



統計的機械学習の基礎と ガラス材料探索への応用

滋賀大学 教育学部

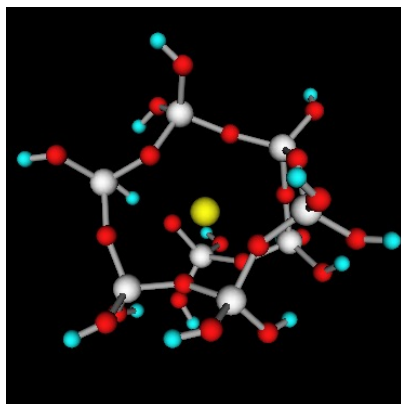
徳田陽明

<https://glass1.net>





・ ガラス材料の高機能化



量子化学計算



固体NMR

・ 震災復興支援

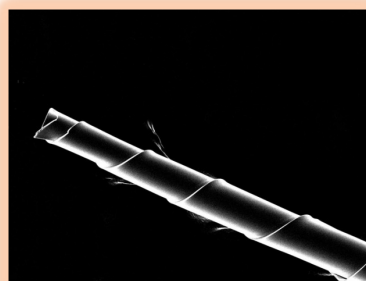


特殊水を用いた除染

・ 新規な材料の創成

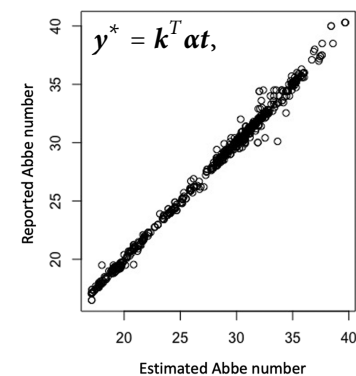


電解質膜



自己組織化

・ 機械学習



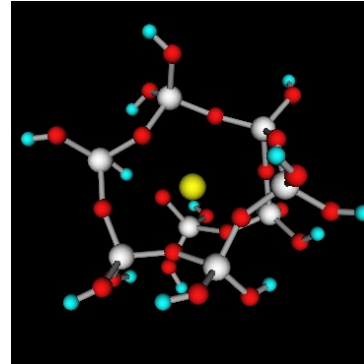
カーネル法



マクロ

イオン伝導性の解明

原子集団の中距離構造の理解



構造解析と電子状態の理解



$$H\psi = E\Psi$$

ミクロ



連載講座「IGOサイエンス」特別座談会 & 全5局解説
新時代到来! ~アルファ碁の驚異の序盤力と弱点に迫る~
 王銘琬九段×マイケル・レドモンド九段×大橋拓文六段

●第40期 棋聖戦七番勝負 第4局 井山裕太 vs 山下敬吾
井山 疾風4連勝 井山裕太棋聖 2大好評連載 漫画 井山裕太のストーリー
自戦解説 井山裕太棋聖 天棋TENKI 第2部 松田一輝

●第17回 農心辛ラーメン杯世界最強戦・第3ラウンド
 ●第54期 十段戦五番勝負 第1・2局 ●第28期 女流名人戦三番勝負 第1・2局

好評連載

「私と囲碁」 <small>2次ノブアテック 株式会社社長</small> 水野雅生	日本棋院90年史 「小林光一、ライバルを破って頂点に」
囲碁大使 戸島花のドリーム囲碁サロン 戸島花	碁会所70年 坂田栄男・藤沢秀行も碁会所で育った
教えてアツシ先生! 人気定石Q&A 講師・加藤充志九段	大矢布石研究所 所長・大矢浩一九段
シンプル is ベスト 石の強弱でわかる次の一手 講師・熊率六段	創作詰碁にチャレンジ 出題・青木紳一九段
エッセイ&パズル ハッピーワールド	トライアル50
	連載 漫画 「井山裕太物語」 大和正樹
	「天棋 -TENKI-」 松田一輝

別冊付録 次の一手 「7路で上達! ⑧~ヨセ福~」 出題・水間俊文七段

あなたの「碁ライフ」が充実する

月刊碁ワールド **李世乭九段** (人工知能) イ・セドル

4月20日発売 定価885円(税込) **5月号 2016 May** Google Play 日本棋院

年間購読(各種会員特典有り)をご希望の方は日本棋院普及事業部まで 〒102-0076 千代田区五番町7-2 TEL.03-3288-8723

世界に衝撃! 最強AI
 グーグルデイトープマインド
 チャレンジマッチ



https://www.nihonkiin.or.jp/publishing/go_world/goworld_201605.html

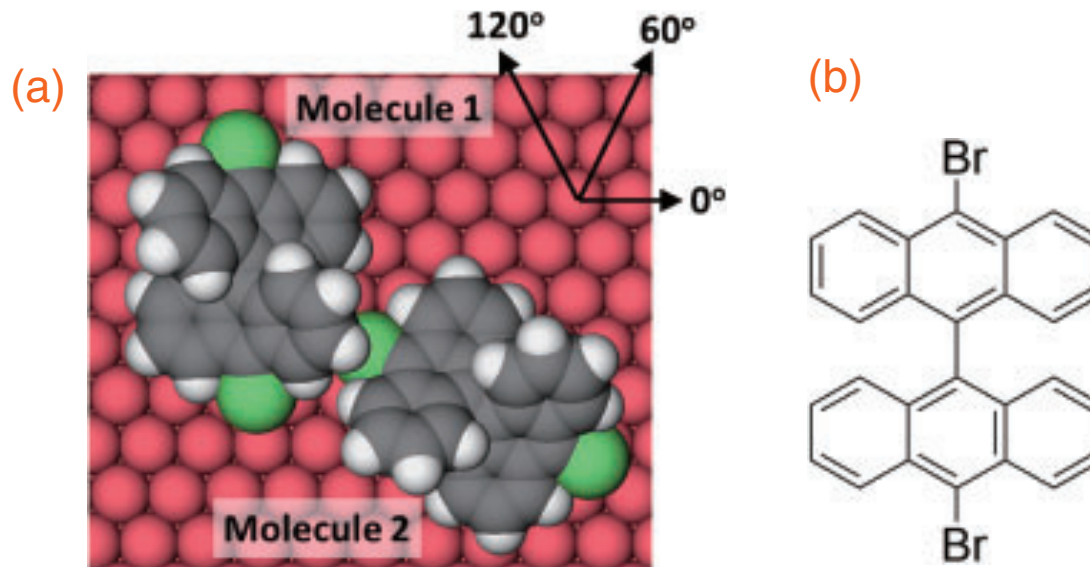
アルファ碁ゼロは、AI同士の対戦で学習し、プロ棋士に勝利

Applied Physics Express **10**, 065502 (2017)

<https://doi.org/10.7567/APEX.10.065502>

Rapid prediction of molecule arrangements on metal surfaces via Bayesian optimization

Daniel M. Packwood^{1,2} and Taro Hitosugi³





- GAFAによる実装
- データ量の増加
 - インターネット利用（画像など）の拡大
 - 実験データの蓄積
- コンピュータの性能の向上
 - ゲーム機が昔のスパコン並みの能力
- 解析アルゴリズムの充実
 - 統計計算，行列計算のパッケージ化

機械学習は魔法の技術？

自動化，自動認識などのニーズに見合うだけの性能を達成可能



マクロ

機械学習による予測

エタノール



データベース化

特徴量の抽出

官能基

バンドギャップ etc



分子の電子状態の理解

$$H\psi = E\Psi$$

ミクロ

高精度な計算が可能



- 凝縮系（固体）
 - 周期性を利用した電子状態の予測（データベース化）
 - 特徴量の抽出
- 凝縮系（液体を含む）
 - 機械学習による原子間ポテンシャルの利用

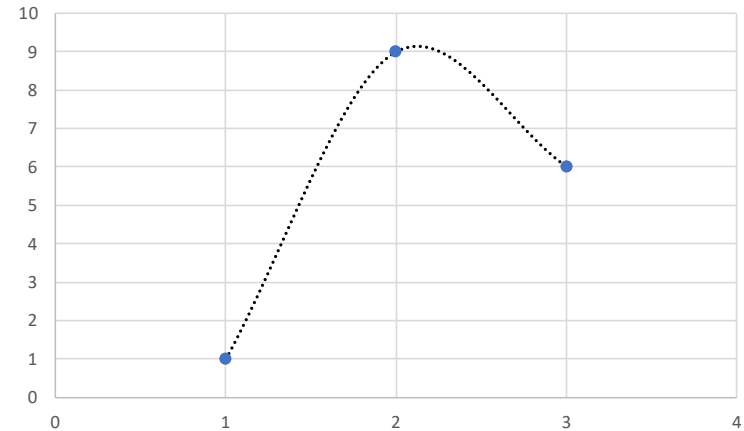
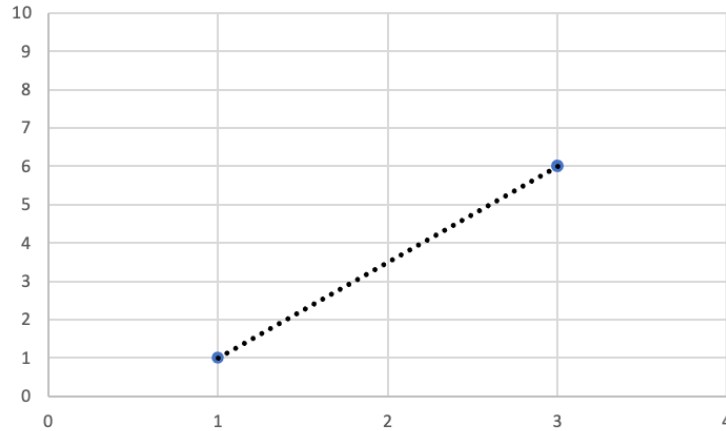
大量のデータを使って、もっと簡単に物性予測できないか

機械学習の基本的な考え方



- 機械（コンピュータ）がデータを用いて自動で学習し、そのデータの背後にあるルールを発見すること。

なんとなく難しそう・・・



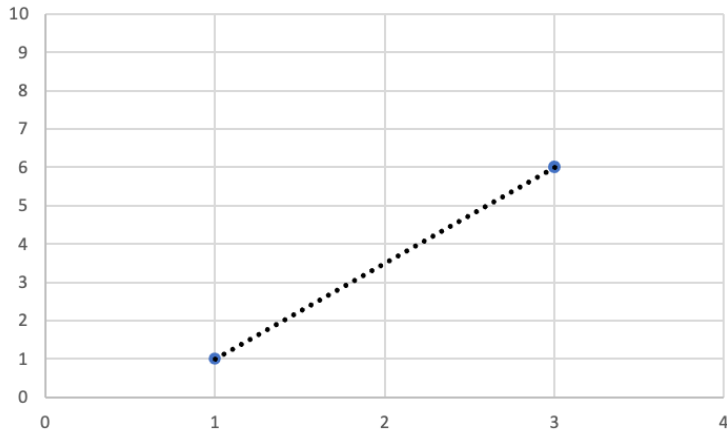
この点線を予想するのが機械学習（の一つ）



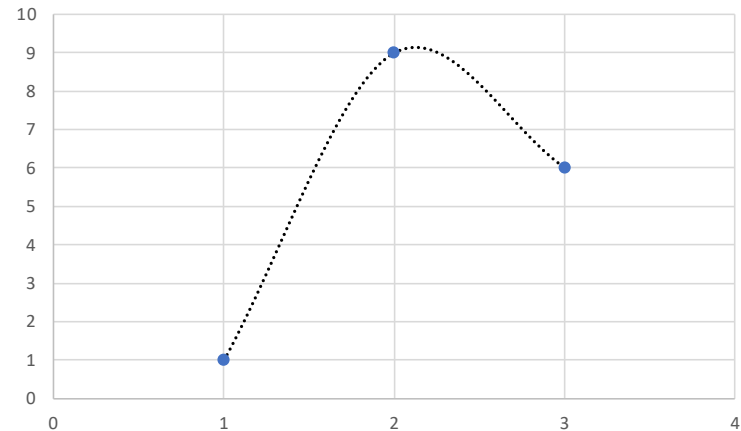
学習：与えられた入力 x と出力 y に合うように関数の形を変えること

$$y = ax^2 + bx + c$$

データが増えたと a, b, c が変わる

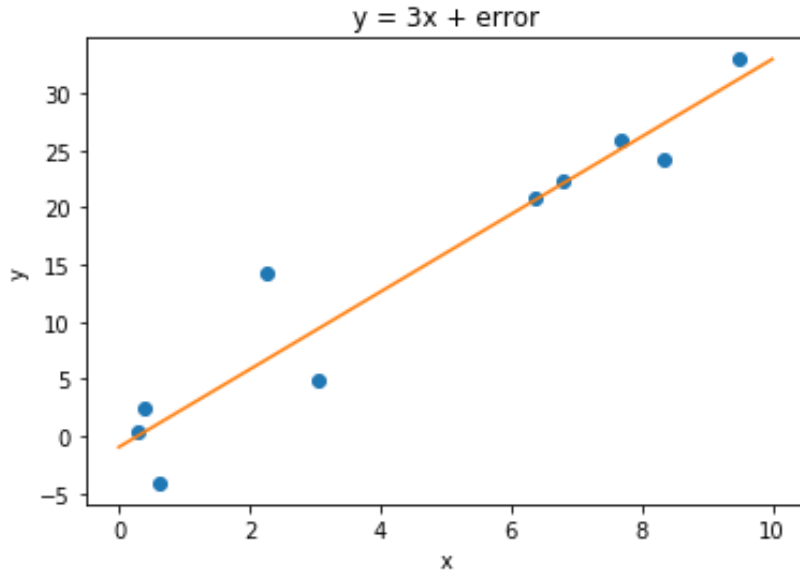


$$y = 0x^2 + 2.5x - 1.5$$

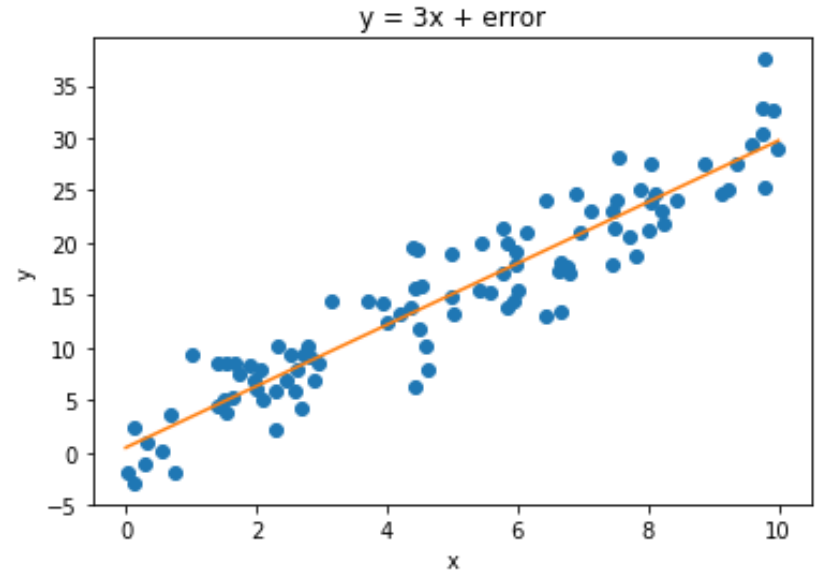


$$y = -6x^2 + 25x - 18$$

データ点が増えると・・・



標準偏差：0.3639



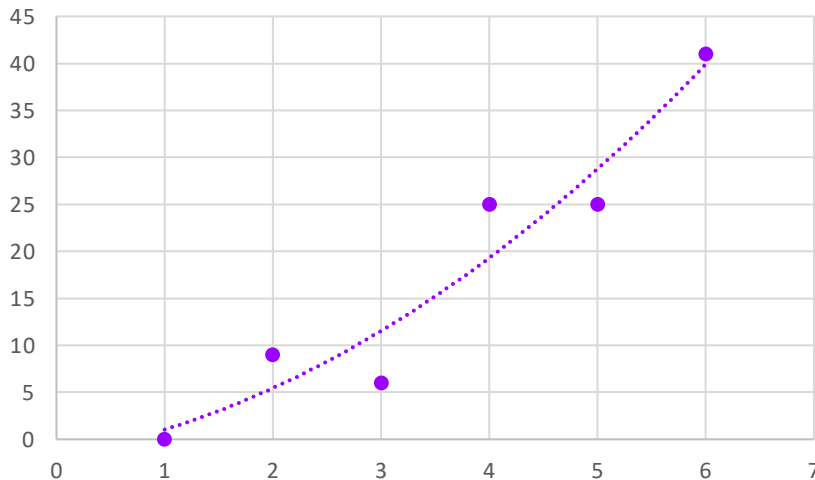
標準偏差：0.1084

データ数の増加 → 予測の精度の向上

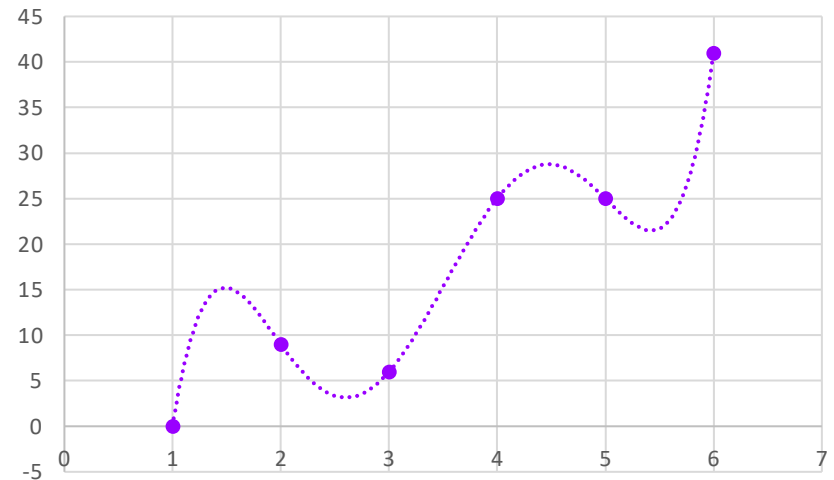


- パラメータの数に依存
- パラメータが多すぎると, 合いすぎる (過学習)

2次で近似



6次で近似



どのようなモデルを用いるか
 どれだけデータを集められるか

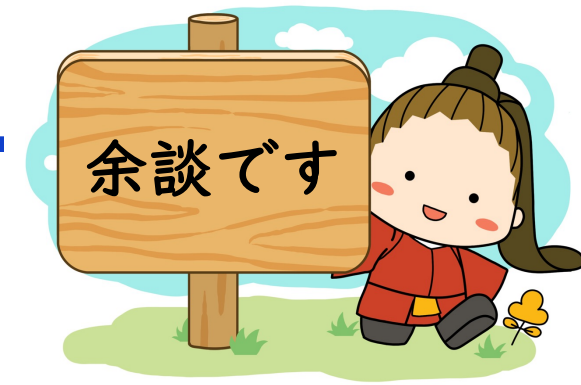
ChatGPTについて



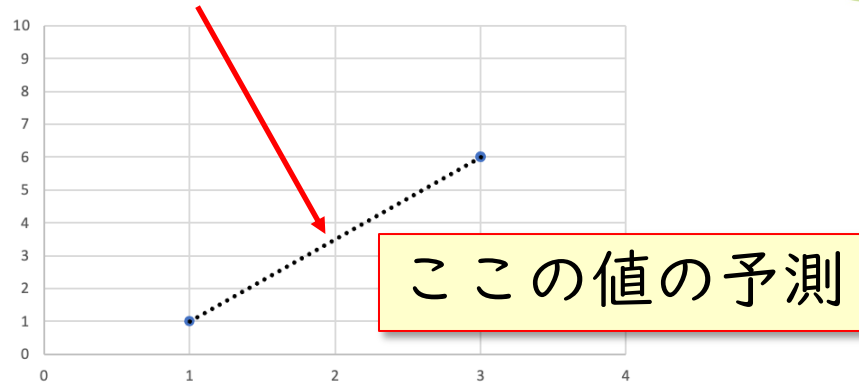
- 現時点で詳細は非公開
- 「非常に多くの」情報(数十Tb)をインターネットで取得し、自動的に学習
- 用いられたパラメータは2000億個
- ある語の次に来る確率の高い語をつなげて表示

例：「富士山」という単語が出ると、次に「は」、その次に「日本」・・・が来ると予測し、最終的に「富士山は日本で一番高い山です」という文章を生成する

ChatGPTについて



- 巨大な空間での「**内挿**」を行なっている。



- 大きなデータが入手できるので、実用的になった
- **データの無いもの, 少ないものには非力**=何でもできるわけではない
 - 固有名詞に関する情報など

* 正解が明確な場合には無から有を作れる。その例が囲碁

- 「**人類**」の作ったものを超えることはありえない (**個人**を超えることは起きている)



- 物性から組成を予測 (逆問題)
- 複数物性を同時に予測 (組成の提案)
- 工場ラインでの異常検知
- 因果関係の推論

物性から組成を予測

逆問題を解きたい



ある性質 y が材料の合成条件 x で決まるとする。

$$y = f(x)$$

例：結晶の組成 x と物性 y など

順問題とは？

x から y を見つける

逆問題とは？

y を示す x を見つける



- 融点 550°C (y)を示す結晶の組成(x)を知りたい
- 超伝導のチャンピオンデータ(y)を示す結晶(x)を見つけたい

これらはまさに材料開発でやっていること
&
簡単にできたら研究で苦勞しない・・・

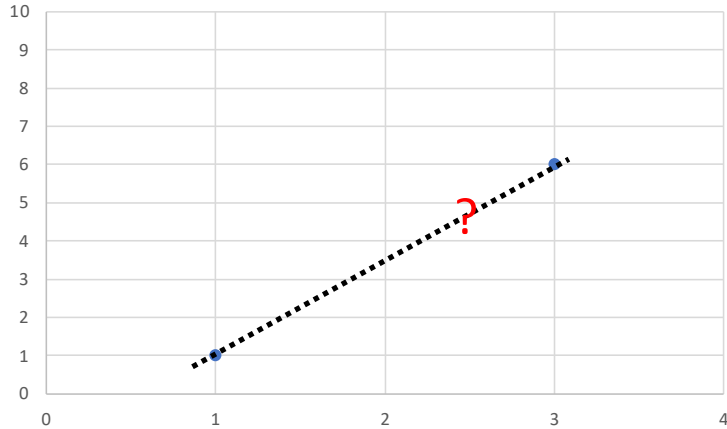


逆問題を直接解くのではなく、既存のデータから探すことはできないだろうか

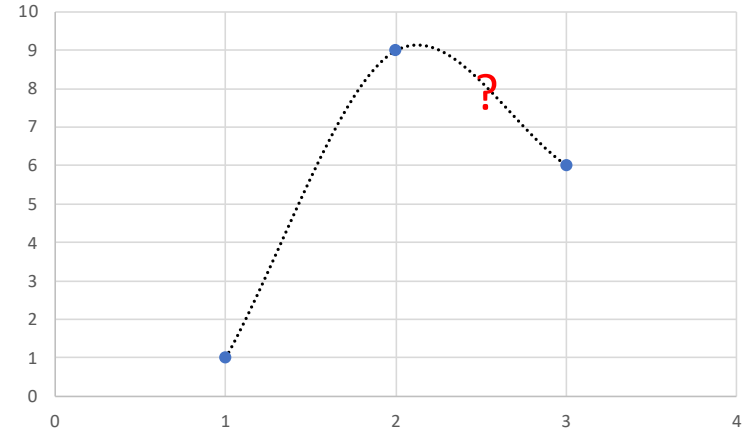
ベイズ最適化



目的: $x = 2.5$ の時の y の値を推定したい

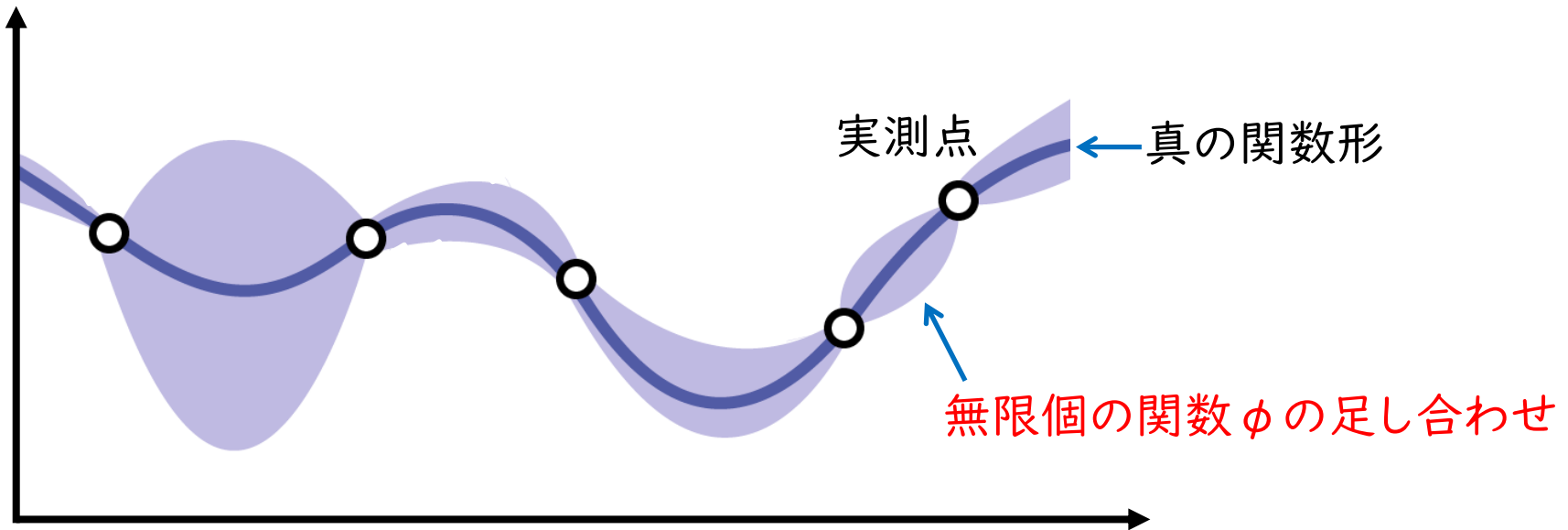


データ増



データ点が増えれば増えるほど, 予測しやすくなる

ガウス過程回帰では点同士の共分散を定義して予測できる



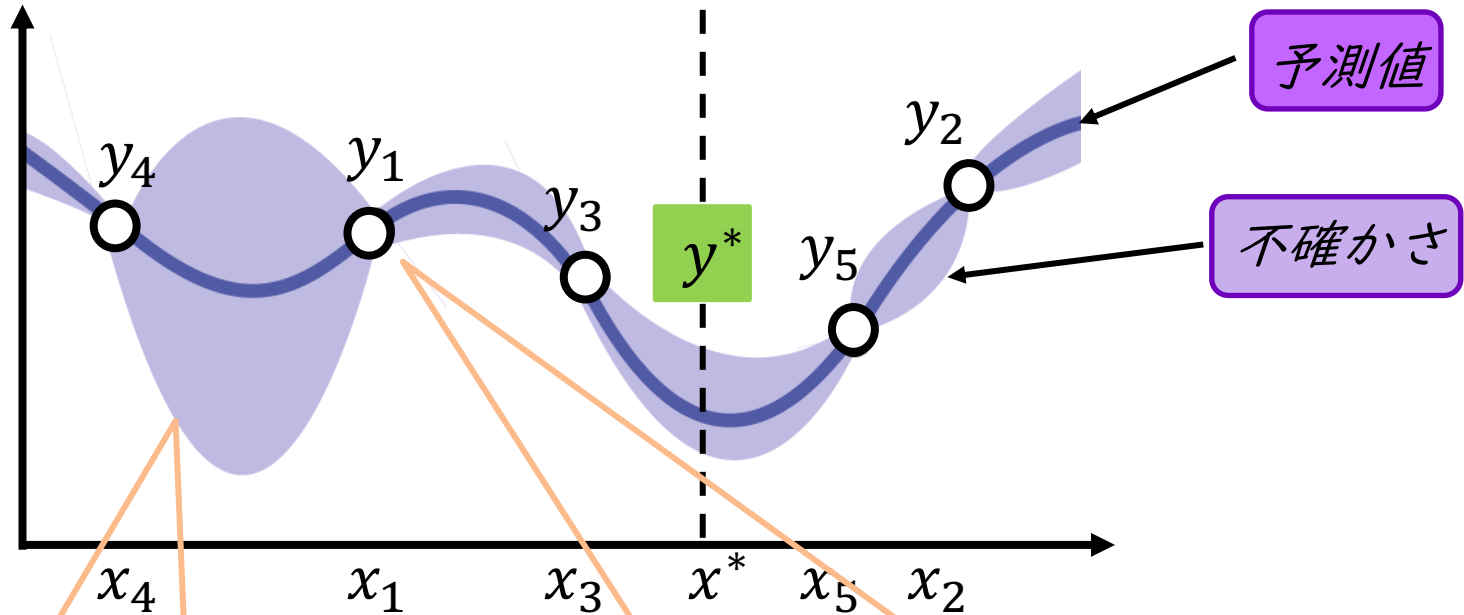
・実測点 \rightarrow 無限個の関数 ϕ の和を推定

特徴

- ・予測の不確かさも一緒に推定可能
- ・過学習が起きづらい



入力から出力を予測する回帰関数の確率モデル



データが十分に存在する場所
→ 自信のある出力 (不確かさが小さい)

データの少ない場所
→ 自信のない出力 (不確かさが大きい)



計算方法（省略）

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \\ y^* \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K & k_* \\ k_*^T & k_{**} \end{pmatrix} \right)$$

x_1, \dots, x_n, x^*

元のデータ同士の類似度

新しいデータと元のデータの類似度

新しいデータと元のデータの類似度

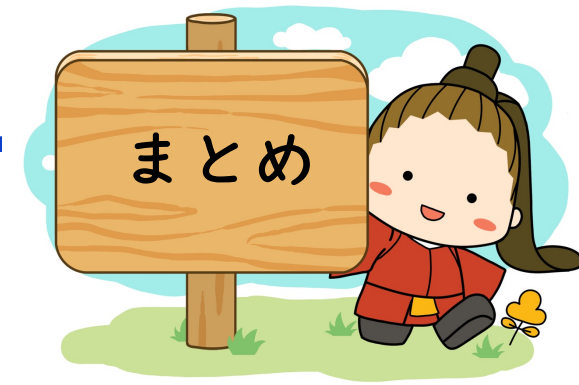
正規分布

・ガウス過程の予

$$p\{y^* | x^*, D\} = (k_*^T K^{-1} y, k_{**} - k_*^T K^{-1} k_*)$$

ベクトルや微分積分を使います
表計算ソフトでも計算できます

難しそうだけど、これだけで予測できるのは凄い！



- データによる予測と実験の組み合わせの繰り返し
 - 探索回数の減少
- 最大値(最小値)をみつける方法
- 材料開発以外でも, 様々な応用が可能

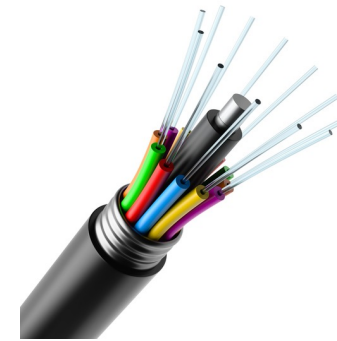
実施例



シェフィールド大聖堂
(イギリス)



江戸切子
Sunrising / PIXTA



光ファイバ

Make / PIXTA



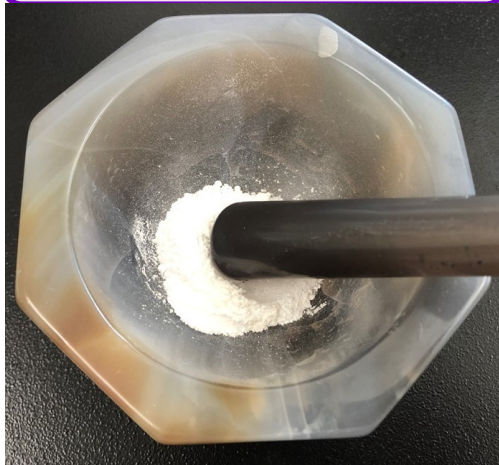
アンプル

panoramaimages / PIXTA

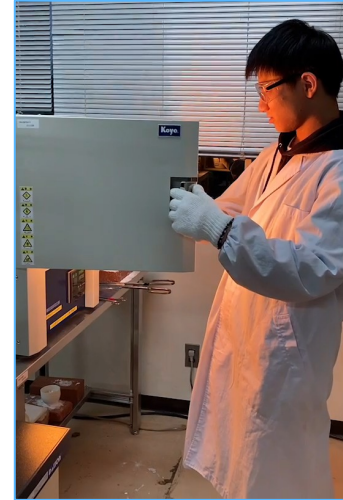
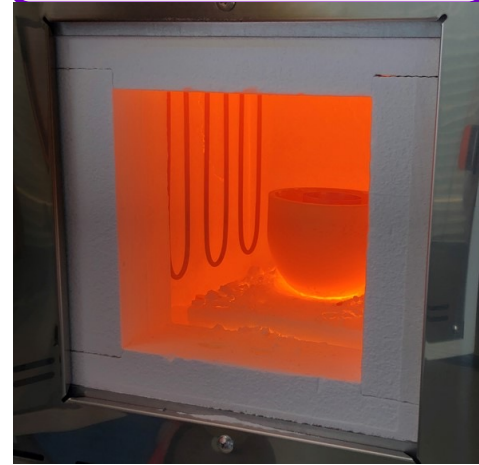
特徴

- ・組成が連続的
- ・多くの元素を取り込む溶媒

秤量



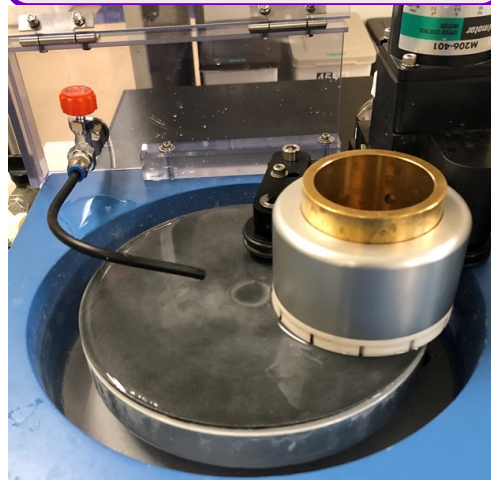
熔融



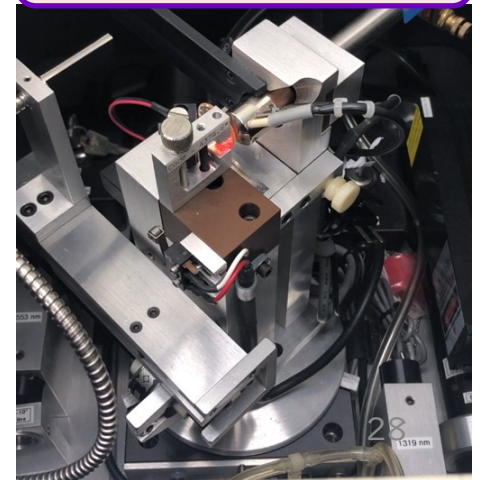
形成



研磨



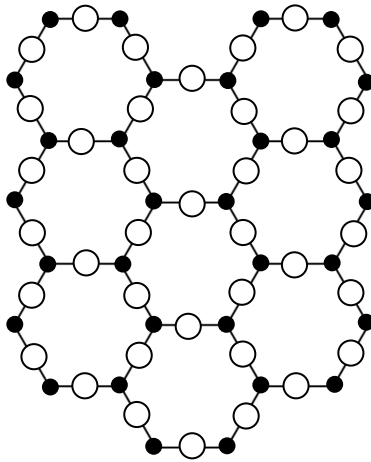
測定





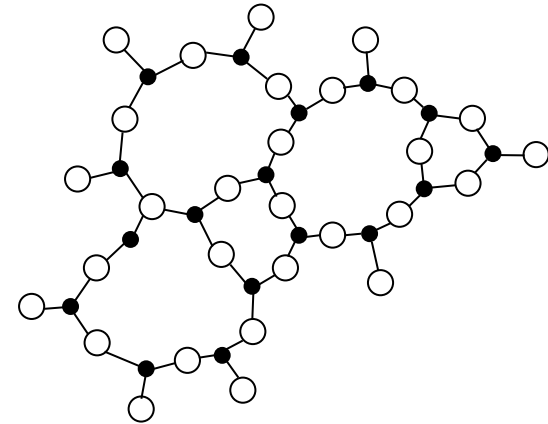
- 結晶

- 規則正しい
- 固体



- ガラス

- ばらばら
- 固体 (液体に似ている)





ガラスに関する物性の例

熱膨張係数
耐熱性
微細加工時に影響

ガラス転移温度
融点の目安
低いと加工が容易

軟化温度
低いと加工が容易

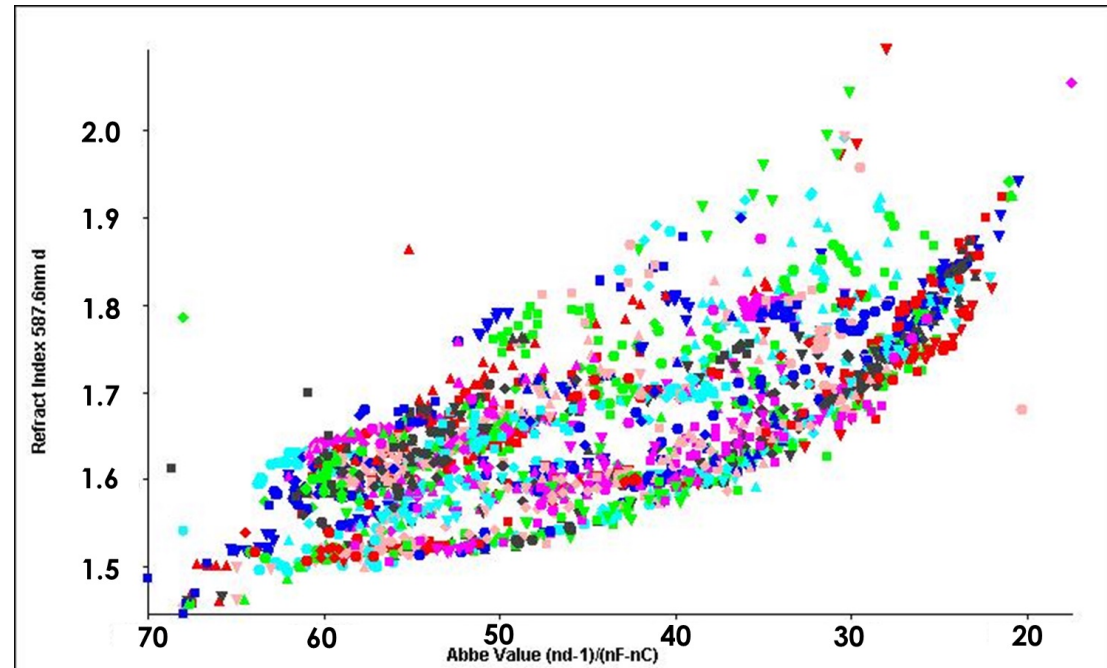
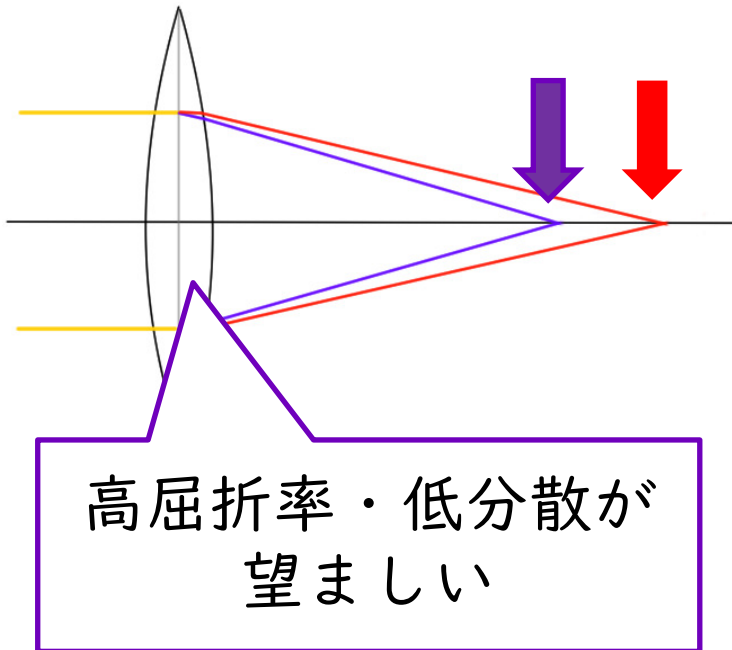
屈折率
反射・屈折に影響

ガラスの物性値を予測できれば、
用途に合わせたガラス作製が可能





光学ガラス 屈折率や分散は組成に依存

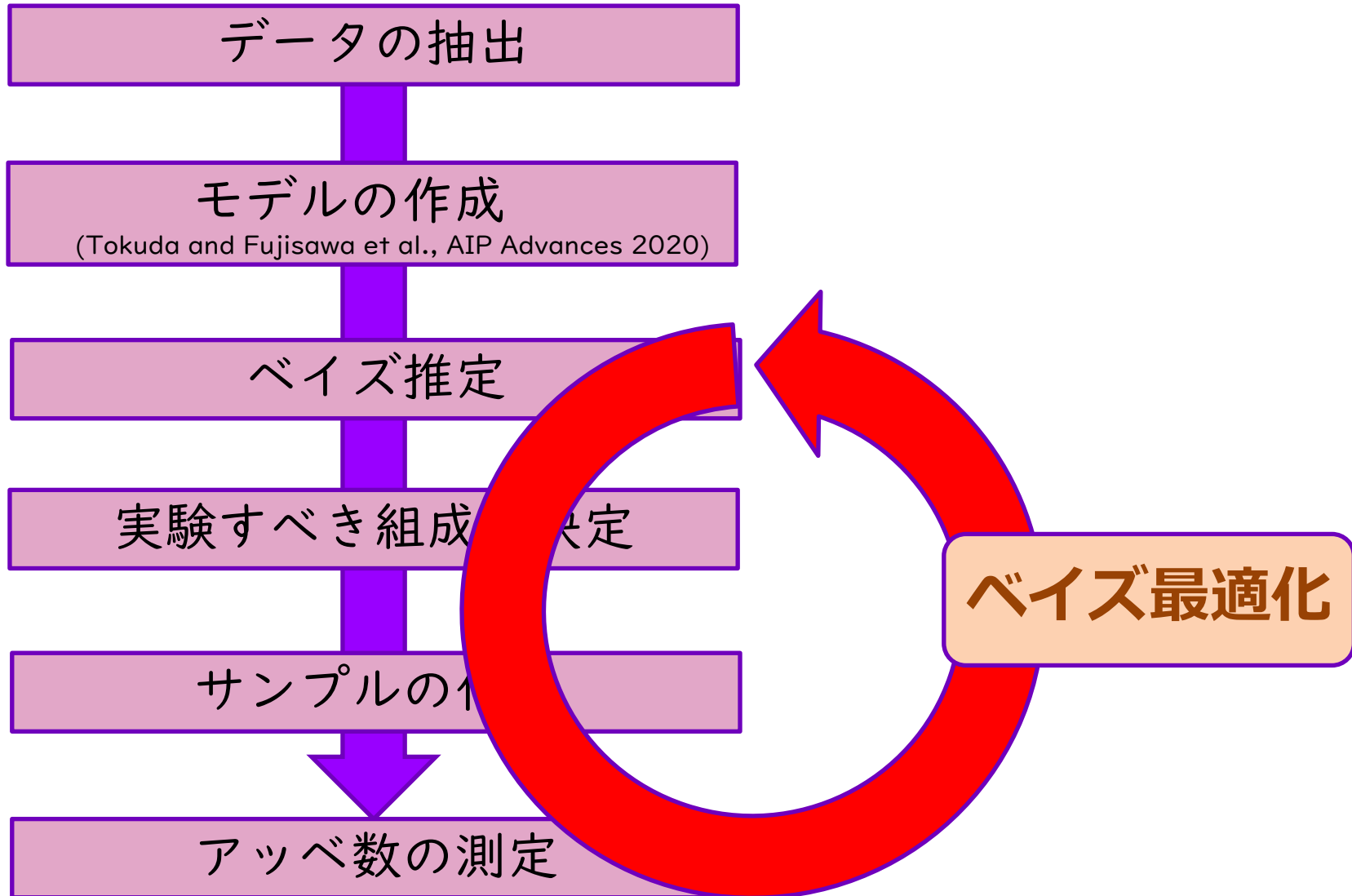


➡ 組成とアッベ数の関係の予測モデル



- 国内外の論文, 特許全てのガラス情報を掲載
- 38万種ガラス, 100万個の物性データ

ガラス番号	SiO ₂	B ₂ O ₃	Al ₂ O ₃	...	MoO ₂	屈折率	アッベ数
1	2.8	3.5	0...		0	1.9967	26.2
2	4.86	19.36	5.6...		0	1.9995	28.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
878	0.5	4.6	0.2...		0	1.9567	29.9
879	20.1	5.6	2.4...		0	1.9435	27.6



実験すべき組成(1回目)



- ・ベイズ最適化をRで実装
- ・EIの値を基に次に探索すべき実験条件(Pbを含まない)を決定

成分	物質質量比 (mol%)	成分	物質質量比 (mol%)
SiO ₂	17.77	ZnO	4.4
B ₂ O ₃	21.39	SrO	7.99
CaO	9.08	Bi ₂ O ₃	4.23
BaO	7.3	TeO ₂	1.05
Li ₂ O	26.79	La ₂ O ₃	5

屈折率
1.7315

アッベ数
36.7





成分	物質質量比 (mol%)	成分	物質質量比 (mol%)
SiO ₂	15.77	ZnO	3.91
B ₂ O ₃	18.98	SrO	7.87
CaO	13.44	Bi ₂ O ₃	3.75
BaO	5.39	TeO ₂	0.93
Li ₂ O	29.96	Na ₂ O	5

屈折率
1.6824

アッベ数
38.3



新たな探索点の決定(3回目)



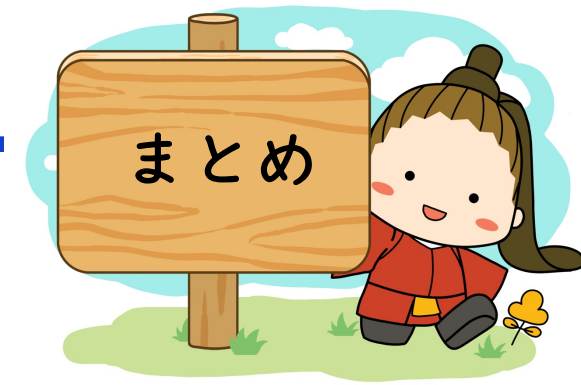
成分	物質質量比 (mol%)	成分	物質質量比 (mol%)
La_2O_3	26.4	Al_2O_3	0.1
Nb_2O_5	2.43	SiO_2	10.87
Ta_2O_5	11.07	Sb_2O_3	0.10
B_2O_3	25.83	Gd_2O_3	11.36
ZrO_2	6.99	TiO_2	4.8

屈折率
1.95

アッベ数
30



探索の結果



最初のガラス



2番目のガラス



3番目のガラス



機械学習で新しいガラスを見つけることができた

- ・この方法は一般的な方法なので、様々な対象に使える
- ・誰でも（“簡単に”とは言わないが）挑戦できる

複数物性の同時予測

組成も提案したい



予算は？

どんな
ガラス？

何色
になる？

何に使う？

熱膨張係数は？

加工
しやすい？

何度で
柔らかく
なる？

屈折率は？



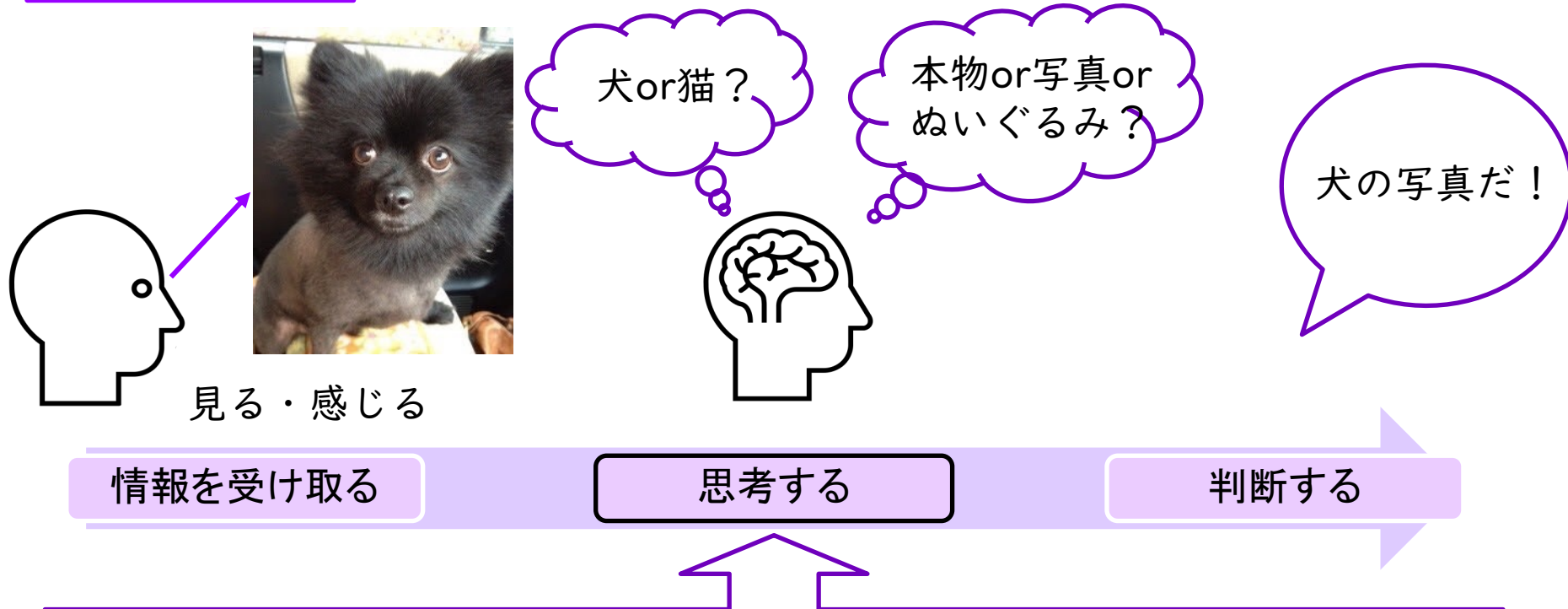
複数の制約条件を満たすものを作りたい！

複数の物性を同時に予測



- 脳機能の特性のいくつかをコンピュータ上で表現するために作られた数学モデル

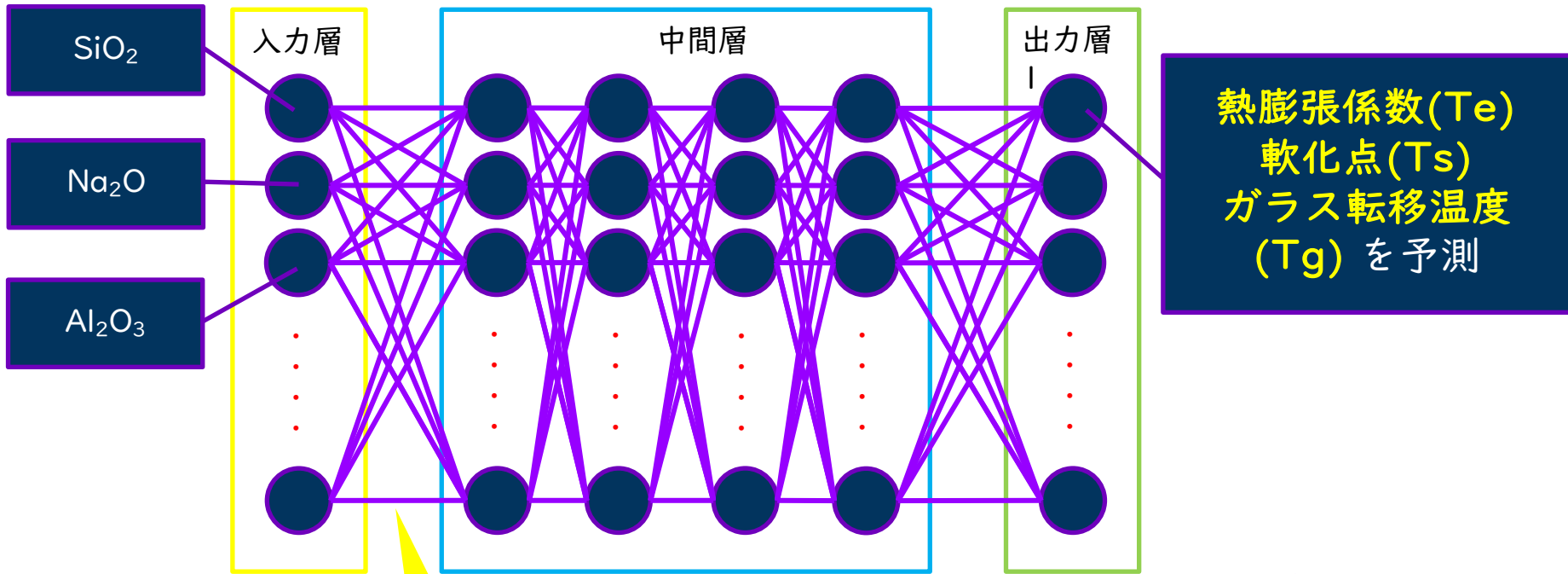
思考のプロセス



情報（データ）量が膨大なので、「思考」をコンピューターで処理
 =ディープラーニング
 ディープラーニングで、2つ以上の物性値を予測することが可能



- 脳機能の特性のいくつかをコンピュータ上で表現するために作られた数学モデル



情報を受け取る

思考する

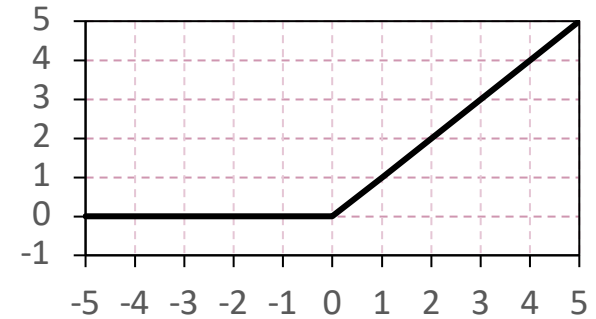
判断する

ニューロン間で関数を使い、必要な情報・不要な情報かを判断している



- ReLU (ランプ関数) は入力が0以上の場合にはそのまま出力され、入力が0以下の場合には0が出力される関数

$$f(u) = \max(0, u) \begin{cases} u & (u \geq 0) \\ 0 & (u < 0) \end{cases}$$



ReLUの特徴

予測可能

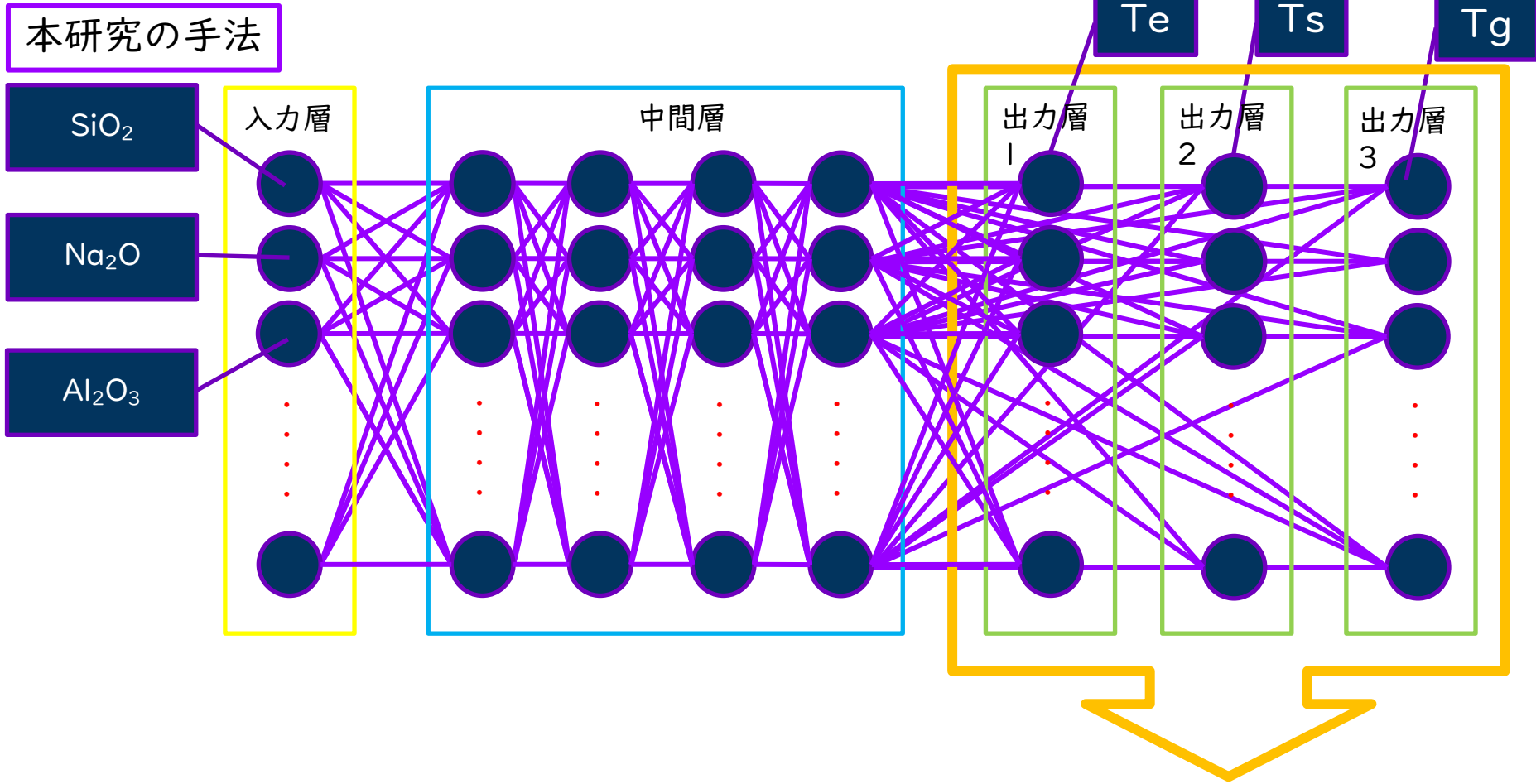
$\begin{cases} u \text{が} 0 \text{以上} : \text{必要} \\ u \text{が} 0 \text{以下} : \text{不要} \end{cases}$

高い学習効率

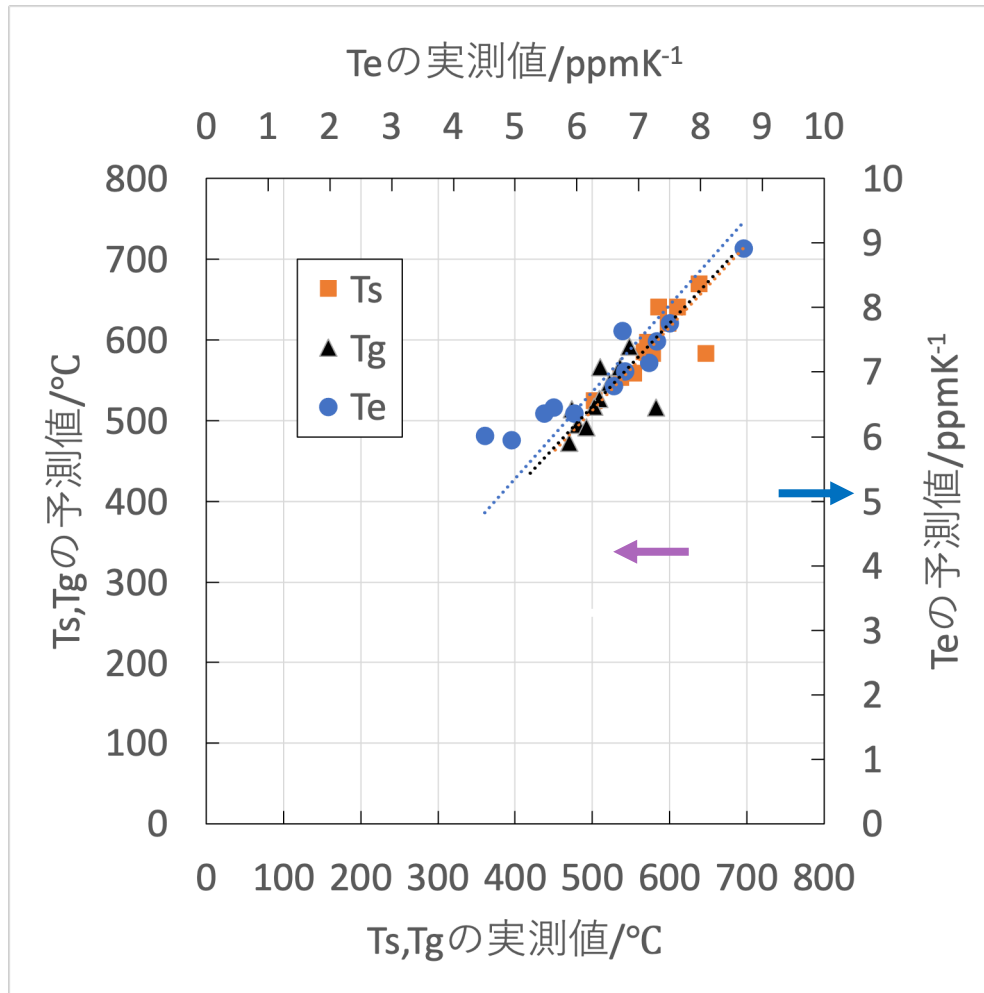
計算式がシンプルなので、処理が速く、精度が向上しやすい



- 脳機能の特性のいくつかをコンピュータ上で表現するために作られた数学モデル



ガラスの熱膨張係数(Te), 軟化点(Ts), ガラス転移温度(Tg)の2つ以上の物性値を同時に予測する



Te:熱膨張係数
 Ts:軟化点
 Tg:ガラス転移温度

- 中間層が(100, 100, 100, 100, 100, 100)かつ $\alpha = 0.0001$ の時に, 最適化した($R^2 = 0.83$)



実証結果

- ▶ Bi_2O_3 を5%添加して作製した試料を測定したところ、 T_e は同程度の値を示し、 T_s は10度程下がり、 T_g は40度程下がった

表1 試料1とBi添加した試料の実測値

物性値	試料1	Bi_2O_3 添加
$T_e(\text{ppm/K})$	4.94	5.62
$T_s(^{\circ}\text{C})$	586	578
$T_g(^{\circ}\text{C})$	510	473

予測した通り T_e の値が同程度で T_s と T_g の値が小さい

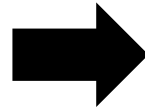


- データ活用は強力な手法
- データ活用して材料開発を加速できる時代
 - 大量のデータを入手可能(データの質が重要)
 - パソコンの性能向上
 - パッケージ化された開発ツール(少々の勉強は必要)
- データ活用が全てではないことに注意
 - 共存の時代



原理原則に基づく理解

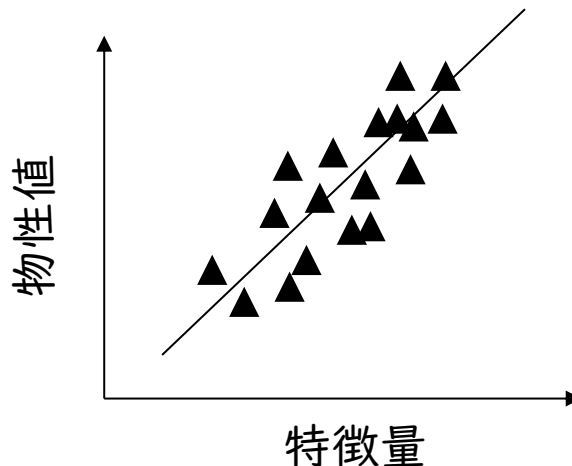
$$\left\{ -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(r) \right\} \Psi(r) = E\Psi(r)$$



ガラスの特性の理解



経験 (データ) に基づく理解



この関係がわかれば十分
(ということもある)

データ駆動型のアプローチ