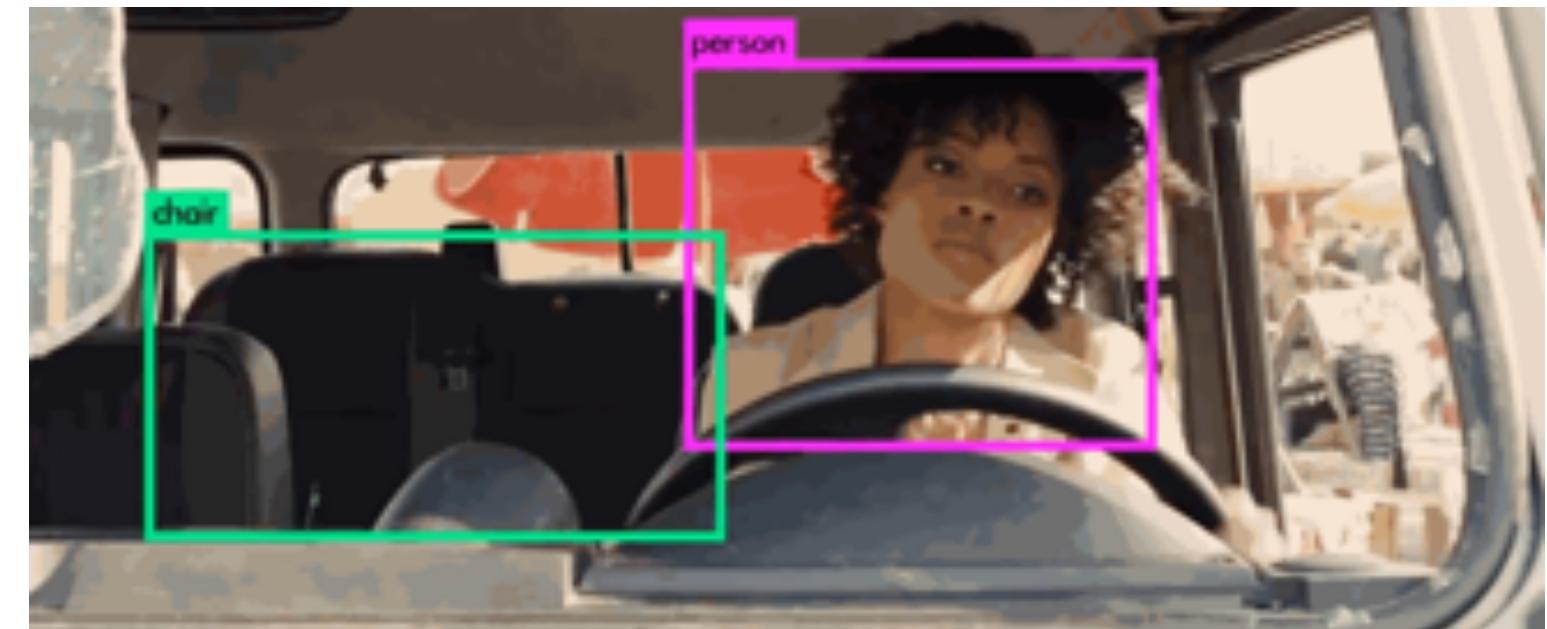
機械学習は「新しい(雑な)コンピュータプログラムの作り方」

この単純なしくみは上手に使うと「めちゃくちゃ強力」でいろいろな楽しいこともできる！



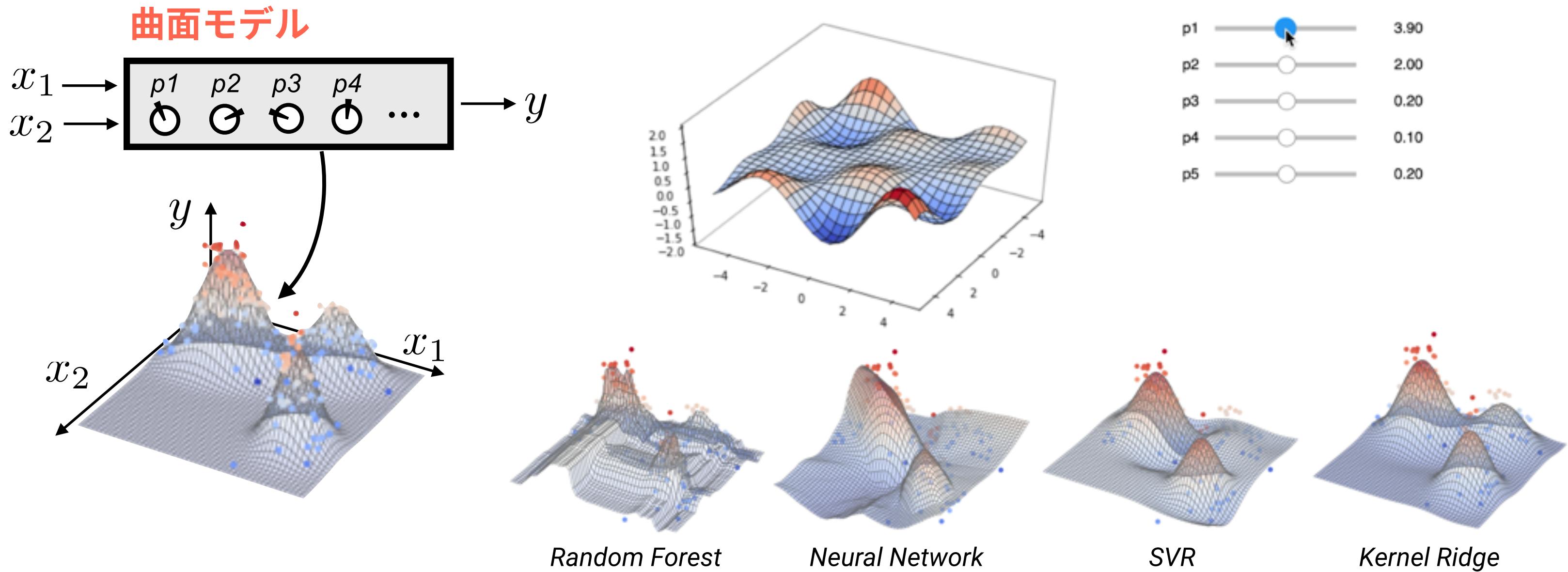
機械学習は「新しい(雑な)コンピュータプログラムの作り方」

この単純なしくみは上手に使うと「めちゃくちゃ強力」でいろいろな楽しいこともできる！



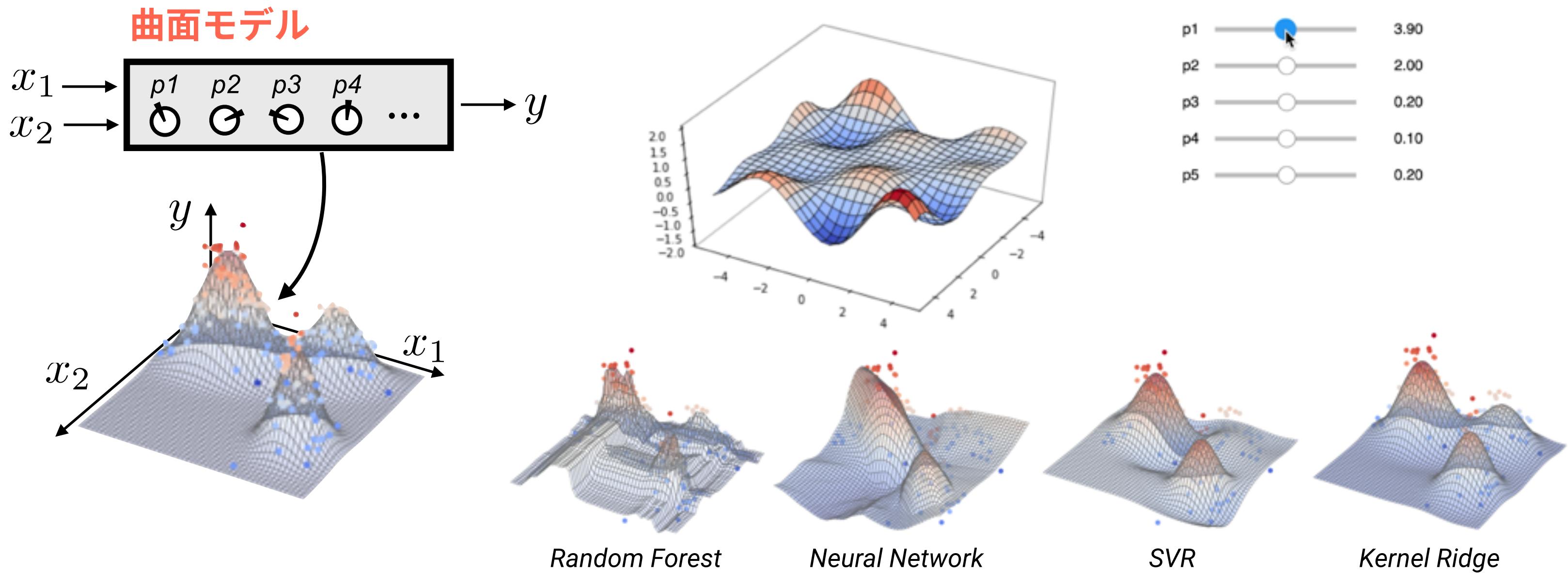
仕組みは曲面フィッティングによるデータ内挿

「曲面モデル」の内部パラメタ値を調整して見本点にあうようフィッティングする



仕組みは曲面フィッティングによるデータ内挿

「曲面モデル」の内部パラメタ値を調整して見本点にあうようフィッティングする



半世紀前に生まれた「希望的な呼称」による幻想にご注意

現在の機械学習は一般に想像するSF的な「人工知能(AI)」とはかなりかけ離れているが、「データを予測に変える」機能があまりに強力なため、私たちの日常生活から今後の社会のカタチにまで影響を及ぼそうとしている…

「人工知能」「機械学習」などの希望的な呼称は本質をミスリードしやすいのでご注意を！

https://spectrum.ieee.org/files/11920/10_Spectrum_2021.pdf

SIGART Newsletter No. 57 April 1976

ARTIFICIAL INTELLIGENCE MEETS NATURAL STUPIDITY

Drew McDermott

MIT AI Lab Cambridge, Mass 02139

As a field, artificial intelligence has always been on the border of respectability, and therefore on the border of crackpottery. Many critics <Dreyfus, 1972>, <Lighthill, 1973> have urged that we are over the border. We have been very defensive toward this charge, drawing ourselves up with dignity when it is made and folding the cloak of Science about us. On the other hand, in private, we have been justifiably proud of our willingness to explore weird ideas, because pursuing them is the only way to make progress.

Leo Breimanの3レッスン: Rashomon, Occam, Bellman

<https://projecteuclid.org/euclid.ss/1009213726> (Open Access)

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

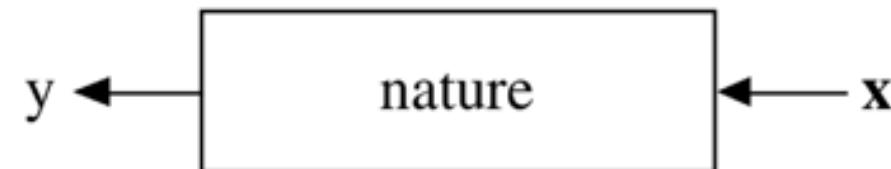
Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

Leo Breiman (1928-2005)



- UC Berkeley名誉教授
- 2005 SIGKDD Innovation Award
- Probability Theorist → Consultant → Statistician

Leo Breimanの3レッスン: Rashomon, Occam, Bellman



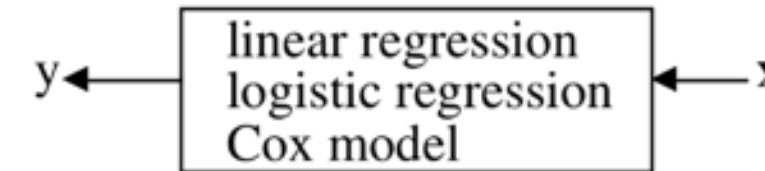
There are **two goals** in analyzing the data:

Prediction. To be able to predict what the responses are going to be to future input variables;

Information. To extract some information about how nature is associating the response variables to the input variables.

- 決定木アンサンブル
- 自動微分フレームワーク (深層学習)

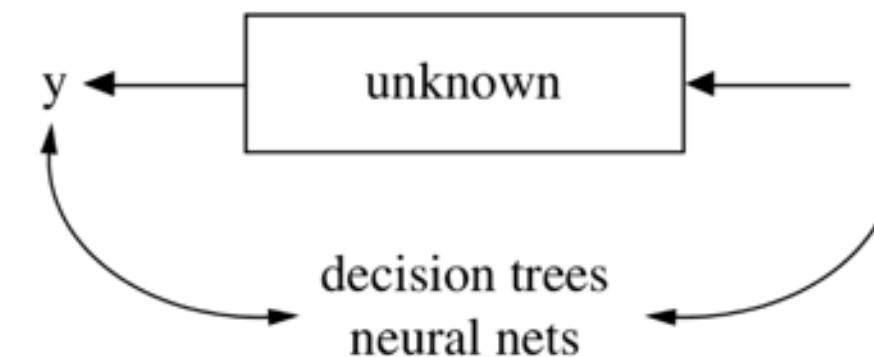
The Data Modeling Culture



Model validation. Yes–no using goodness-of-fit tests and residual examination.

Estimated culture population. 98% of all statisticians.

The Algorithmic Modeling Culture



Model validation. Measured by predictive accuracy.
Estimated culture population. 2% of statisticians, many in other fields.

Leo Breimanの3レッスン: Rashomon, Occam, Bellman

7.3 Recent Lessons

The advances in methodology and increases in predictive accuracy since the mid-1980s that have occurred in the research of machine learning has been phenomenal. There have been particularly exciting developments in the last five years. What has been learned? The three lessons that seem most important to one:

Rashomon: the multiplicity of good models;

Occam: the conflict between simplicity and accuracy;

Bellman: dimensionality—curse or blessing.

Leo Breimanの3レッスン: Rashomon, Occam, Bellman

Rashomon, Occam, Bellman

今でもこの3点はOpenな問題を孕んで機械学習の中心問題として研究されている

- "Rashomon": 良いモデルの多重性(非一意性)

同程度の良い予測精度を持つ全く異なるモデルがたくさん存在する

- "Occam": モデルの解釈性と予測精度のコンフリクト

モデルのシンプルさ(解釈性)と予測精度の両立はとても難しい

- "Bellman": 高次元データが引き起こすメリットとデメリット

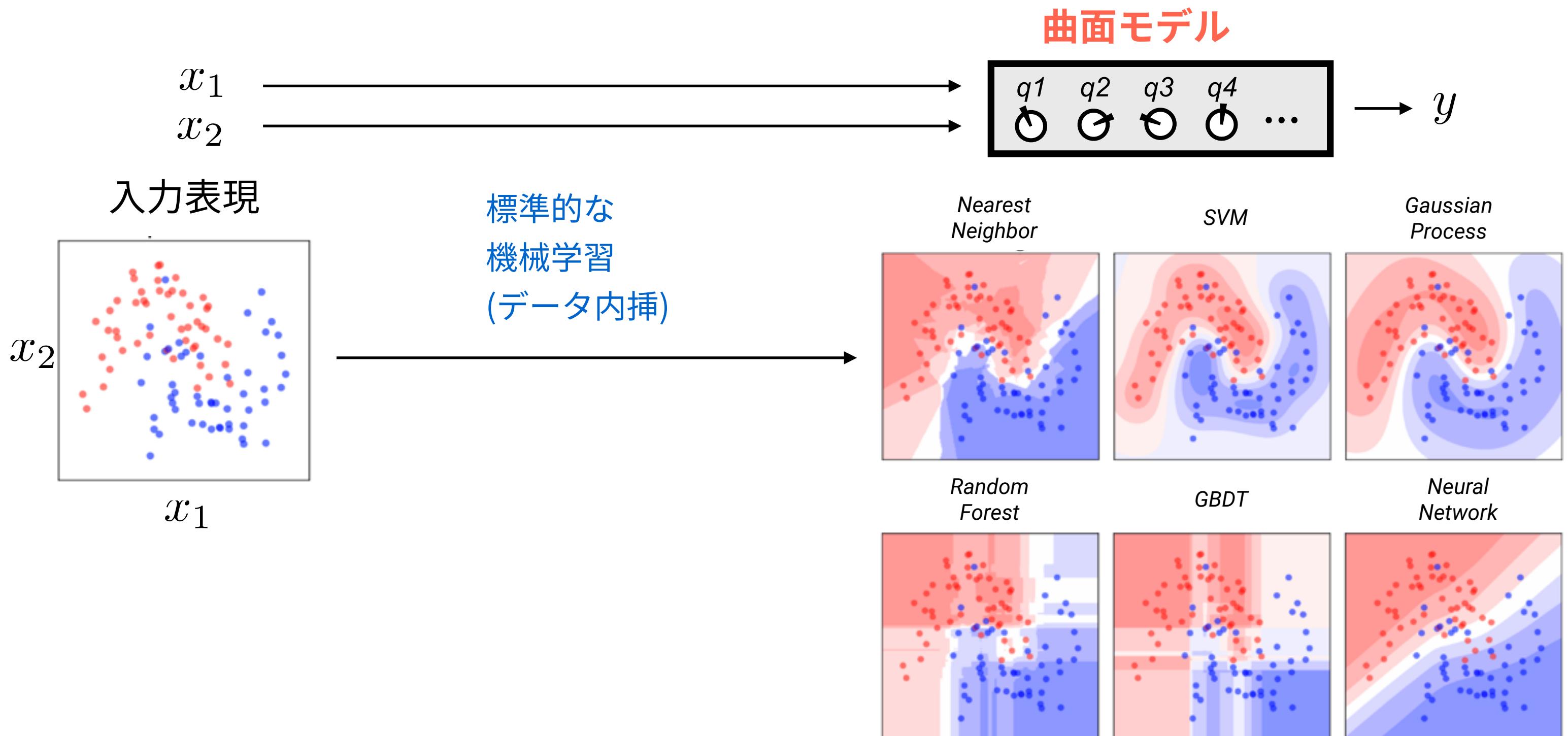
高次元な表現(関係しそうなできるだけ多くの変量)を扱うべきなのか

伝統的な統計学のように支配的な少数の変量を検討し分析すべきなのか

→ キッチンシンク回帰(思いつく変数全部入りモデル)・特徴量

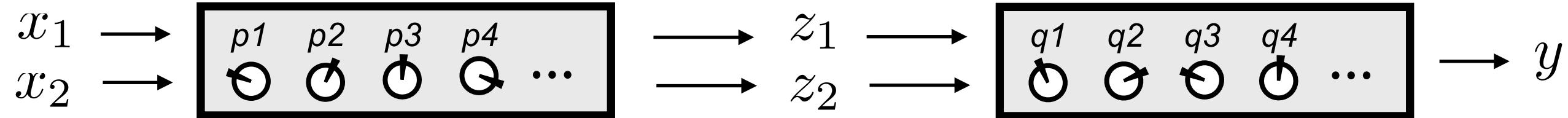
エンジニアリングと擬似相関・Rashomon効果の増大リスク

深層学習と表現学習

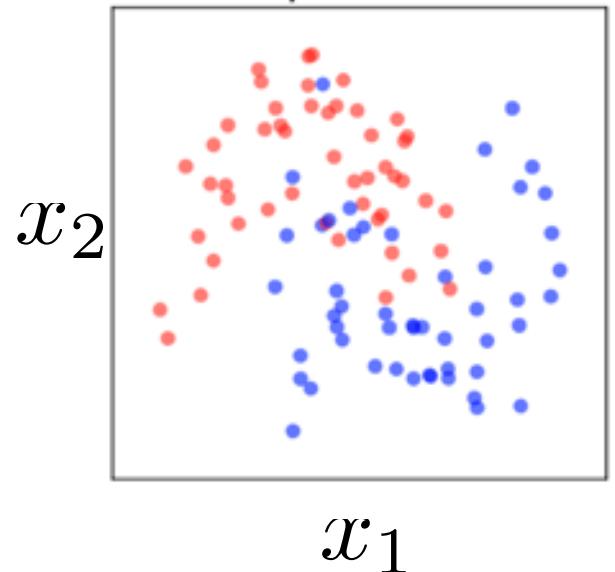


深層学習と表現学習

変数変換(表現学習)

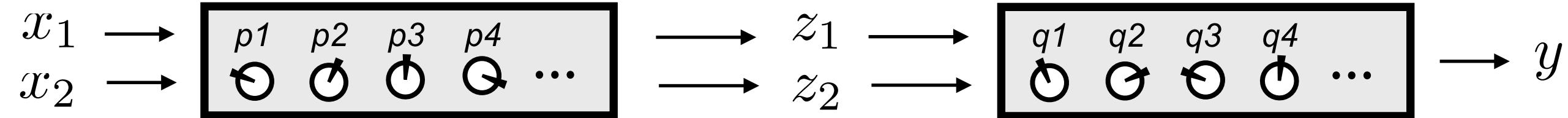


入力表現



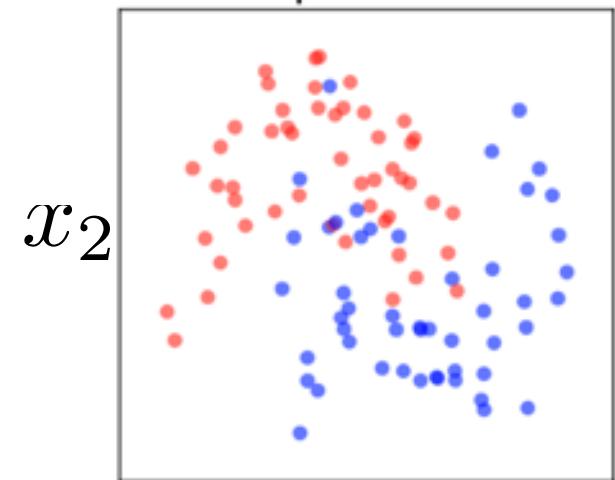
深層学習と表現学習

変数変換(表現学習)

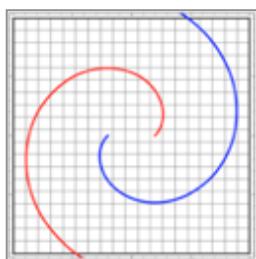
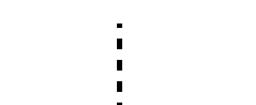


曲面モデル

入力表現

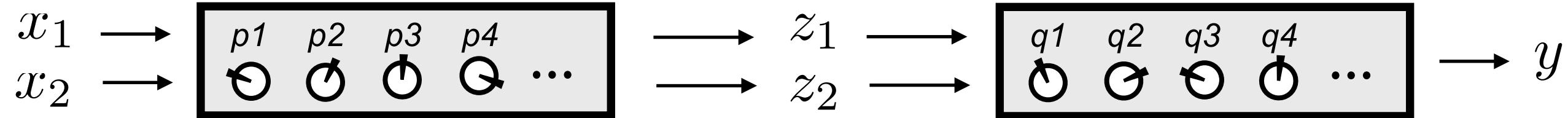


x_1



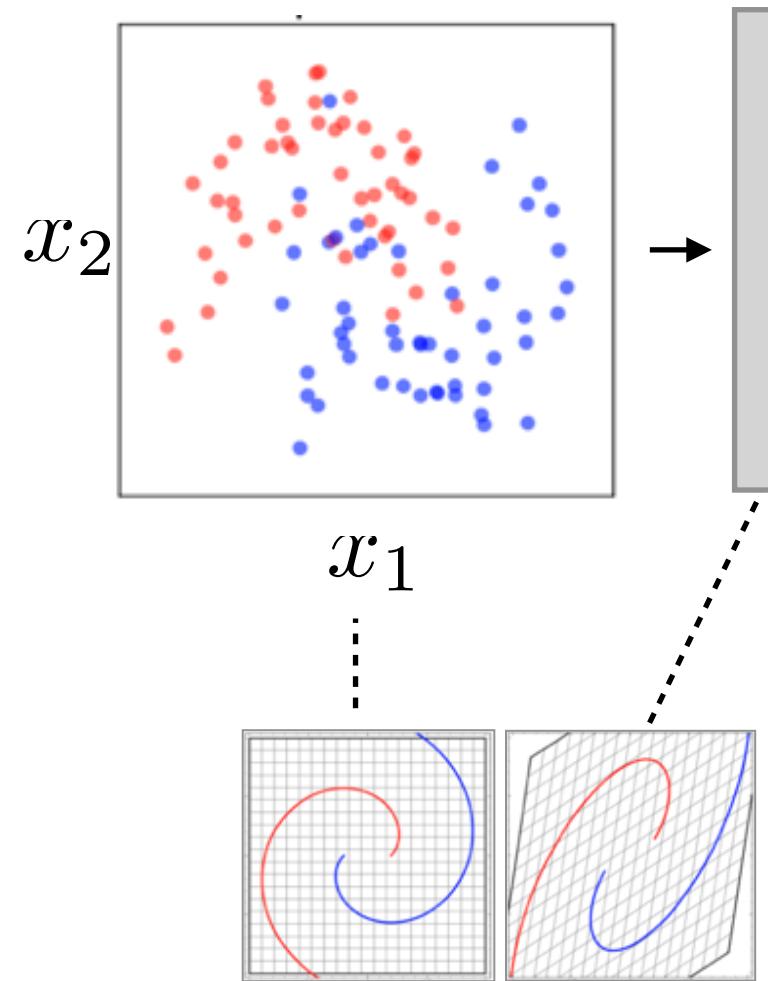
深層学習と表現学習

変数変換(表現学習)



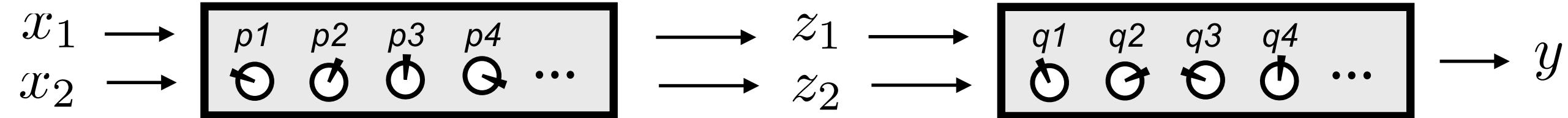
曲面モデル

入力表現



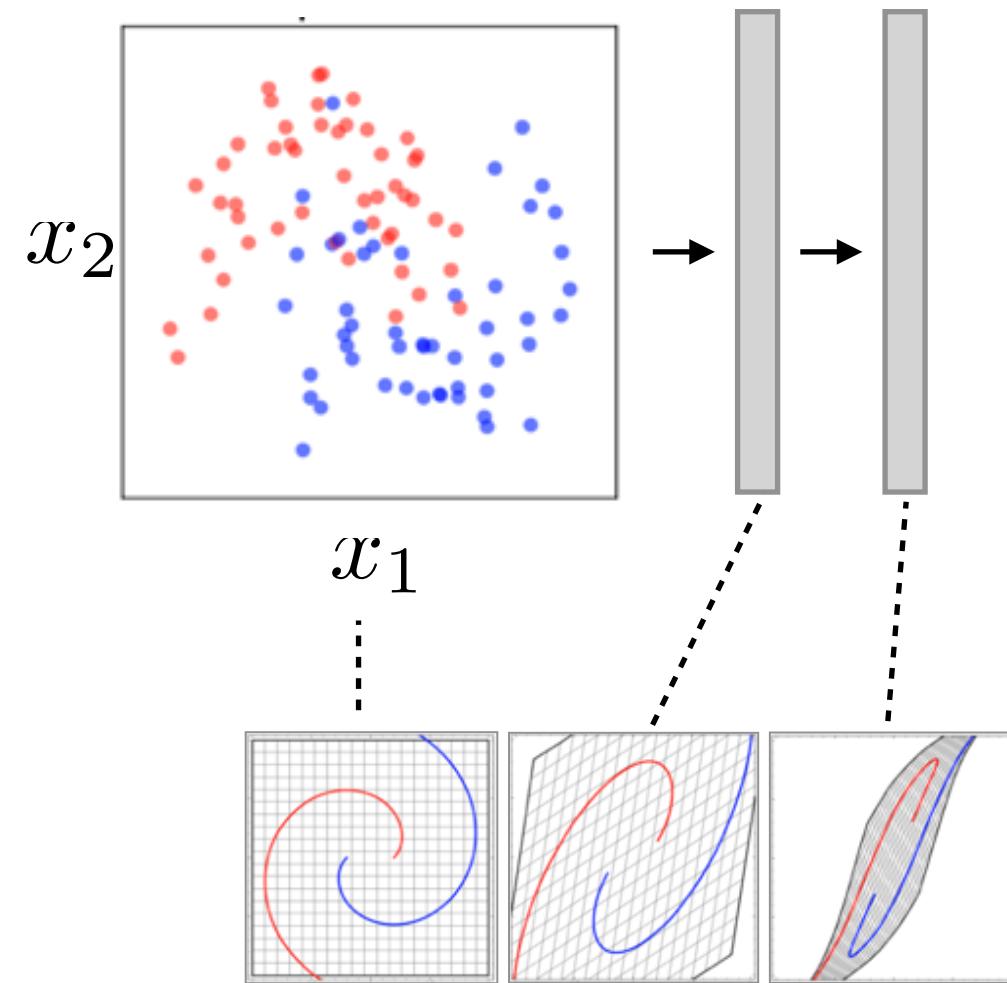
深層学習と表現学習

変数変換(表現学習)



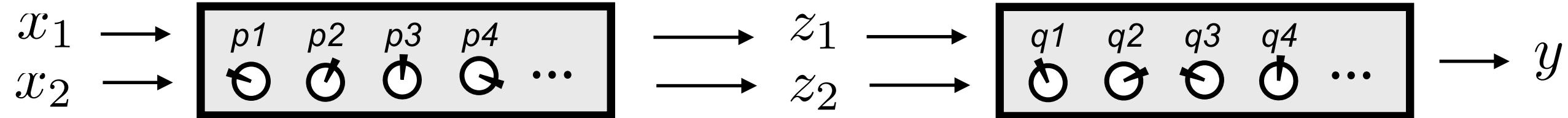
曲面モデル

入力表現



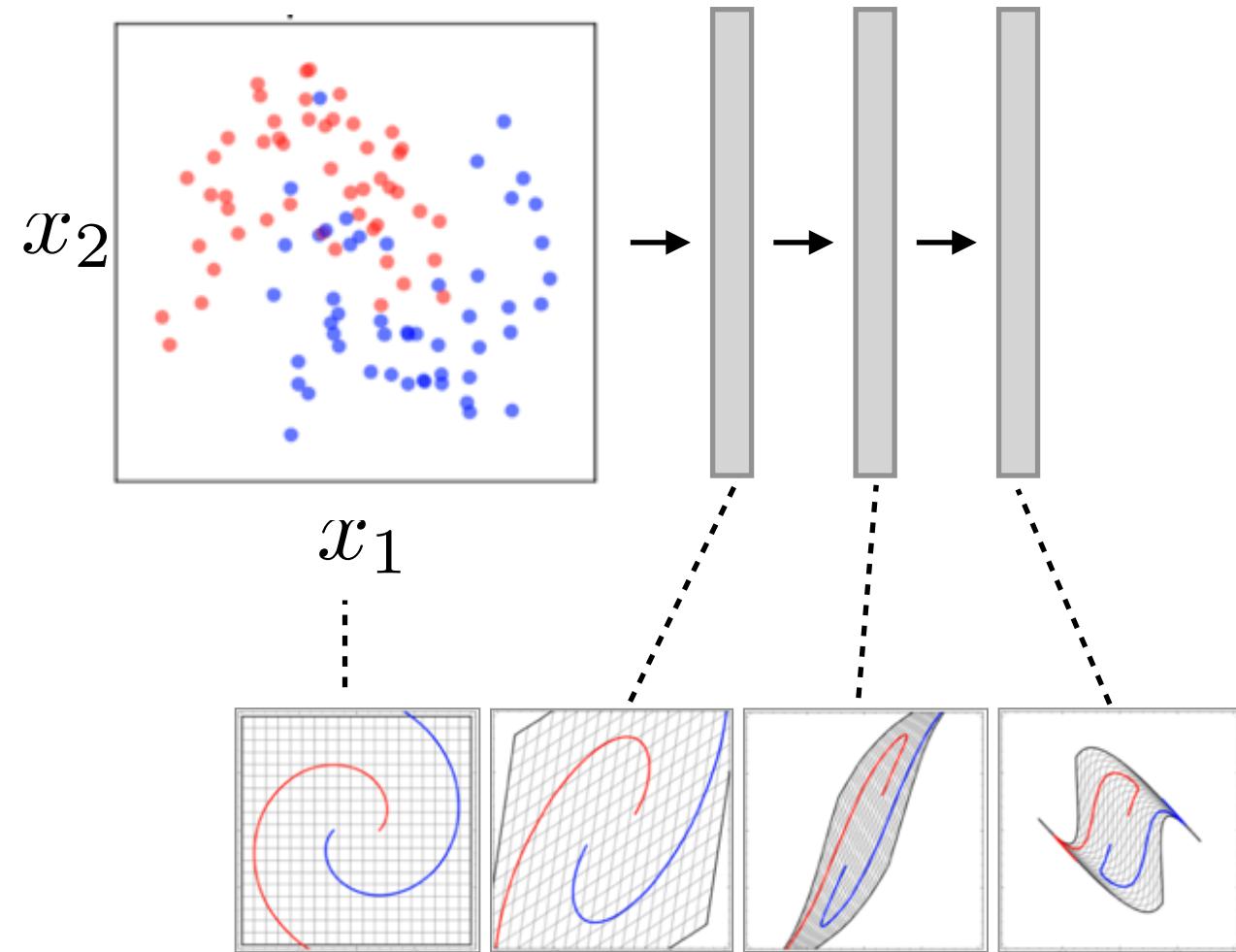
深層学習と表現学習

変数変換(表現学習)



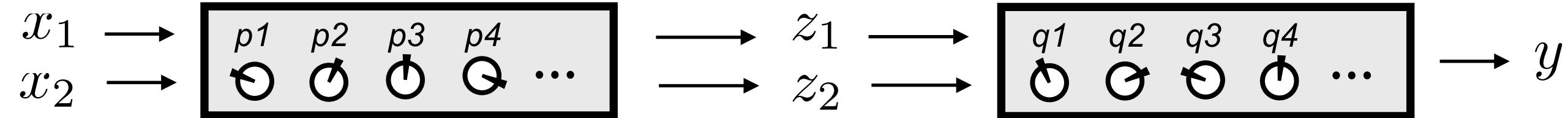
曲面モデル

入力表現

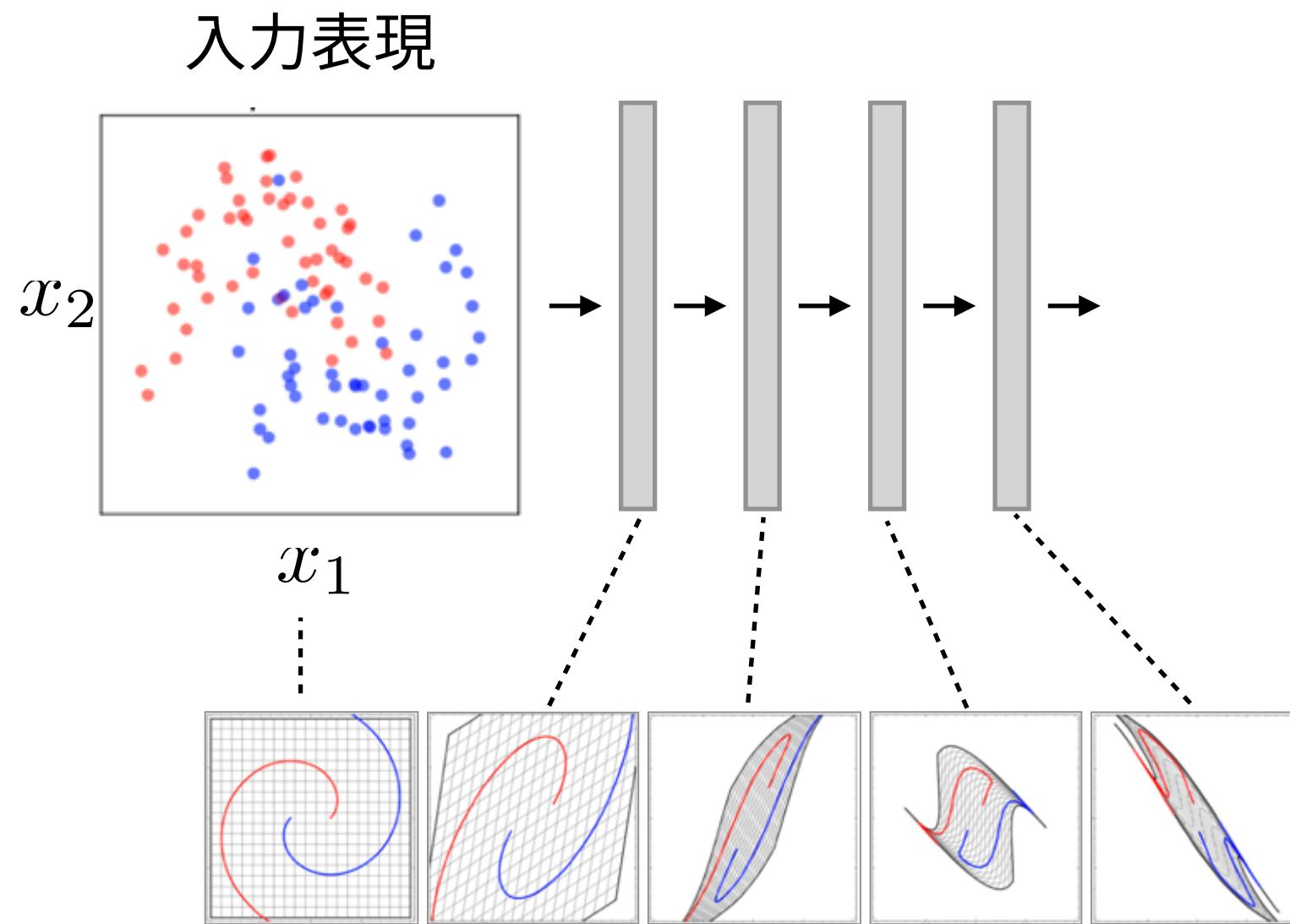


深層学習と表現学習

変数変換(表現学習)

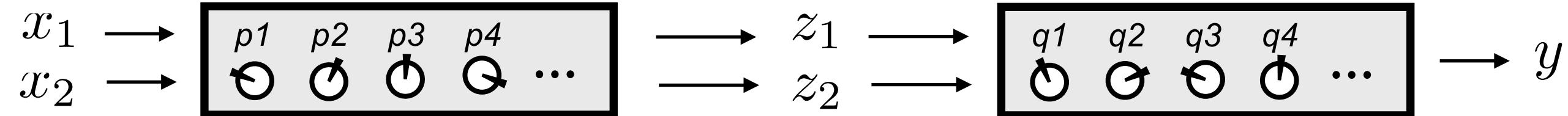


曲面モデル

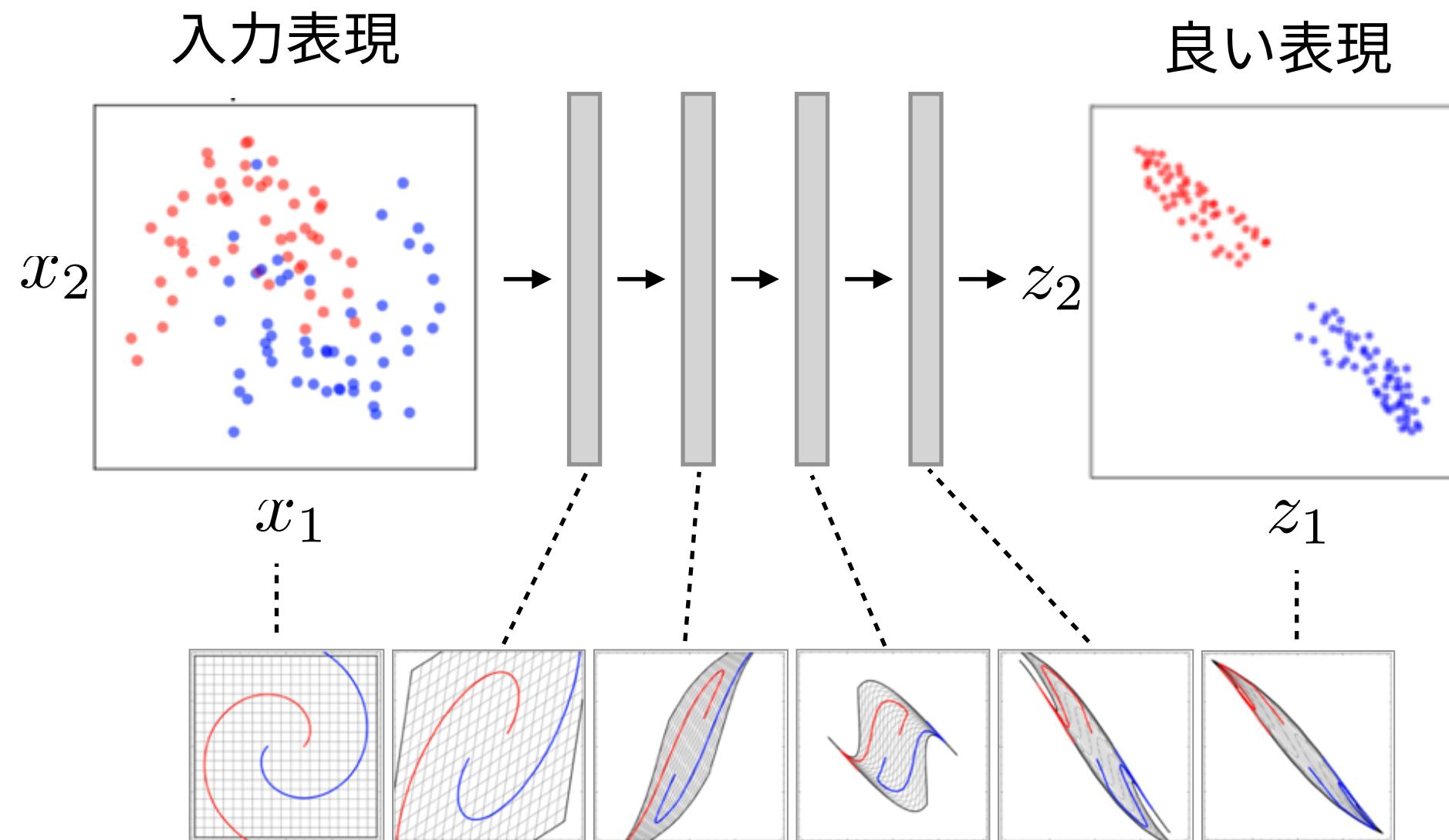


深層学習と表現学習

変数変換(表現学習)

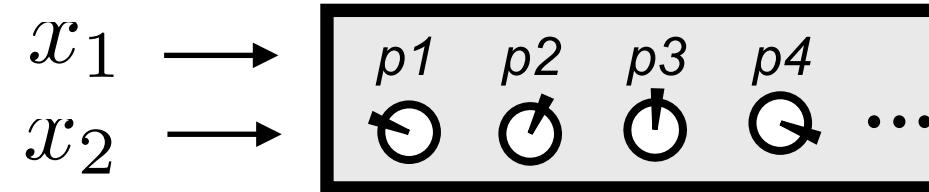


曲面モデル

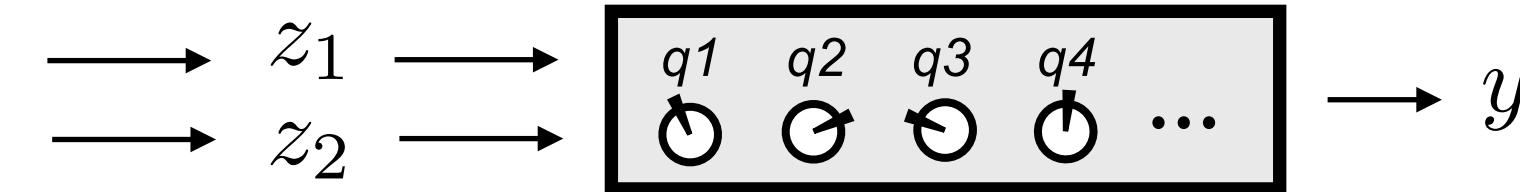


深層学習と表現学習

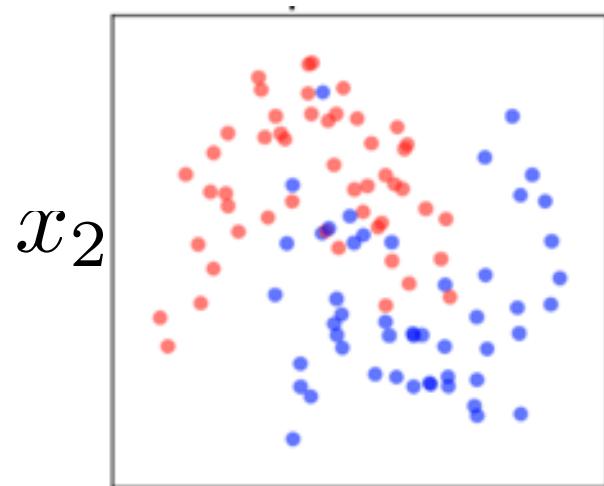
変数変換(表現学習)



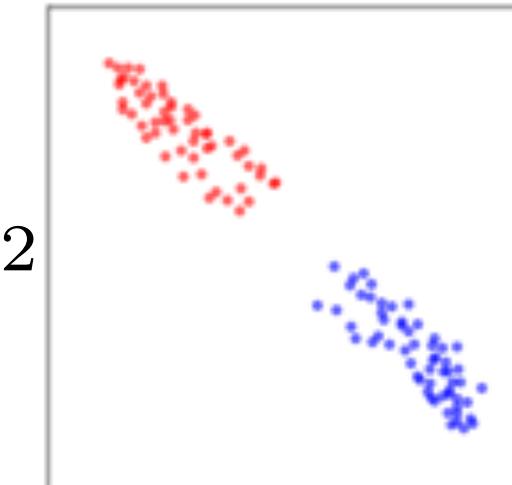
曲面モデル



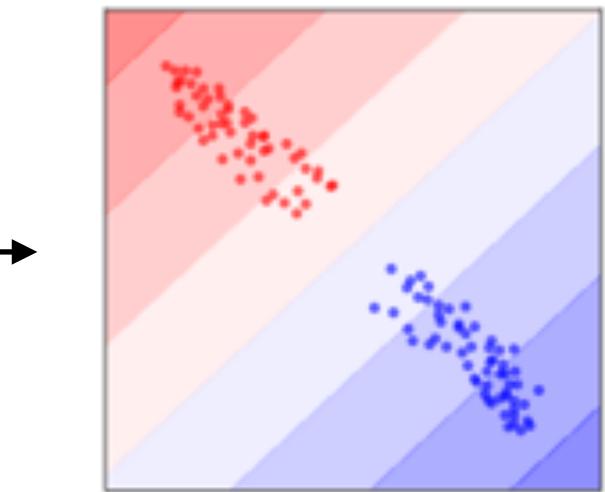
入力表現



良い表現



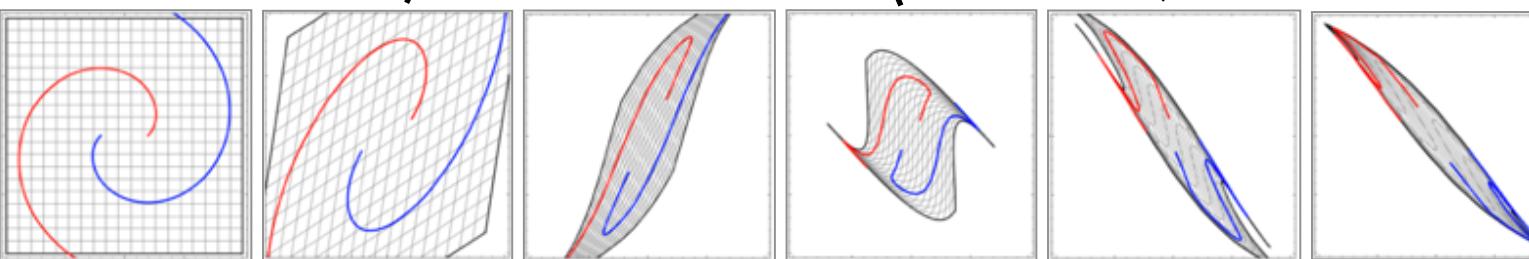
標準的な
機械学習
(データ内挿)



x_1

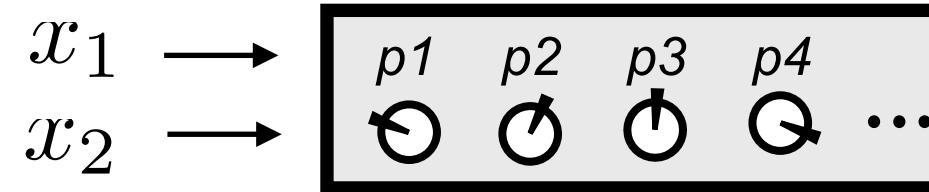
z_1

シンプル(線形)で十分

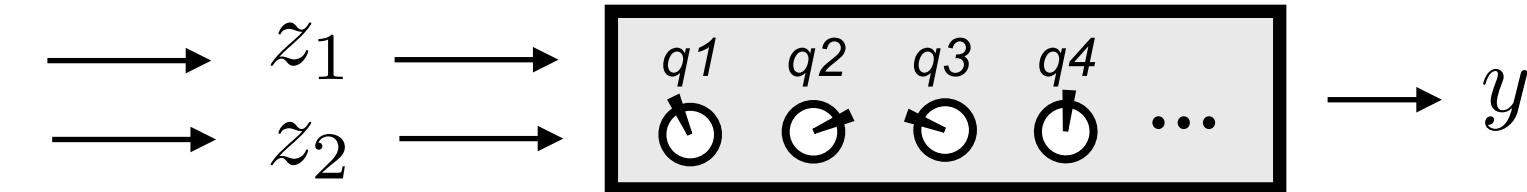


深層学習と表現学習

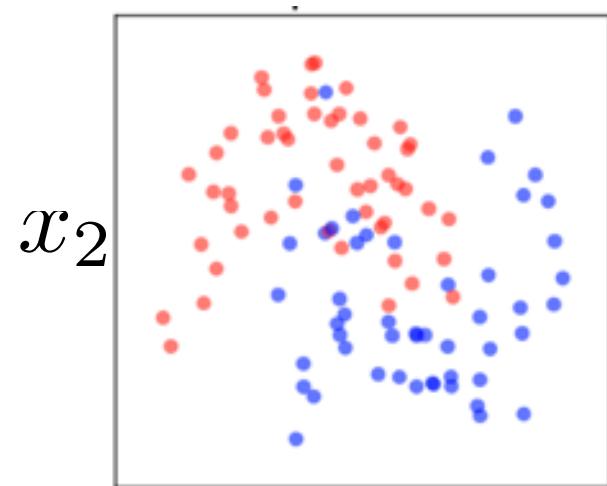
変数変換(表現学習)



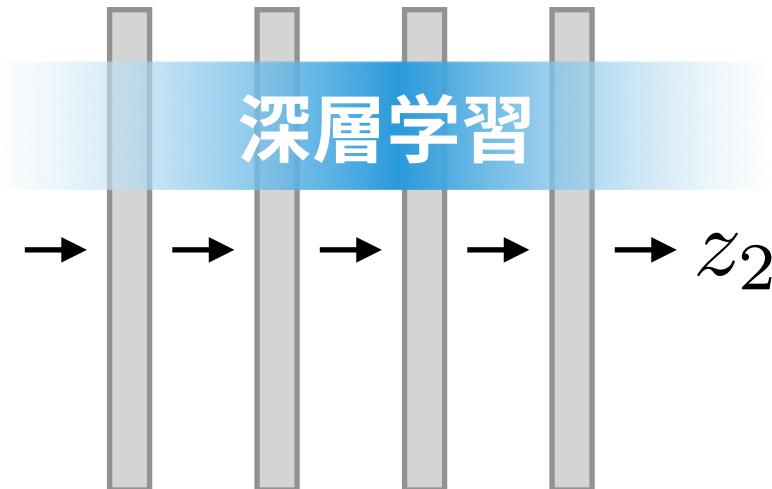
曲面モデル



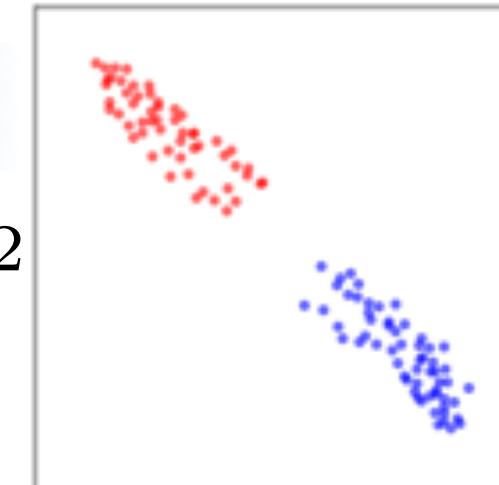
入力表現



深層学習

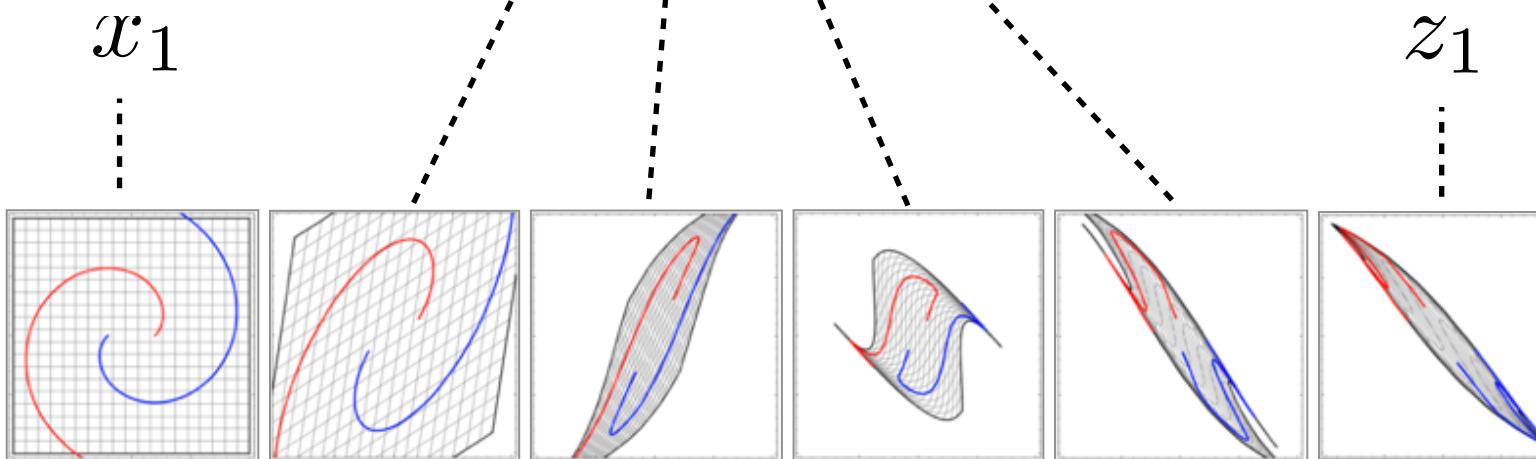


良い表現



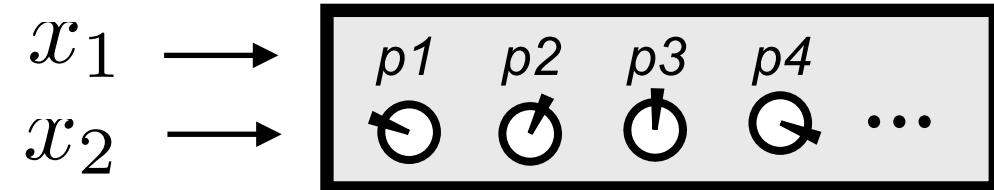
標準的な
機械学習
(データ内挿)

シンプル(線形)で十分

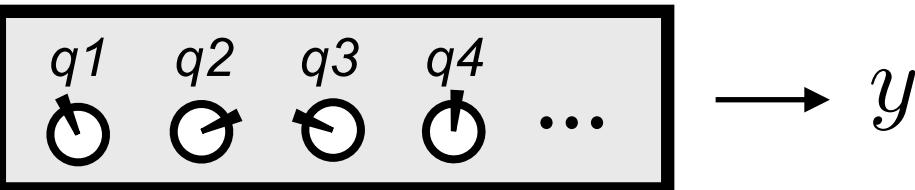


深層学習と表現学習

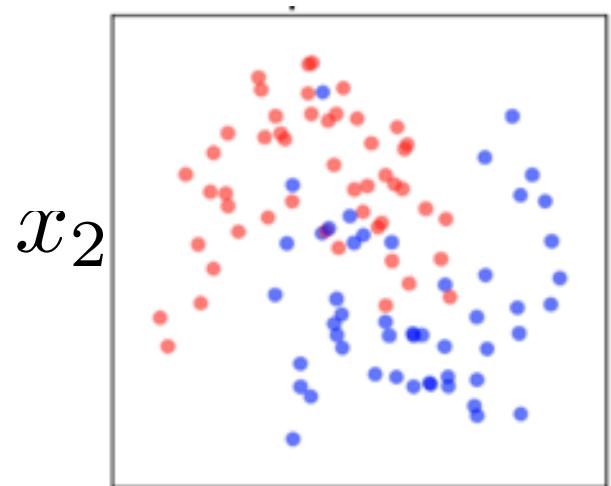
変数変換(表現学習)



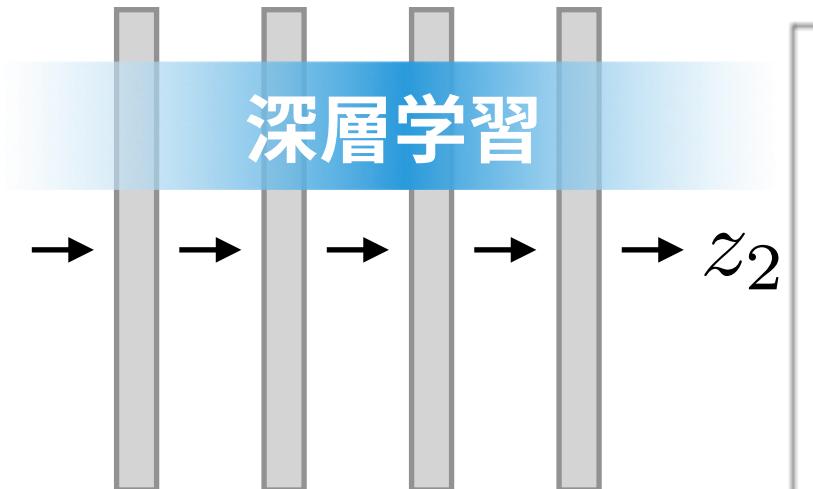
曲面モデル



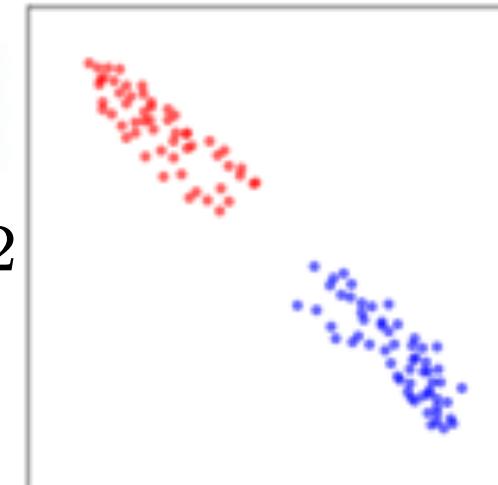
入力表現



深層学習

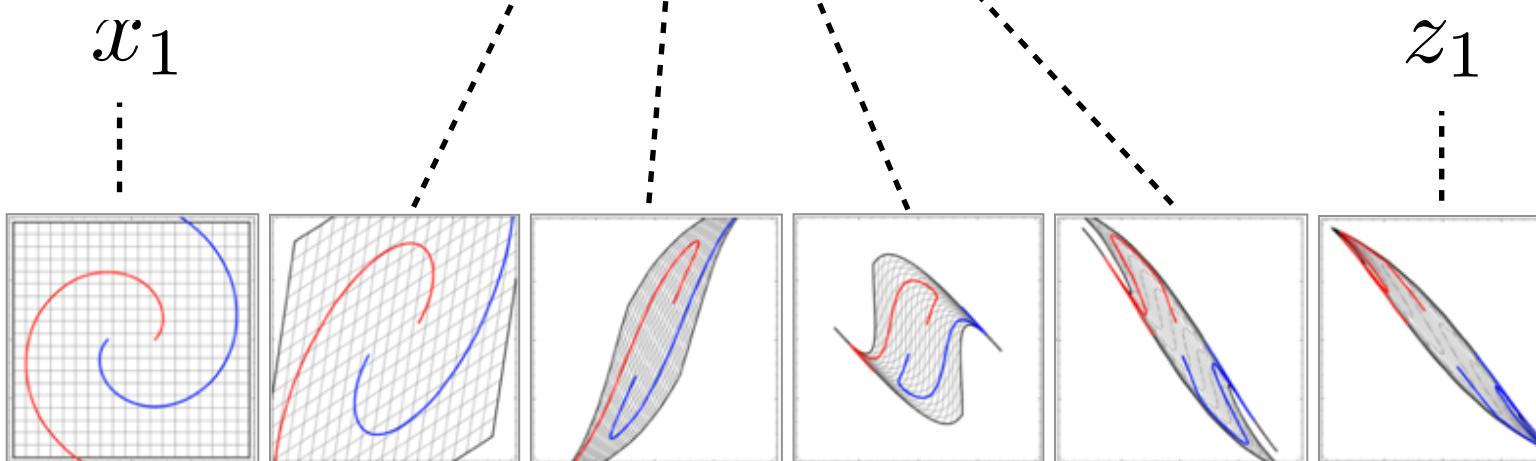


良い表現



標準的な
機械学習
(データ内挿)

シンプル(線形)で十分

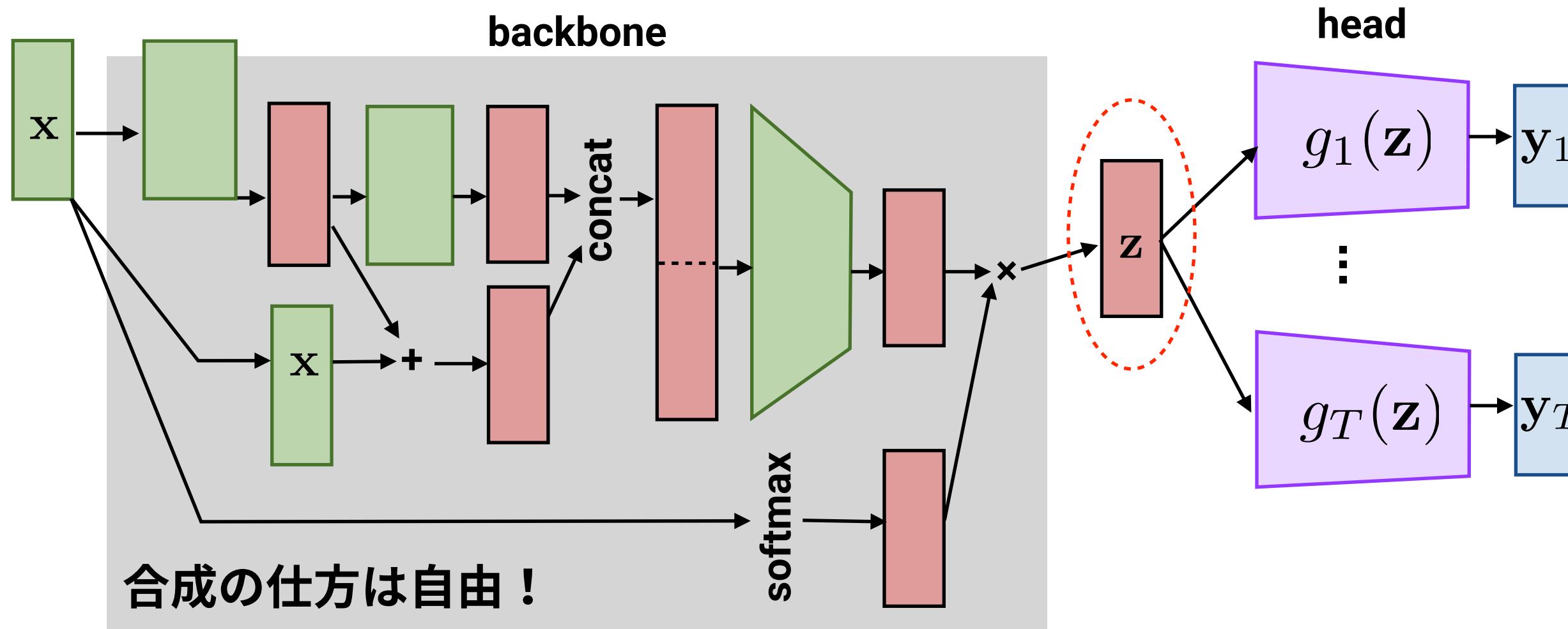


データ内挿は入力表現ではなく
「良い表現」で行う

微分可能プログラミング

https://en.wikipedia.org/wiki/Differentiable_programming

合成関数を計算グラフで表現すれば勾配は自動微分(aka backprop)で計算可 → 勾配法で最適化

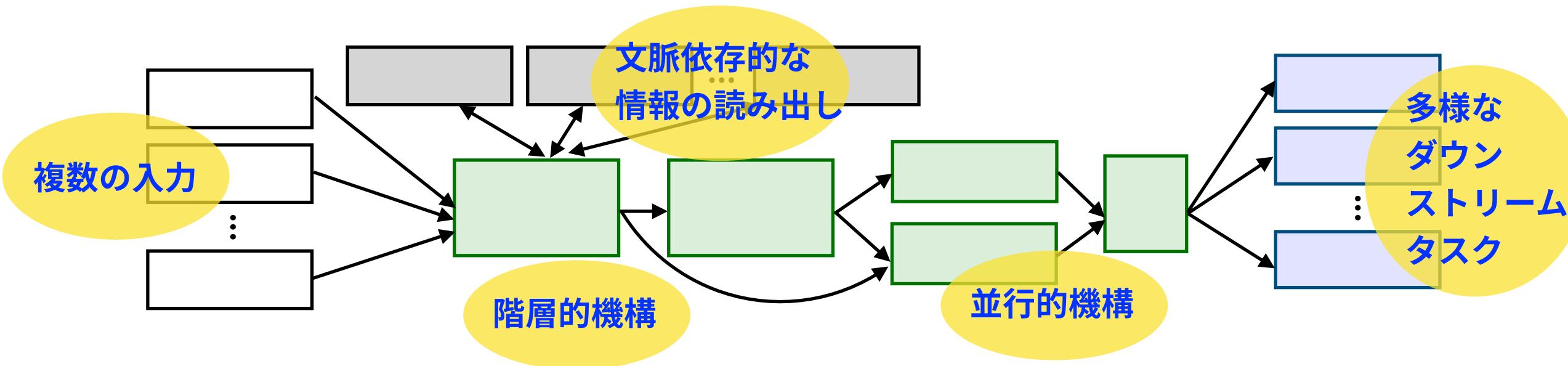


"All numerical computations are ultimately **compositions of a finite set of elementary operations** for which derivatives are known (Verma, 2000; Griewank and Walther, 2008)"

注意機構と構成的学习

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

- 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



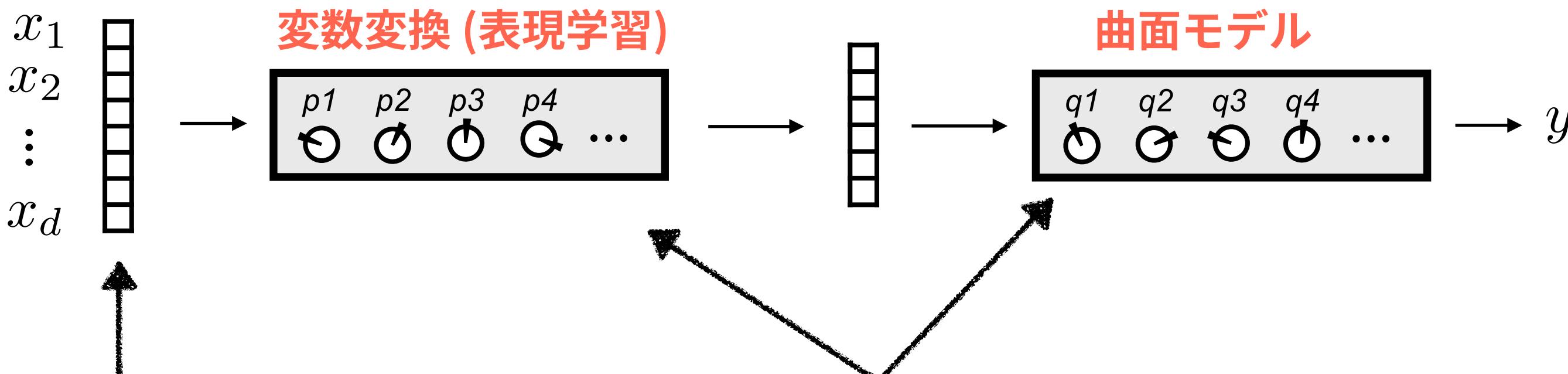
$$\begin{array}{c} \text{Query} \quad \text{Key} \quad \text{Value} \\ \downarrow \quad \downarrow \quad \downarrow \\ \alpha(x, x_i) y_i \end{array}$$

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

困難① 現代の機械学習モデルは多量のデータを必要とする



① 高次元性：入力変数が多すぎ！

- ✓ 機械学習は入力されてない情報を全く考慮してくれない… (擬似相関リスク)
- ✓ とりあえず色々な変数を入れがち

画像そのままを入力する場合

20×20 ピクセルのカラー画像 → 1200変数

1000×1000 ピクセルのカラー画像 → 300万変数

② 過剰パラメタ化：パラメタ数が多すぎ！

画像 ResNet50: 2600万パラメタ
ResNet101: 4500万パラメタ
EfficientNet-B7: 6600万パラメタ
VGG19: 1億4400万パラメタ

言語 12-layer, 12-heads BERT: 1億1000万パラメタ
24-layer, 16-heads BERT: 3億3600万パラメタ
GPT-2 XL: 15億5800万パラメタ
GPT-3: 1750億パラメタ

困難② 羅生門効果とUnderspecification

羅生門効果：良い機械学習モデルの多重性（非一意性）

高い予測精度を持つ機械学習モデルは一つのデータセットからたくさん作れる！

有限データから見積もる予測精度では指数的広さのモデル探索空間をspecifyしきれない…

困難② 羅生門効果とUnderspecification

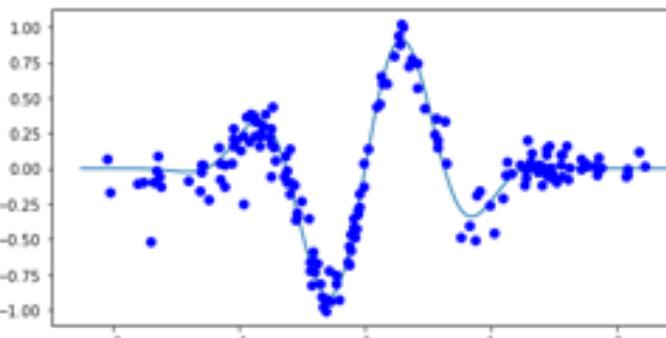
羅生門効果：良い機械学習モデルの多重性（非一意性）

高い予測精度を持つ機械学習モデルは一つのデータセットからたくさん作れる！

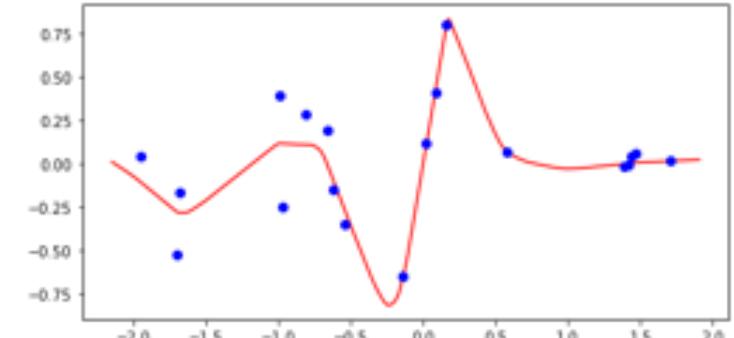
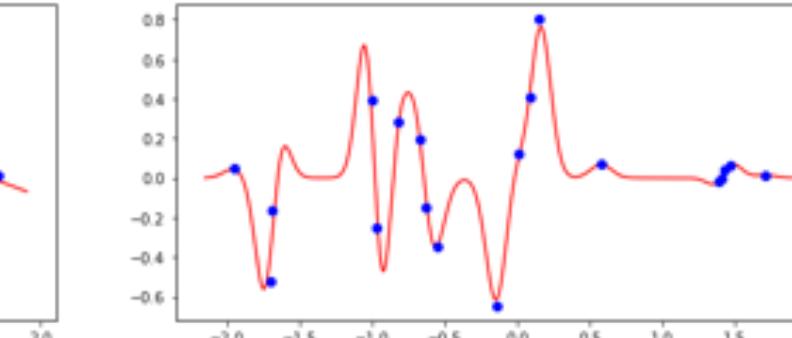
有限データから見積もる予測精度では指数的広さのモデル探索空間をspecifyしきれない…

- ✓ 「どのモデルが求める真実なの？」と考えてしまうと、まさに真実は「**藪の中**」…
→ **複数の手法による多角的解釈**が鉄則 "all models are wrong but some models are useful"
- ✓ 実際には**本質的にデータが足りてない(Underspecification)**ことで多重性はさらに悪化

だいたいの方法で類似



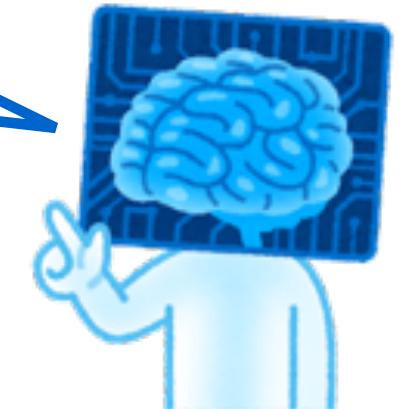
手法やモデルによって予測時の挙動にかなり差が出てしまう



Underspecificationは"ビッグデータ"事例でも広く起こっている？

ショボい認知能力のおまえら人間にとったら「ビッグ」データかもしらんけど、
ホンマに必要な情報量からしたらハナクソみたいなもんやな！

by ディープラーニング様



[https://ai.googleblog.com/2021/10/
how-underspecification-presents.html](https://ai.googleblog.com/2021/10/how-underspecification-presents.html)



The latest from Google Research

How Underspecification Presents Challenges for Machine Learning

Monday, October 18, 2021

Posted by Alex D'Amour and Katherine Heller, Research Scientists, Google Research

Machine learning (ML) models are being used more widely today than ever before and are becoming increasingly impactful. However, they often exhibit unexpected behavior when they are used in real-world domains. For example, computer vision models can exhibit surprising sensitivity to irrelevant features, while natural language processing models can depend unpredictably on demographic correlations not directly indicated by the text. Some reasons for these failures are well-known: for example, training ML models on poorly curated data, or training models to solve prediction problems that are structurally mismatched with the application domain. Yet, even when

<https://arxiv.org/abs/2011.03395>

arXiv.org > cs > arXiv:2011.03395

Search...

Help | Advanced S

Computer Science > Machine Learning

(Submitted on 6 Nov 2020 ([v1](#)), last revised 24 Nov 2020 (this version, v2))

Underspecification Presents Challenges for Credibility in Modern Machine Learning

Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D. Hoffman, Farhad Hormozdiari, Neil Houlsby, Shaobo Hou, Ghassen Jerfel, Alan Karthikesalingam, Mario Lucic, Yian Ma, Cory McLean, Diana Mincu, Akinori Mitani, Andrea Montanari, Zachary Nado, Vivek Natarajan, Christopher Nielson, Thomas F. Osborne, Rajiv Raman, Kim Ramasamy, Rory Sayres, Jessica Schrouff, Martin Seneviratne, Shannon Sequeira, Harini Suresh, Victor Veitch, Max Vladymyrov, Xuezhi Wang, Kellie Webster, Steve Yadlowsky, Taedong Yun, Xiaohua Zhai, D. Sculley

ML models often exhibit unexpectedly poor behavior when they are deployed in real-world domains. We identify underspecification as a key reason for these failures. An ML pipeline is underspecified when it can return many predictors with equivalently strong held-out performance in the training domain. Underspecification is common in modern ML pipelines, such as those based on deep learning. Predictors

現代の技術的関心はこの高次元性をどう手懐けるか

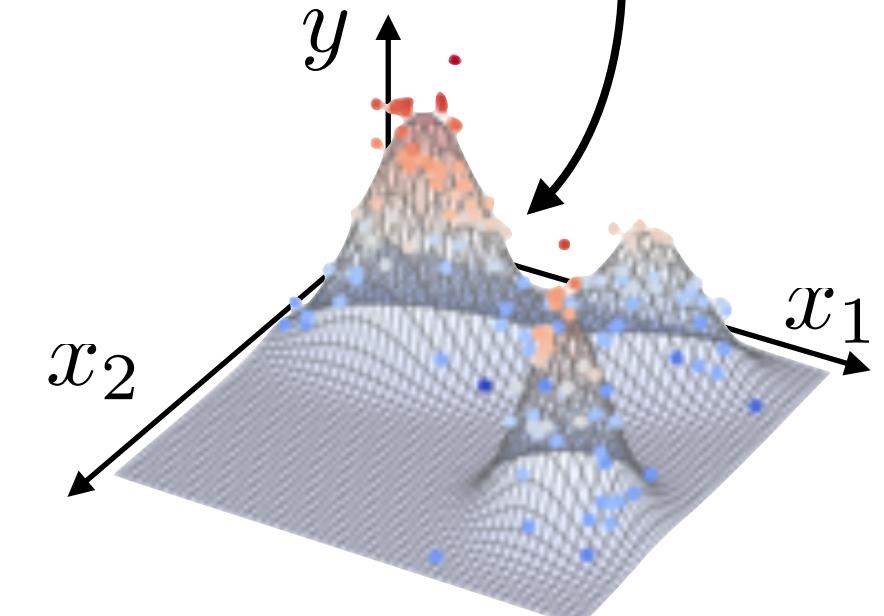
1. 確率的最適化・正則化 → モデルが大きい自由度の中で暴れまくらないよう動ける範囲を何とかして制御・制限・安定化する
2. 事前学習 (Warm Start) の転移 → 事前に得ておいたイイ感じのパラメタ初期値を使う
3. 帰納バイアスの設計

曲面モデルがどんな入出力関係でも表現できることが逆に擬似相関やUnderspecificationの問題を悪化させている



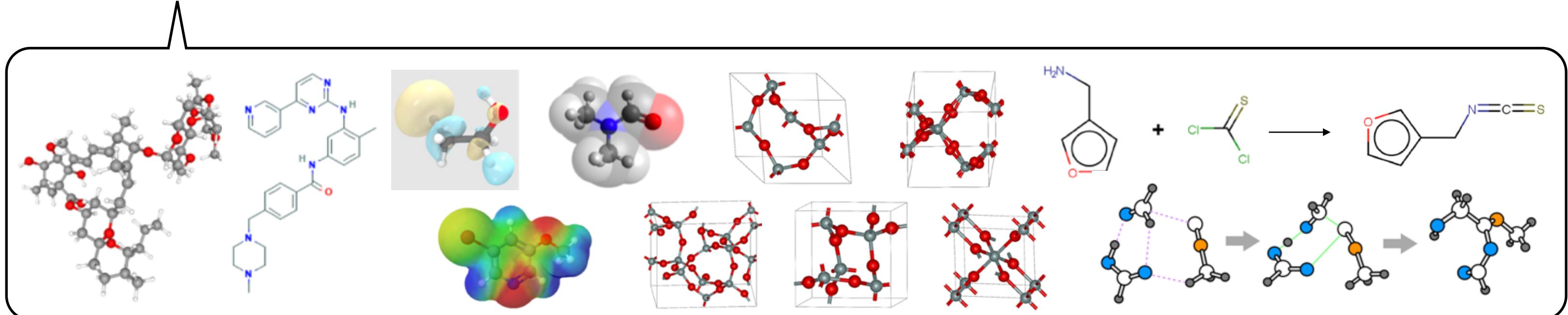
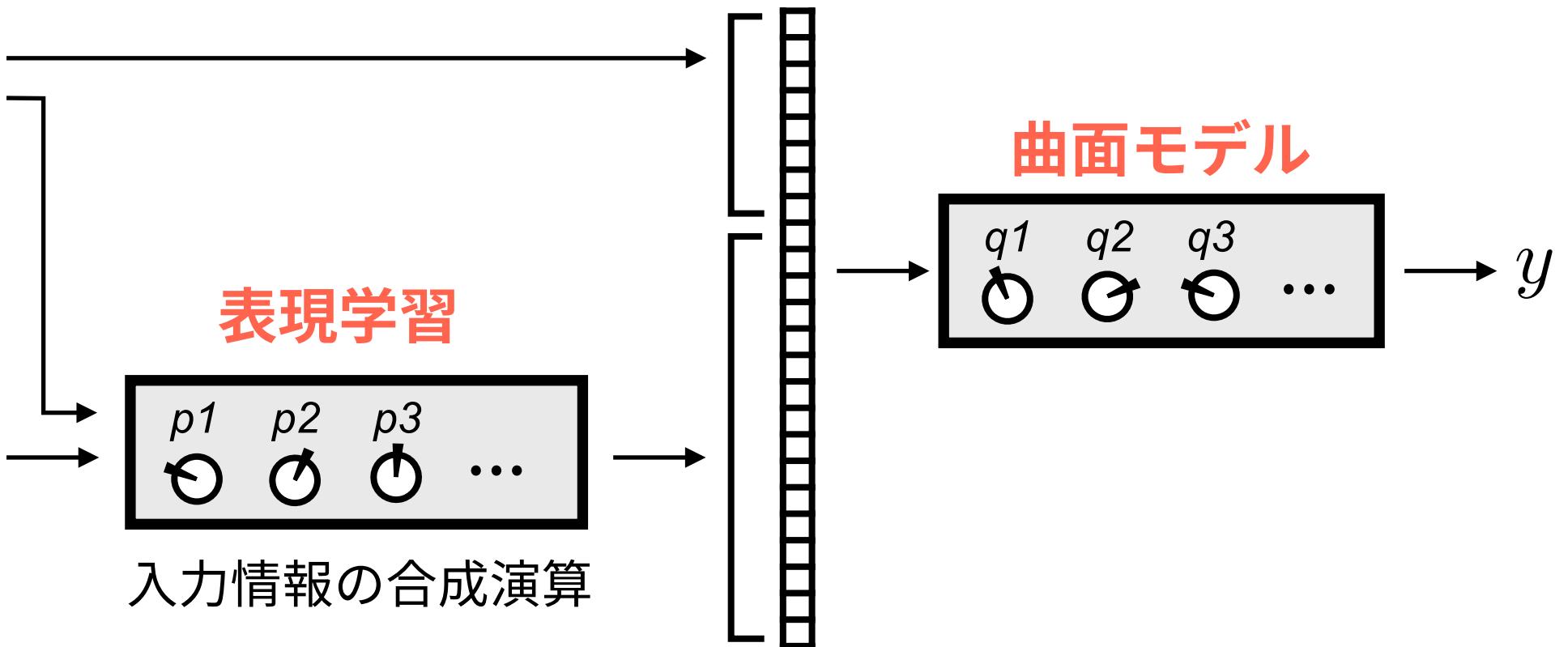
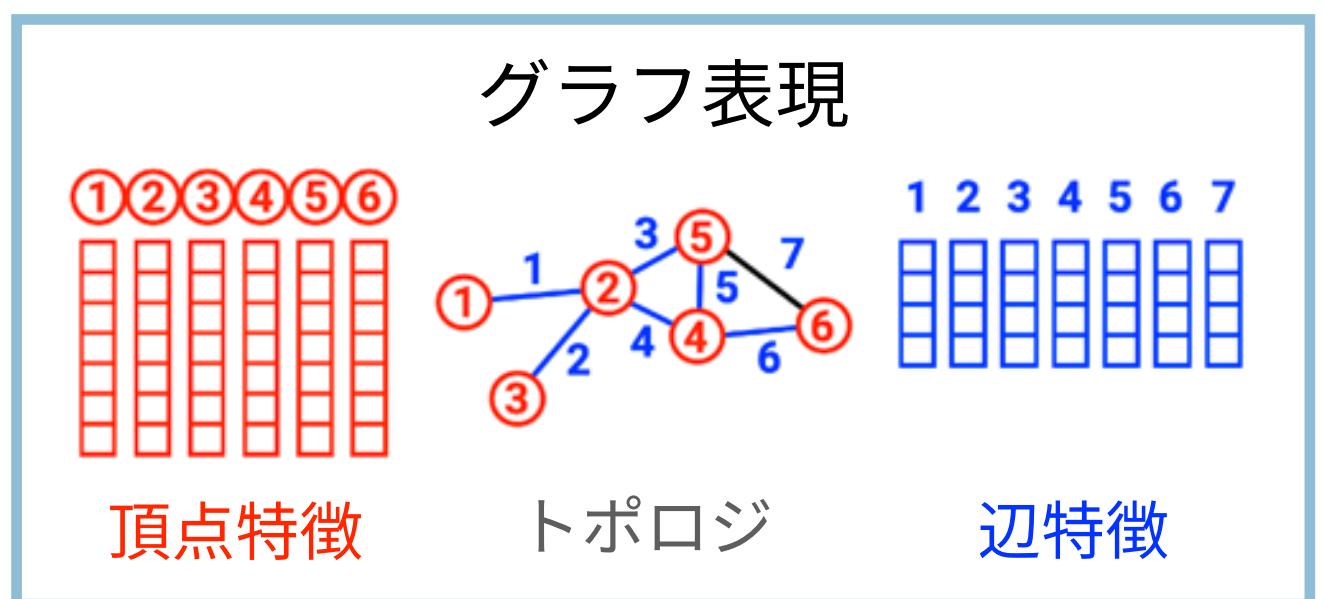
機械学習×化学：化学に適合した帰納バイアスのデザイン

化学的に妥当性を欠くようなモデルが意図せず表現されてしまわないように化学の知識や理論科学・計算化学の知見を総動員して**モデルの自由度を技術的に制限**する！



実例：分子の表現学習とGraph Neural Networks (GNNs)

分子・材料の環境/条件/標的/相互作用…



幾何的深層學習

GNNは幅広い幾何構造を統一的に扱える枠組み (機械學習のエルランゲン・プログラム!?)

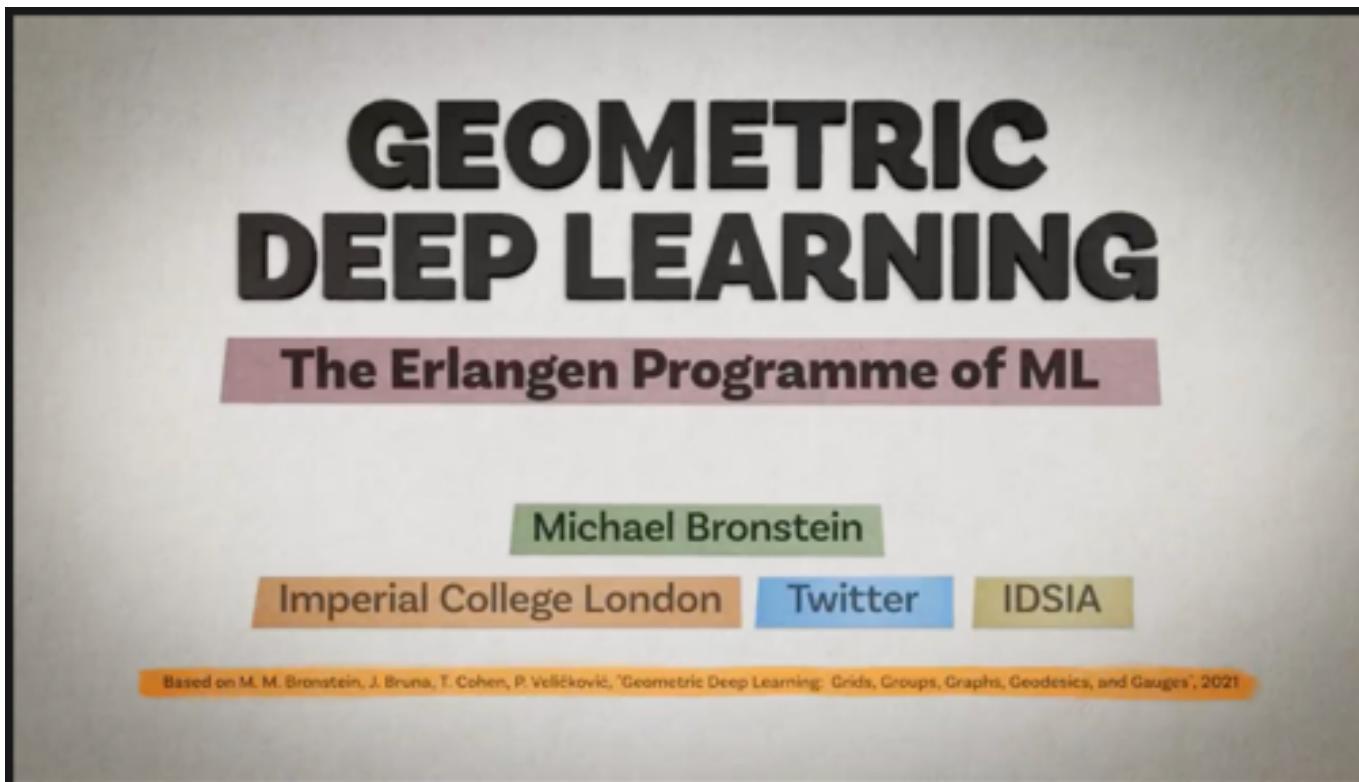
[Submitted on 27 Apr 2021 (v1), last revised 2 May 2021 (this version, v2)]

Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges

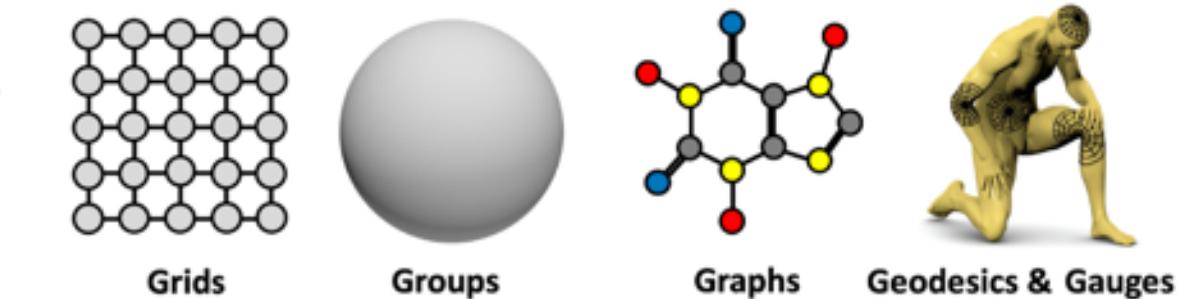
Michael M. Bronstein, Joan Bruna, Taco Cohen, Petar Veličković

<https://arxiv.org/abs/2104.13478>

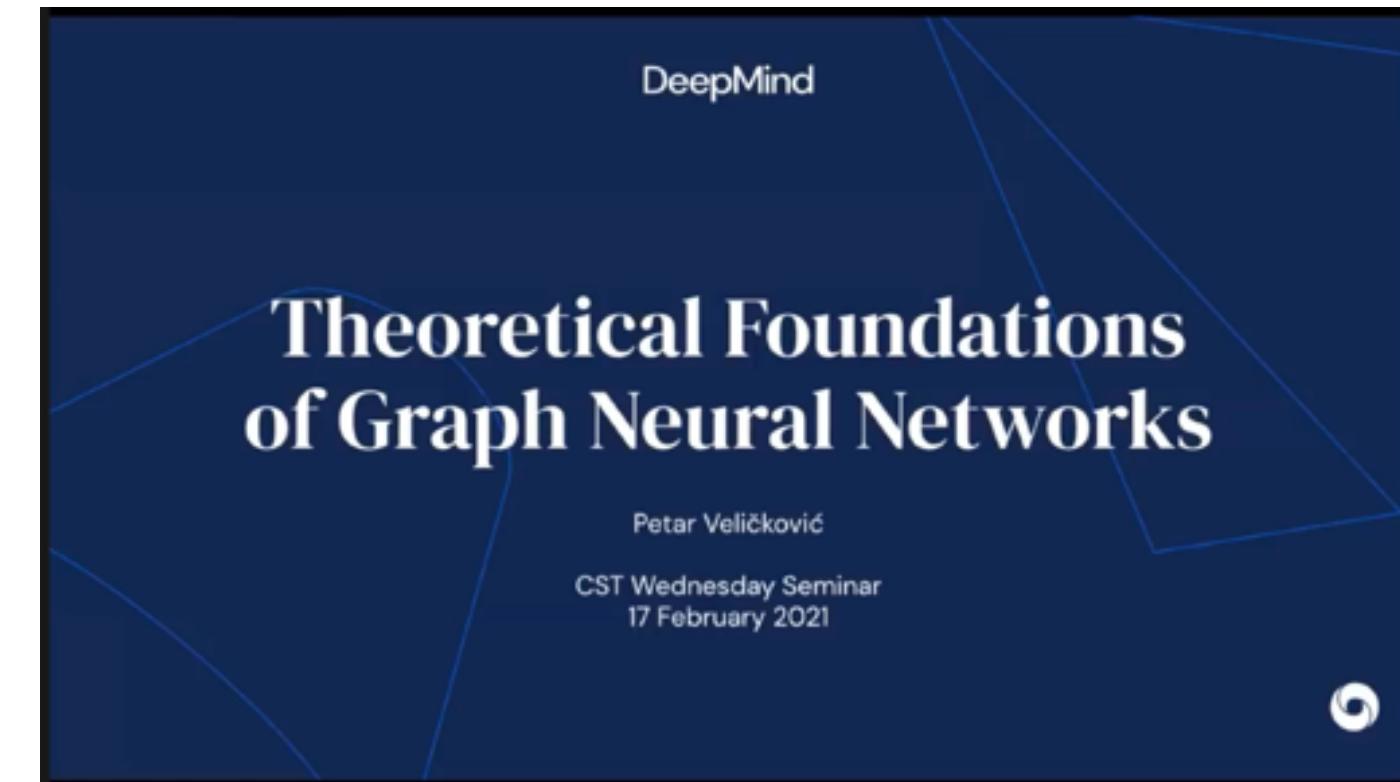
ICLR 2021 Keynote (Michael Bronstein)



<https://youtu.be/w6Pw4MOzMuo>



Seminar Talk (Petar Veličković)

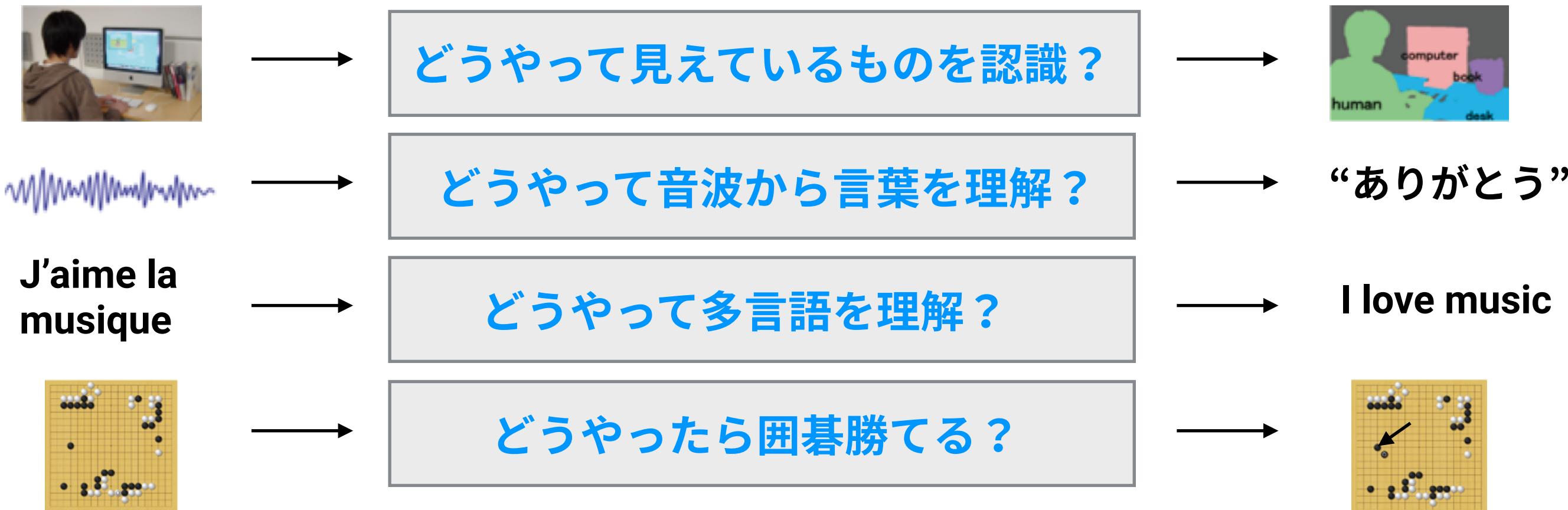


<https://youtu.be/uF53xsT7mjc>

機械学習×化学の真の問題

「予測ができる」ことは「理解」や「発見」ができるることを直接は意味しない！！

下記はどれも機械学習でかなり高精度な予測ができますが、それは私たちがその仕組みを理解できたことを少しでも意味するでしょうか？



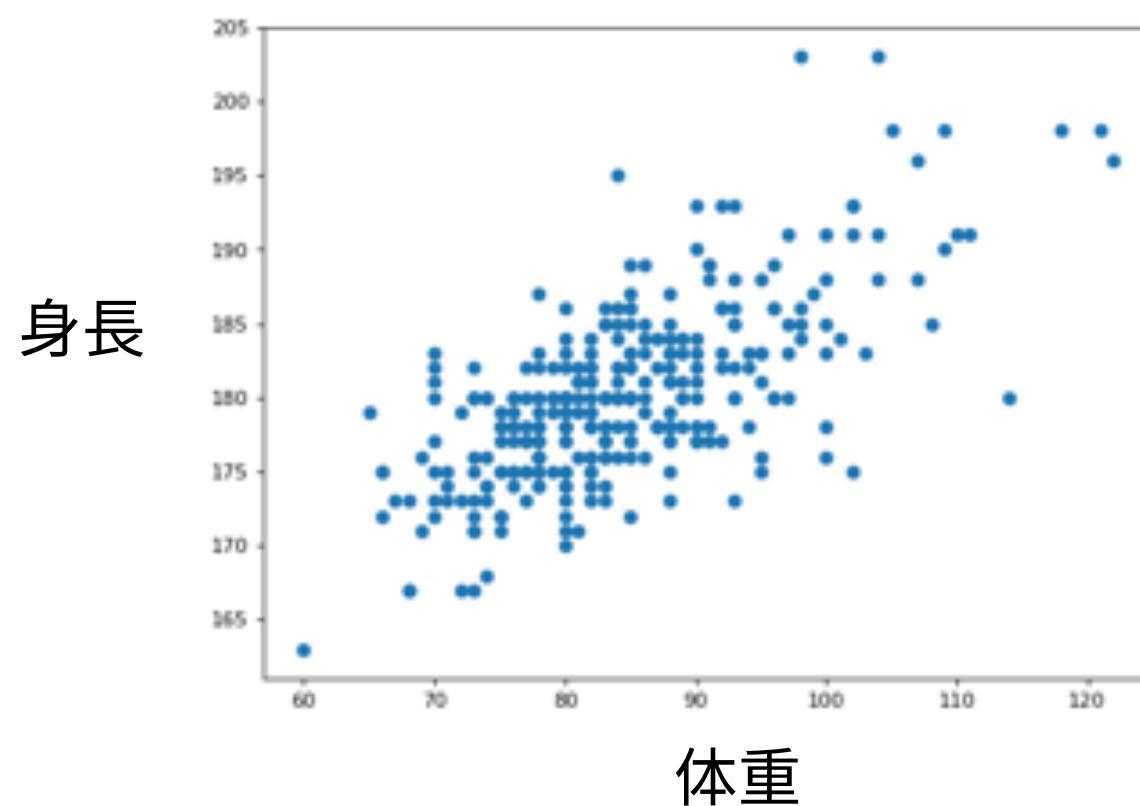
因果の理解には実験研究(介入研究)が必要不可欠

機械学習はあくまでデータの中の多次元相関を捉え、それによって予測する技術

→ 観察された相関が本当に因果性を含むのかを確かめるためには実験するしかない！

日本プロ野球開幕一軍選手の身長・体重データ

(2016年球団公式サイト選手データより自作)



「体重を増やせば身長も伸びる」が正しいかは
この観察データだけからは決して分からぬ

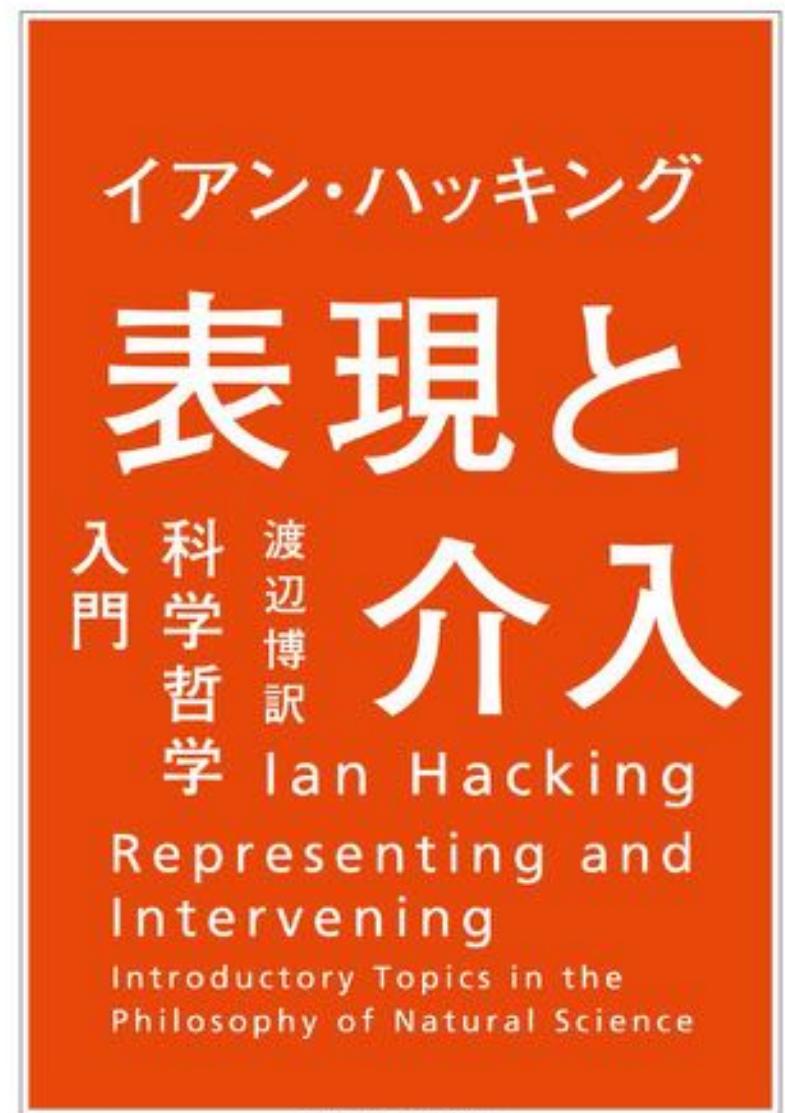
応用統計学の基本のキ

相関関係は必ずしも因果関係を意味しない

→ 「予測ができる」ことは「理解」や「発見」ができる^{ことを直接は意味しない！！}

予測から理解・発見へ：頭でっかちを脱し、現場に出よう！

事件はコンピュータ(機械学習)の中で起きてるんじゃない、**現場**で起きているんだ！ by 僕



教訓：科学研究とは結局人間の営み！

「理解」や「発見」したいのは機械ではなく私たち人間

つまり、自然法則の問題ではなく**私たち自身の精神と世界のあり方の問題**を問うことになる！

教訓：科学研究とは結局人間の営み！

「理解」や「発見」したいのは機械ではなく私たち人間

つまり、自然法則の問題ではなく **私たち自身の精神と世界のあり方の問題** を問うことになる！

- **解釈性**：私たちの**ショボい認知能力**に収まるような「平易な理解」が求められている。

教訓：科学研究とは結局人間の営み！

「理解」や「発見」したいのは機械ではなく私たち人間

つまり、自然法則の問題ではなく **私たち自身の精神と世界のあり方の問題** を問うことになる！

- **解釈性**：私たちの**ショボい認知能力**に収まるような「平易な理解」が求められている。
- **時間性**：無限に時間があれば「学習」は不要。**有限の時間**しか生きられない私たちに「発見」をお膳立てするためのヒント出しが求められている（人類絶滅のタイムリミット内に）。

教訓：科学研究とは結局人間の営み！

「理解」や「発見」したいのは機械ではなく私たち人間

つまり、自然法則の問題ではなく **私たち自身の精神と世界のあり方の問題** を問うことになる！

- **解釈性**：私たちの**ショボい認知能力**に収まるような「平易な理解」が求められている。
- **時間性**：無限に時間があれば「学習」は不要。**有限の時間**しか生きられない私たちに「発見」をお膳立てするためのヒント出しが求められている（人類絶滅のタイムリミット内に）。
- **情報の部分性**：データにできる情報は**いつでも世界の情報量のほんのひとかけら**だけ。ゆく河の流れは絶えずして、しかももとの水にあらず。すべてを観測することはできない。

教訓：科学研究とは結局人間の営み！

「理解」や「発見」したいのは機械ではなく私たち人間

つまり、自然法則の問題ではなく **私たち自身の精神と世界のあり方の問題** を問うことになる！

- **解釈性**：私たちの**ショボい認知能力**に収まるような「平易な理解」が求められている。
- **時間性**：無限に時間があれば「学習」は不要。**有限の時間**しか生きられない私たちに「発見」をお膳立てするためのヒント出しが求められている（人類絶滅のタイムリミット内に）。
- **情報の部分性**：データにできる情報は**いつでも世界の情報量のほんのひとかけら**だけ。ゆく河の流れは絶えずして、しかももとの水にあらず。すべてを観測することはできない。
- **観測バイアス**：人間が一生懸命集めたデータはどうしたって**何らかの偏り**から逃れられない。与えられたデータの傾向を捉える機械学習の予測も同様にその偏りから逃れられない。

教訓：科学研究とは結局人間の営み！

「理解」や「発見」したいのは機械ではなく私たち人間

つまり、自然法則の問題ではなく **私たち自身の精神と世界のあり方の問題** を問うことになる！

- **解釈性**：私たちの**ショボい認知能力**に収まるような「平易な理解」が求められている。
- **時間性**：無限に時間があれば「学習」は不要。**有限の時間**しか生きられない私たちに「発見」をお膳立てするためのヒント出しが求められている（人類絶滅のタイムリミット内に）。
- **情報の部分性**：データにできる情報は**いつでも世界の情報量のほんのひとかけら**だけ。ゆく河の流れは絶えずして、しかももとの水にあらず。すべてを観測することはできない。
- **観測バイアス**：人間が一生懸命集めたデータはどうしたって**何らかの偏り**から逃れられない。与えられたデータの傾向を捉える機械学習の予測も同様にその偏りから逃れられない。
- **因果性の理解**：「**因果性**」は直接観測できない。人間がアクセスできるのは「相関」だけ。

機械学習から機械発見へ

実験自動化の技術的発展：化学でも非効率な労働がいずれ自動化されるのは歴史的必然



Science 363 (2019)



Nature 583 (2020)



Nature Reviews Drug Discovery 17 (2018)



機械学習から機械発見へ

実験自動化の技術的発展：化学でも非効率な労働がいずれ自動化されるのは歴史的必然



Science 363 (2019)



Nature 583 (2020)



Nature Reviews Drug Discovery 17 (2018)



機械学習から機械発見へ

実験自動化の技術的発展：化学でも非効率な労働がいずれ自動化されるのは歴史的必然



Science 363 (2019)



Nature 583 (2020)



Nature Reviews Drug Discovery 17 (2018)



- **機械発見技術の研究基盤として非常に重要**：再現性・属人性などデータの質と量の確保 + Negativeデータを取る実験やランダム実験はデータ科学上は必要だが人間はやりたくない…

機械学習から機械発見へ

実験自動化の技術的発展：化学でも非効率な労働がいずれ自動化されるのは歴史的必然



Science 363 (2019)



Nature 583 (2020)



Nature Reviews Drug Discovery 17 (2018)



- **機械発見技術の研究基盤として非常に重要**：再現性・属人性などデータの質と量の確保 + Negativeデータを取る実験やランダム実験はデータ科学上は必要だが人間はやりたくない…
- 実験自動化が実現されても 「常にひとかけらの部分情報しか手に入らない」 本質は**変わらない**

機械学習から機械発見へ

実験自動化の技術的発展：化学でも非効率な労働がいずれ自動化されるのは歴史的必然



Science 363 (2019)



Nature 583 (2020)



Nature Reviews Drug Discovery 17 (2018)



- **機械発見技術の研究基盤として非常に重要**：再現性・属人性などデータの質と量の確保 + Negativeデータを取る実験やランダム実験はデータ科学上は必要だが人間はやりたくない…
- 実験自動化が実現されても 「常にひとかけらの部分情報しか手に入らない」 本質は**変わらない**
- **発見が自動化できるか**はAI分野にとっても積年の未解決問題。「人工知能」を作りたいなら私たちが日々小さな「発見」と「学習」を繰り返して世界を理解していく過程の理解は不可避

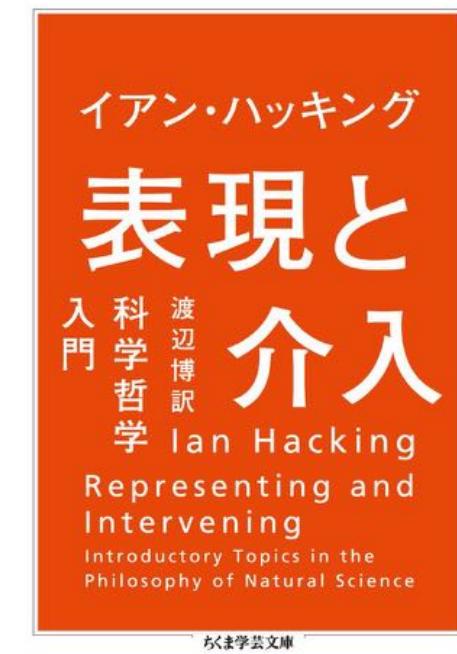
Reality Bites! 道具は違えど、みんな同じ問題と戦っている…

**All models are wrong,
but some are useful (George Box)**

**If data can speak for themselves,
they can also lie for themselves
(David Hand)**

So it's critically important to

- exercise caution
- do not claim too much
- understand the data
- and its quality



~~With enough data, the numbers speak for themselves.~~ (Chris Anderson)

CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

WIRED

経験論(機械学習)と合理論(理論・論理推論・記号処理)の融合

グレイボックス最適化

データ同化

表現学習

モデルベース最適化

モデルベース探索

逐次実験計画

メタ学習

模倣学習

統計的論理合成

生成的モデル化

:

Theory-driven 【合理(先駆)論】

(人工知能分野)

- 対象現象や要求の複雑化
理論化で使う仮定と実現象の解離
- シミュレーション技法の複雑化
恣意性の高いパラメタや近似技法

- 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

Data-driven 【経験論】

(人工知能分野)

- 小サンプル・低カウントの問題
羅生門効果とUnderspecification
- 帰納バイアスのデザイン
- 予測と理解(相関と因果)の溝
- 組合せ的探索空間の確率的探索

- 「発見/創発/創造」や「理解」の論理
- 明示的知識の獲得の論理
- パターン認識(経験と勘)と論理的思考
(論理推論)のギャップ
- モデル適用範囲と信頼性・安全性

余談：言語など過程が元々観測できない対象だと、より顕著？

人工知能学会誌 2012 年 27 卷 3 号 p. 273-283 (オープンアクセス)

https://doi.org/10.11517/jjsai.27.3_273

合理主義と経験主義のはざまで—内的な処理の計算モデル—

273

特 集 「ポスト経験主義の言語処理」

合理主義と経験主義のはざまで —内的な処理の計算モデル—

Rationalism and Empiricism
—Computation Model of Mental Processes—

辻井 潤一
Jun-ichi Tsujii

マイクロソフト研究所アジア
Microsoft Research Asia.
jtsujii@microsoft.com

Keywords: rationalism, empiricism, natural language understanding, natural language processing, computational modeling.

余談：最近の実例1 「AlphaFold v2」

nature

Explore content ▾ About the journal ▾ Publish with us ▾

View all journals

Search

Login

nature > articles > article

Article | Open Access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

John Jumper , Richard Evans, ... Demis Hassabis + Show authors

Nature 596, 583–589 (2021) | Cite this article

489k Accesses | 493 Citations | 2952 Altmetric | Metrics

Download PDF



Associated Content

Highly accurate protein structure prediction for the human proteome

Kathryn Turjansuvunakool, Jonas Adler ... Demis Hassabis

Article | Open Access | 22 Jul 2021

Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort^{1,2,3,4}, the structures of around 100,000 unique proteins have been determined⁵, but this represents a small fraction of the billions of known protein sequences^{6,7}. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single

Protein-structure prediction revolutionized

Mohammed AlQuraishi

News & Views | 23 Aug 2021

Sections

Figures

References

余談：最近の実例2 「機械学習にPure Mathやらせてみた！」

nature

Explore content ▾ About the journal ▾ Publish with us ▾

View all journals

Search 

Login 

Sign up for alerts 

RSS feed

nature > articles > article

Article | Open Access | Published: 01 December 2021

Advancing mathematics by guiding human intuition with AI

Alex Davies , Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomasev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, Marc Lackenby, Geordie Williamson, Demis Hassabis & Pushmeet Kohli 

Nature 600, 70–74 (2021) | [Cite this article](#)

122k Accesses | 1553 Altmetric | [Metrics](#)

Abstract

The practice of mathematics involves discovering patterns and using these to formulate and prove conjectures, resulting in theorems. Since the 1960s, mathematicians have used computers to assist in the discovery of patterns and formulation of conjectures¹, most famously in the Birch and Swinnerton-Dyer conjecture², a Millennium Prize Problem³. Here we provide examples of new fundamental results in pure mathematics that have been discovered with the assistance of machine learning—demonstrating a method by which machine learning can aid mathematicians in discovering new conjectures and theorems. We propose a process of using machine learning to discover potential patterns and relations

Download PDF 

Associated Content

Artificial intelligence aids intuition in mathematical discovery

Christian Stump

News & Views | 01 Dec 2021

Sections

Figures

References

Abstract

Main

[Guiding mathematical intuition with AI](#)

Topology

[Representation theory](#)

Conclusion

Methods

[Code availability](#)

[Data availability](#)

余談：最近の実例3 「DM21汎関数」

Science

Current issue

First release papers

Archive

About

Submit manuscript

GET OUR E-ALERTS

HOME > SCIENCE > VOL. 376, NO. 6573 > PUSHING THE FRONTIERS OF DENSITY FUNCTIONALS BY SOLVING THE FRACTIONAL ELECTRON PROBLEM

REPORT | QUANTUM CHEMISTRY



Pushing the frontiers of density functionals by solving the fractional electron problem

JAMES KIRKPATRICK · BRENDAN MCNICHOLS · DAVID H. P. TURPIN · ALEXANDER L. SAINT · JAMES S. SPENCER · ALEXANDER S. D. G. MATTHEWS

ANNETTE LOISEKA · LOUIS THIERRY MOREL FORTIN · ANDREW J. COHEN

+8 authors

Authors Info & Affiliations

SCIENCE · 9 Dec 2021 · Vol 374, Issue 6573 · pp. 1389–1399 · DOI:10.1126/science.adb6033



GET ACCESS

Improving DFT with deep learning

In the past 30 years, density functional theory (DFT) has emerged as the most widely used electronic structure method to predict the properties of various systems in chemistry, biology, and materials science. Despite a long history of successes, state-of-the-art DFT functionals have crucial limitations. In particular, significant systematic errors are observed for charge densities involving mobile charges and spins. Kirkpatrick *et al.* developed a framework to train a deep neural network on accurate chemical data and fractional electron constraints (see the



CURRENT ISSUE



Pan-cancer single-cell landscape of tumor-infiltrating T cells

BY LIANGTAO ZHENG, SHISHANG QIN, ET AL.