

スパースモデリングと エンジニアリングデータアナリティクス

東京大学 新領域創成科学研究科 複雑理工学専攻
物質・材料研究機構 統合型材料開発・情報基盤部門
岡田真人

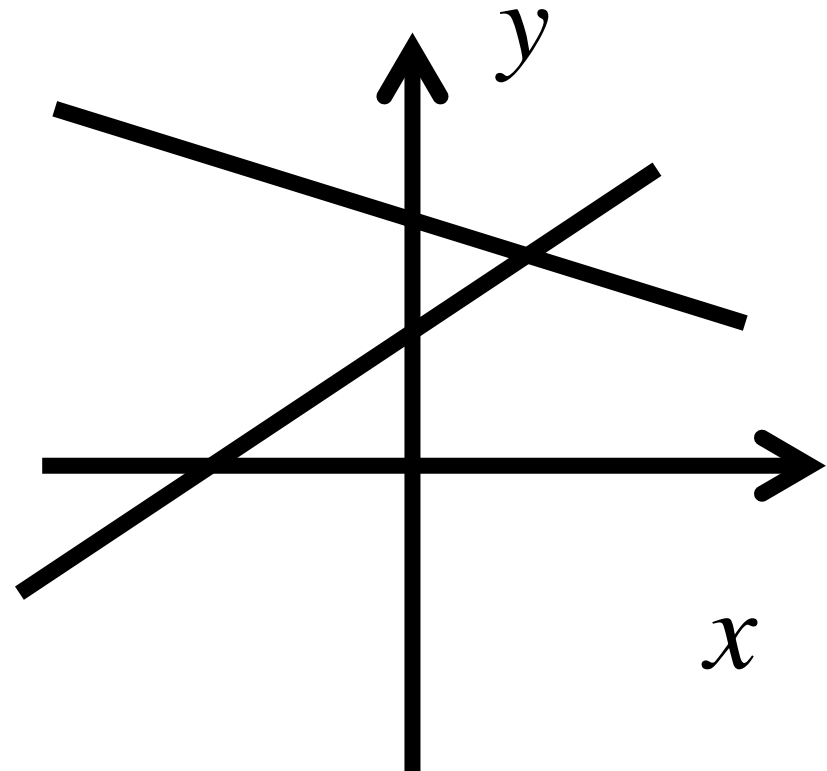
データ駆動科学の肝は 連立方程式と一次関数

連立方程式とその応用

鶴亀算 食塩水 寝坊
して追いかける問題

連立方程式への変換

加減法, 代入法



一つの方程式が一本の線
二本の線の交点が解になる

(五十嵐, 竹中, 永田, 岡田, 応用統計学, 2016)

自己紹介(理論物理学者)

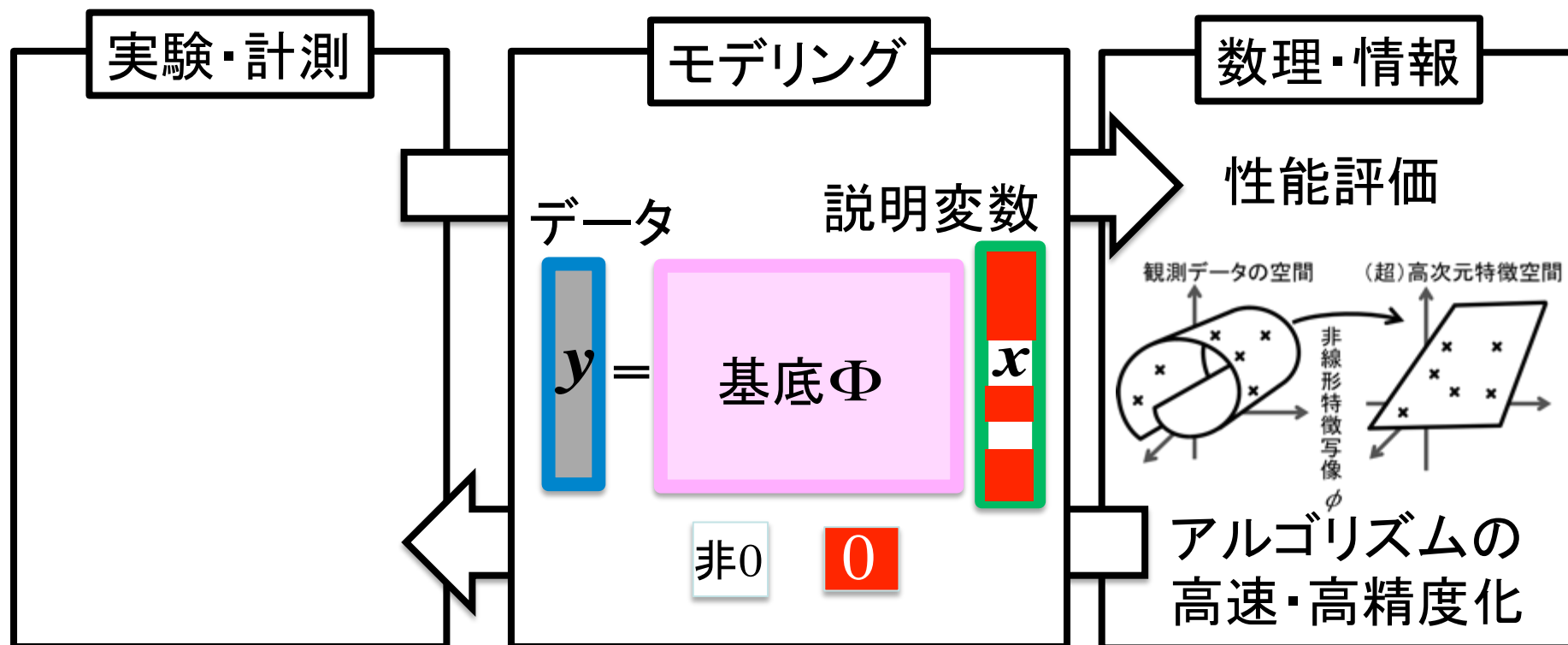
- 大阪市立大学理学部物理学科 (1981 - 1985)
 - アモルファスシリコンの成長と構造解析
- 大阪大学大学院理学研究科(金森研) (1985 - 1987)
 - 希土類元素の光励起スペクトルの理論
- 三菱電機 (1987 - 1989)
 - 化合物半導体(半導体レーザー)の結晶成長
- 大阪大学大学院基礎工学研究科生物工学 (1989 - 1996)
 - 理論脳科学, 計算論的神経科学
 - 情報統計力学(ベイズ推論と統計力学の数理的等価性)
- JST ERATO 川人学習動態脳プロジェクト (1996 - 2001)
 - 計算論的神経科学
- 理化学研究所 脳科学総合研究センター 甘利チーム (2001 - 04/06)
 - ベイズ推論, 機械学習, データ駆動型科学
- 東京大学・大学院新領域創成科学研究科 複雑理工学専攻
 - 物質科学再開(強相関, 表面, 地球惑星科学) (2004/07 -)

新学術領域研究 平成25～29年度 スパースモデリングの深化と高次元データ駆動科学の創成

領域代表 東京大学 岡田真人

個人的な思い

世界を系統的に記述したい
その方法論と枠組みを創りたい



MATLAB EXPO 2017 Japan

グランドニッコー東京 台場 - 10月31日


最新の技術動向を知る

第一線で活躍する
ユーザーから学ぶ

MATLAB/Simulinkのエキスパートと
交流する

注目トラック

幅広い業界での事例や、MATLAB®/Simulink® の最新情報をご紹介します。



エンジニアリングデータアナリ
ティクス
» 講演内容を見る



ディープラーニング
» 講演内容を見る



ADAS/自動運転
» 講演内容を見る

<http://www.matlabexpo.com/jp/>

MONOist IoT Forum 東京 講演レポート：

IoT時代のアナリティクスはエンジニアリングデータが主役になる

企業の関心はエンジニアリングデータ活用へ

- 生産装置の**センサーのデータ**や、**画像、動画**
- 顧客のプロファイルや、売上データ、在庫データなどの**ビジネスデータとは異なる**



MathWorks Japan インダストリーマーケティング部
IA&Mインダストリーマーケティングマネジャーの遠山
巧氏

(Mathworks Japan, 遠山)

- 今後はセンサーや画像など、エンジニアリングデータも使っていきたいと考えている企業が多い

<http://monoist.atmarkit.co.jp/mn/articles/1701/12/news006.html>



IoTシステムを構築するために必要な5条件--エッジとサーバを全体最適化するために

大谷卓也 竹本佳充(MathWorks Japan)

IoTシステムを構築するために必要な5箇条

1. 精度の高い特徴量の抽出
2. 高度なアナリティクスの実装
3. コンピューティングリソースの全体最適化
4. エッジデバイスへのアルゴリズムの実装
5. アグリゲータでのコンピューティング環境

<https://japan.zdnet.com/article/35076390/>



IoTは新たな収益源を生み出すか-- ベンダー座談会 (1)

山田竜司 (編集部) 吉澤亨史 怒賀新也 (編集部) 2016/02/09 07:00

- **センサーデータ**を収集した際にどう上手くデータを前処理するか
- データサイエンティスト人材を見つけるのも大変ですが
- **関連分野のエキスパート**が、機械学習や最適化を簡単に使えるような製品をサポートし、業界全体を底上げする



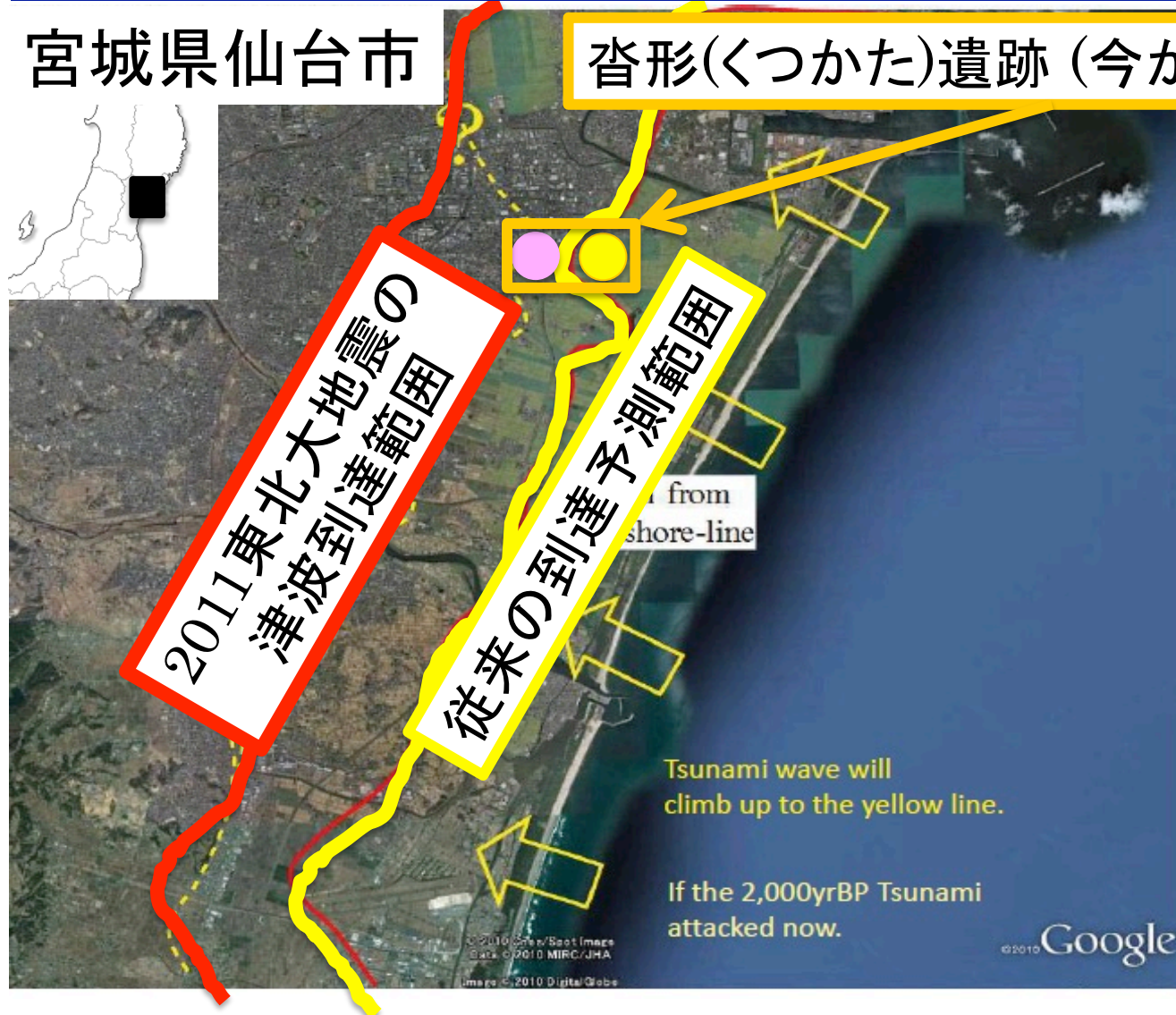
(Mathworks Japan, 阿部)

<https://japan.techrepublic.com/article/35075285.htm>

2011 3.11 津波到達範囲

宮城県仙台市

沓形(くつかた)遺跡 (今から2000年前)



2011 3.11 津波到達範囲

宮城県仙台市

沓形(くつかた)遺跡 (今から2000年前)



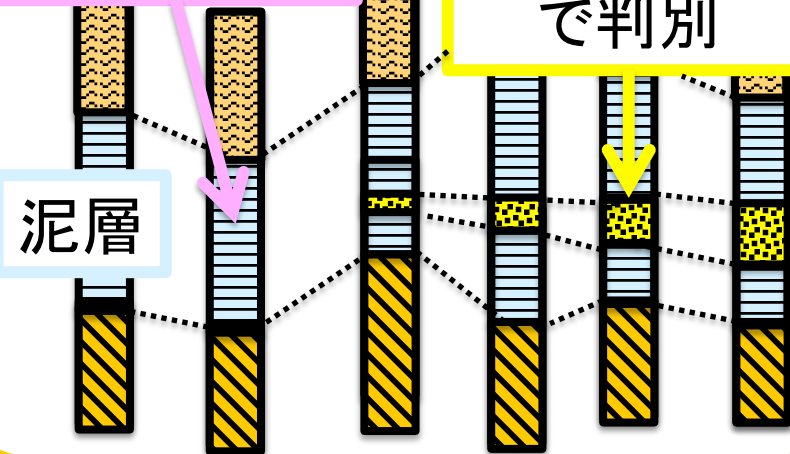
2011東北大地震の
津波到達範囲

従来の到達予測範囲

従来手法: 堆積学的な判別

津波到達?

砂層の存在
で判別



Tsunami wave will
climb up to the yellow line.

If the 2,000yrBP Tsunami
attacked now.

©2010 Google

2011 3.11 津波到達範囲

宮城県仙台市

沓形(くつかた)遺跡 (今から2000年前)



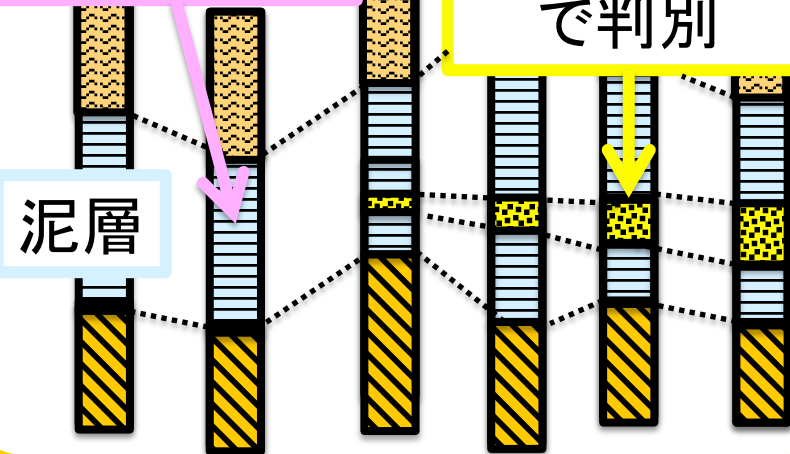
2011東北大地震の
津波到達範囲

従来の到達予測範囲

従来手法: 堆積学的な判別

津波到達?

砂層の存在
で判別



泥層

津波

泥

砂の堆積

従来手法では正確な到達範囲の予測が難しい。

2011 3.11 津波到達範囲

宮城県仙台市

沓形(くつかた)遺跡 (今から2000年前)

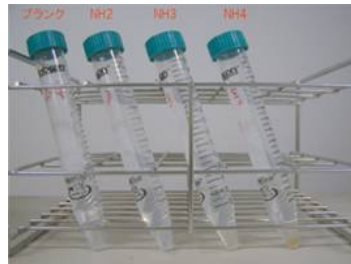


2011東北大地震の
津波到達範囲

従来の到達予測

稠密な試料収集

先端化学分析



膨大な高次元
元素データ

Si	Si	Si
Al	Al	Al
Mg	Mg	Mg
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
Ni	Ni	Ni
Pb	Pb	Pb

元素データから津波堆積物を識別できるか？

If the 2,000yrBP Tsunami

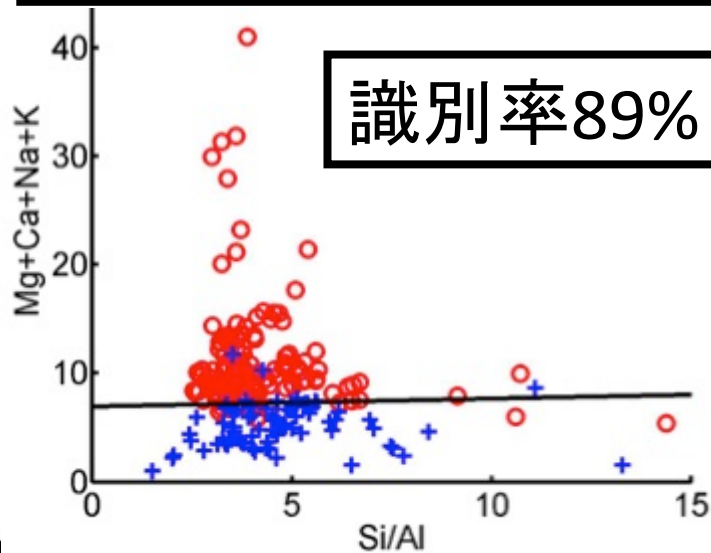
地球科学データによる津波到達範囲推定

2011年東北大地震の津波堆積物
18次元地球化学データ



既知の津波堆積物で判別精度をチェック！

元素を決めうちで判別



全元素を用いた判別

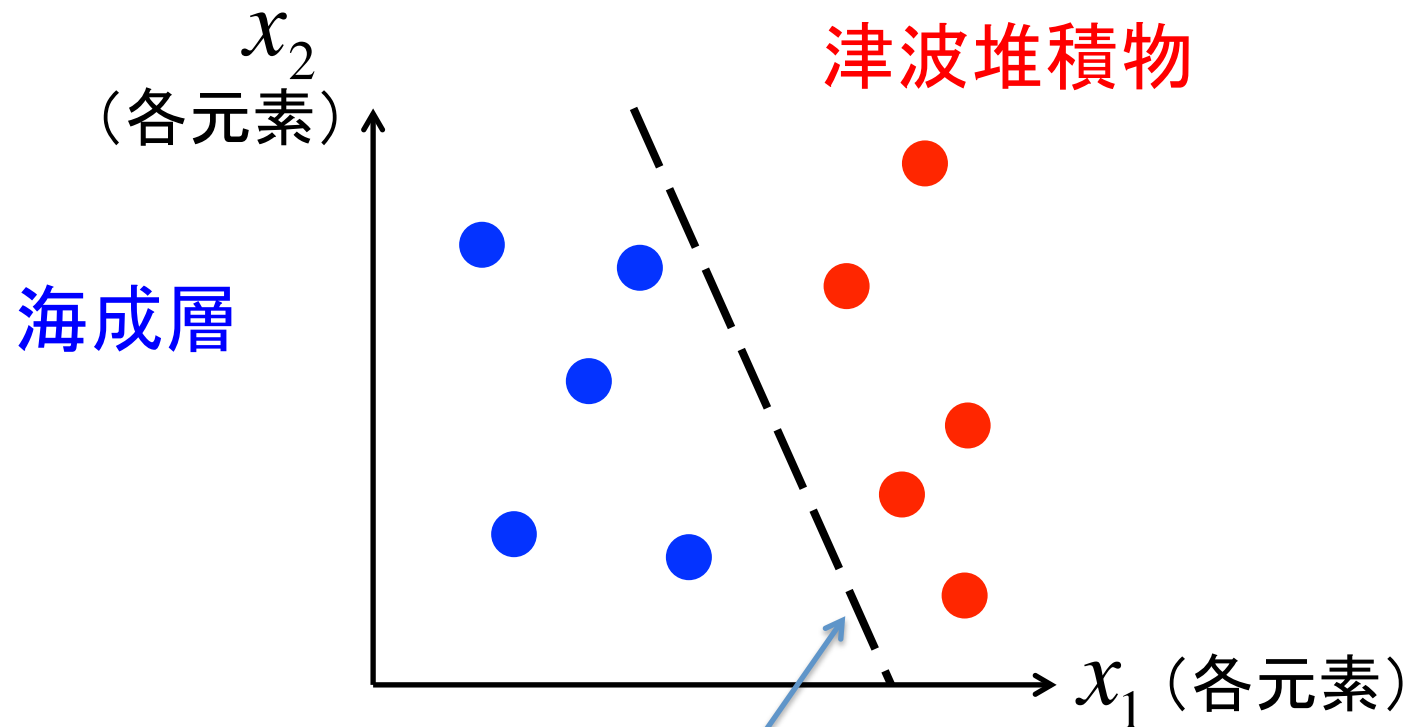
性能があがるはず??

Si	Al	Fe	Mg	Ca	Na
K	Ti	Mn	V	Cr	Ni
Sb	Cu	Zn	As	Cd	Pb

全元素を用いることで高精度に判別??

スパースモデリング入門

線形判別モデル

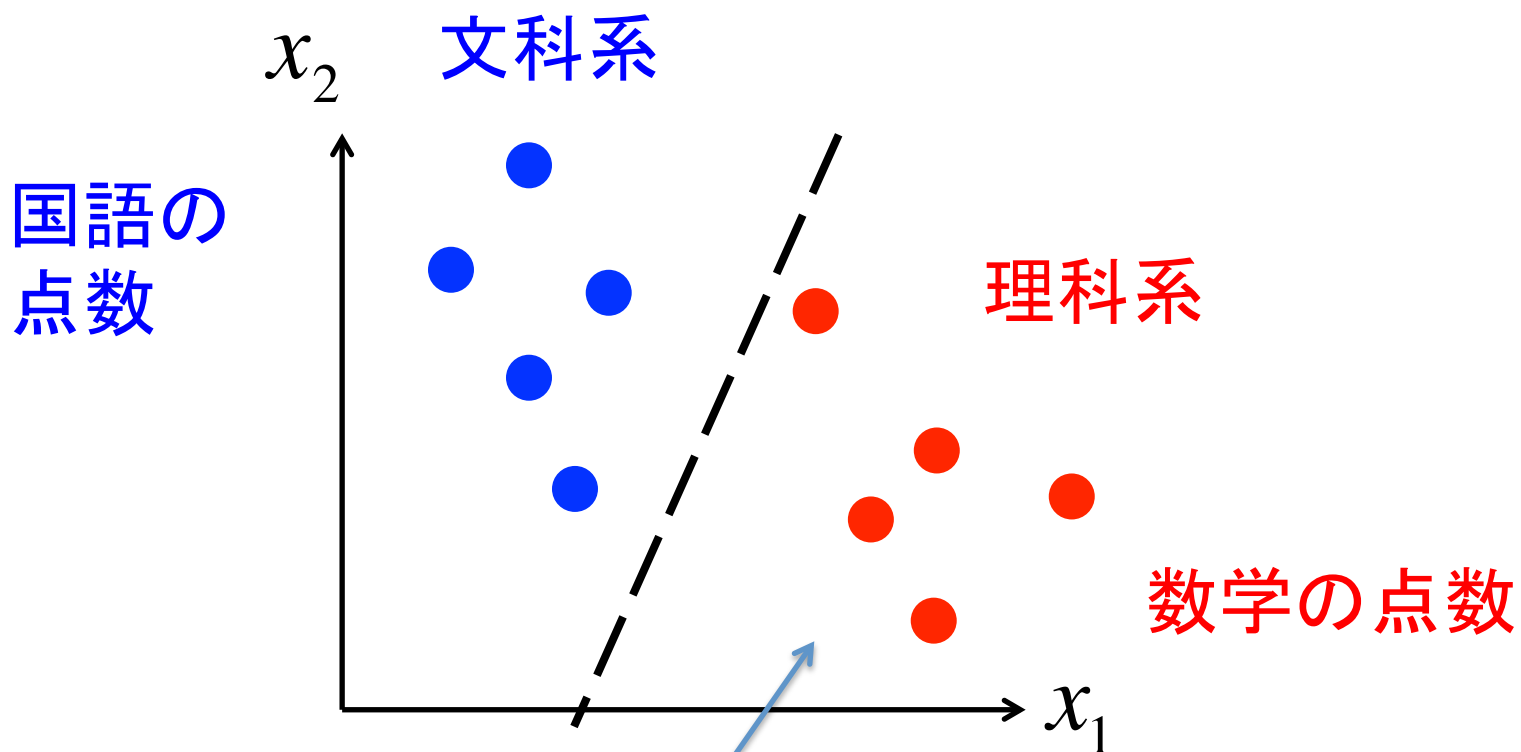


判別面(識別器): $ax_1 + bx_2 + c = 0$

上手く判別できるように, a, b, c を自動調整

スパースモデリング入門

線形判別モデル



判別面: $ax_1 + bx_2 + c = 0$

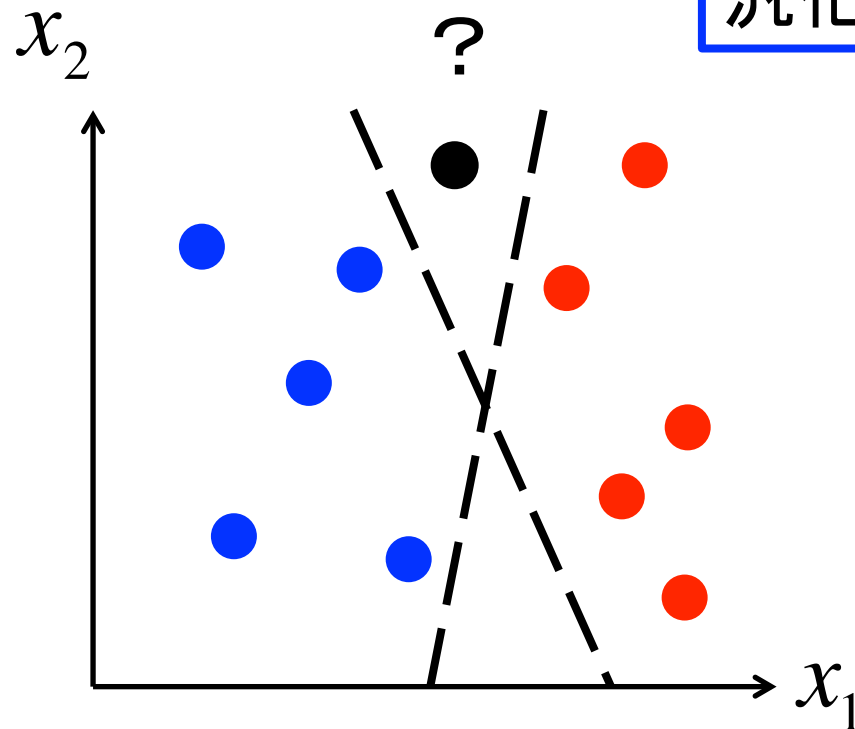
上手く判別できるように, a, b, c を調整

機械学習入門：汎化性能

判別の目的

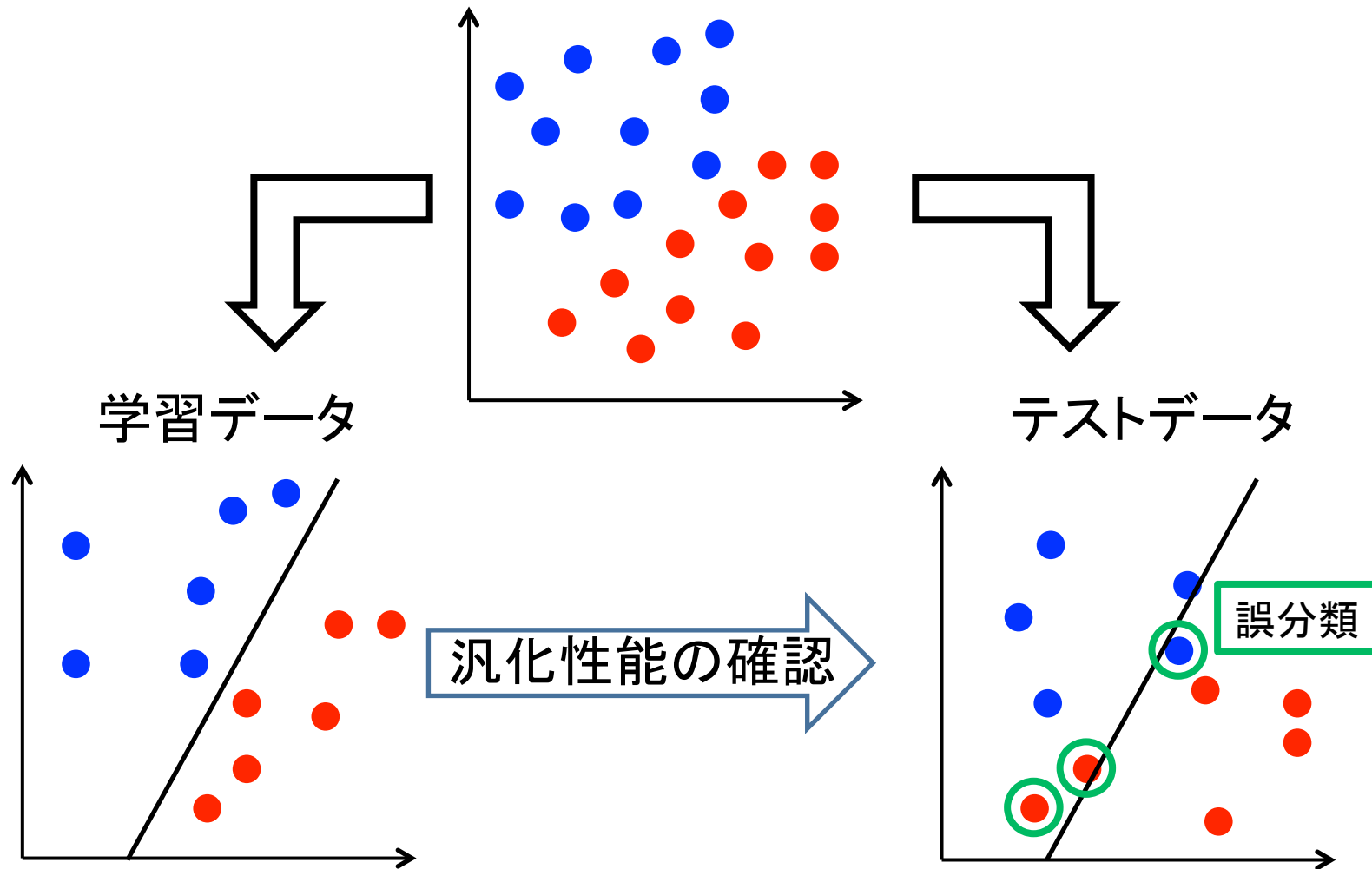
新たな未知データの判別性能をよくすること

汎化性能

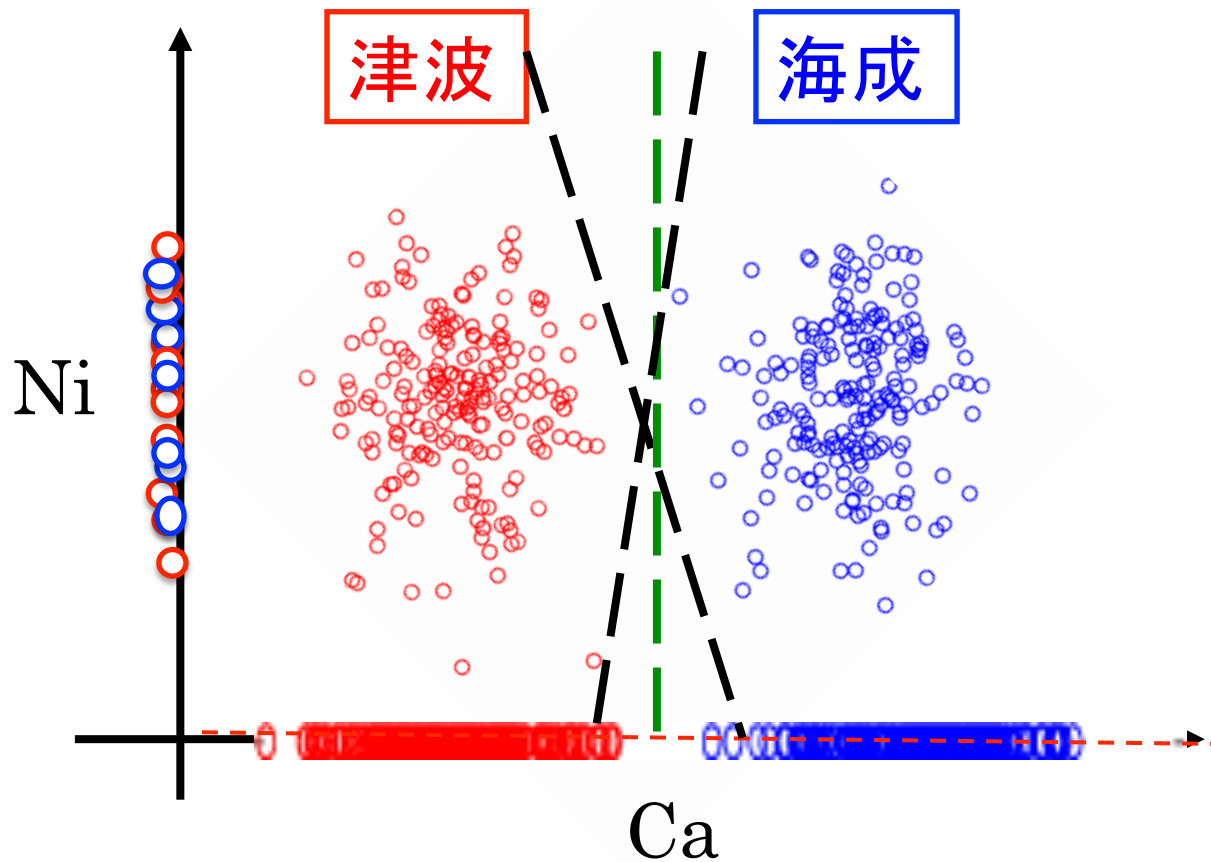


クロスバリデーション (CV): 交差検定法

- データを学習データとテストデータに分ける。



元素の選択(次元選択)

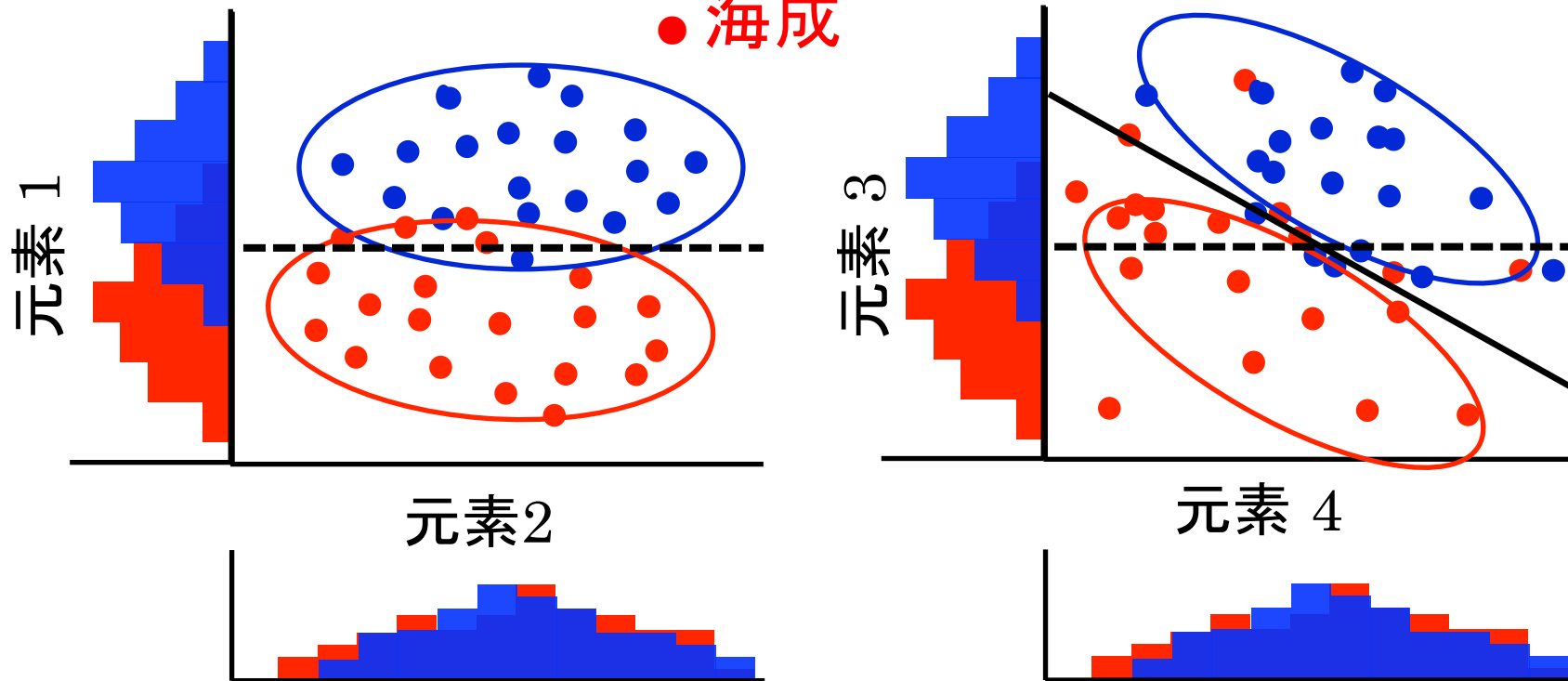


本当はNiは判別に寄与しない成分

⇒結果にばらつきが生じてしまい**汎化性能が悪化**

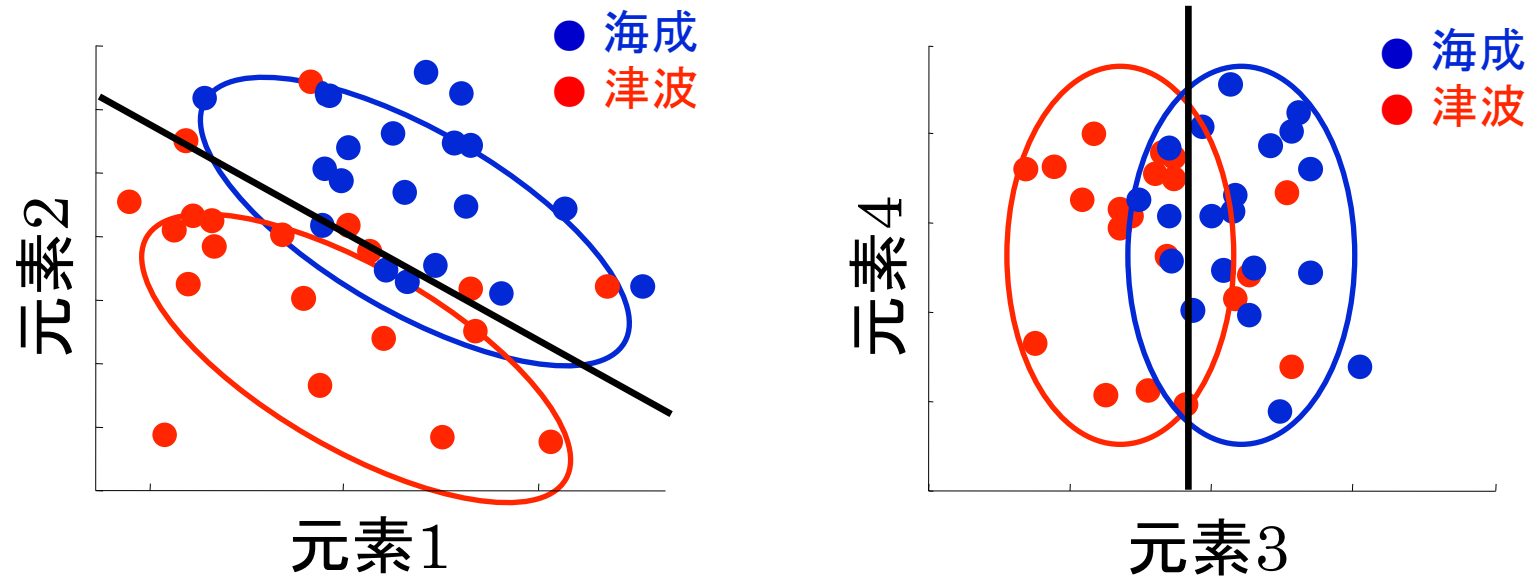
データのスパース化の難しさ 一つの元素だけではわからない

- 津波堆積物
- 海成



元素数 N に対して 2^N 通りの組み合わせが存在する
計算量の指数爆発

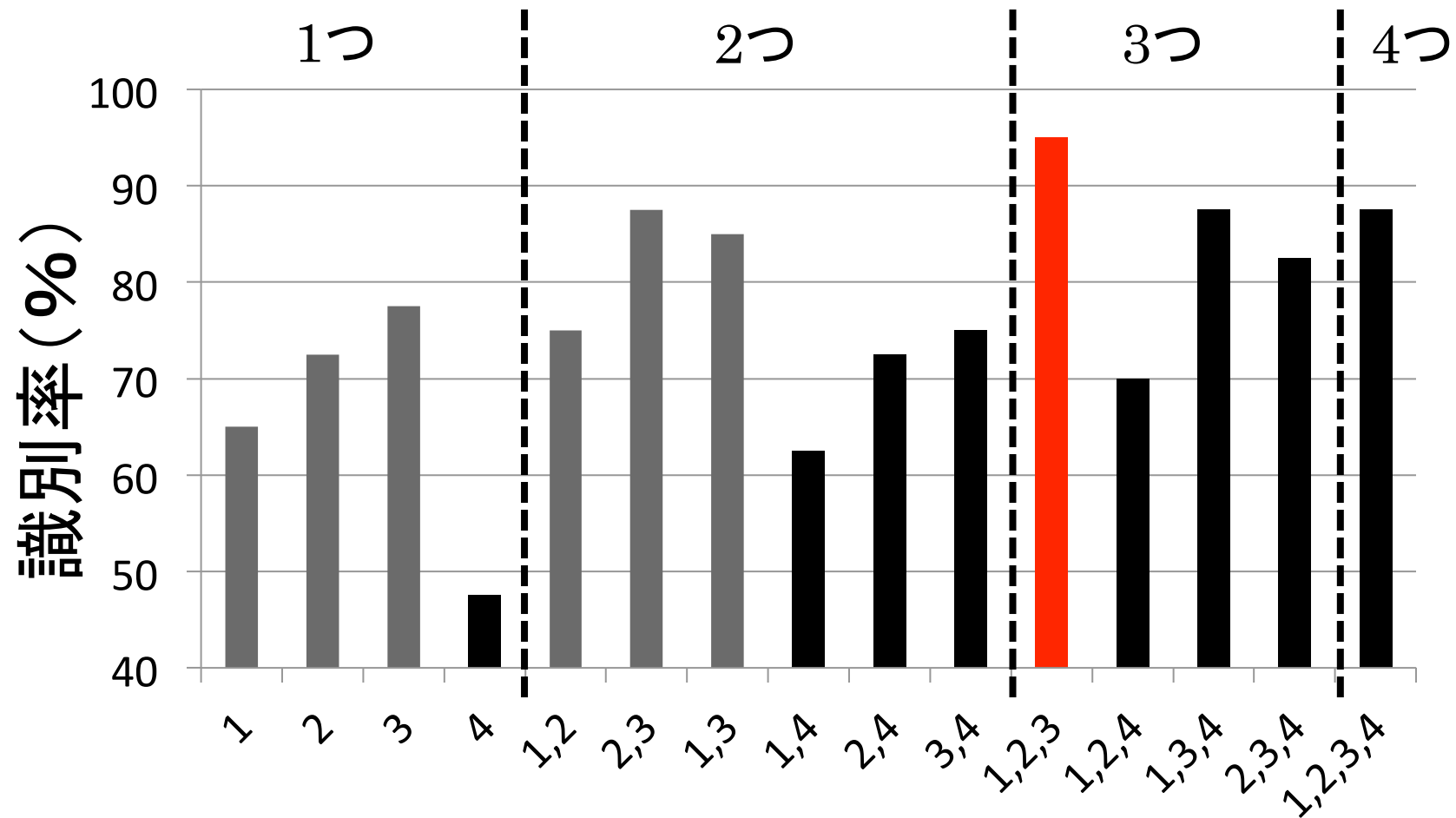
全部の組み合わせでの識別率



- 元素四つ全部使った場合，特定の三つを選んだ場合など，全部の組み合わせについて識別精度を比較する。
- どの組み合わせが一番精度がよくなるか？

Igarashi, Ichikawa, Nakanishi-Ohno, Takenaka, Kawabata, Eifuku, Tamura, Nagata and Okada, submitted to *Proc. of HD3-2017*, 2017.

全部の組み合わせでの識別率



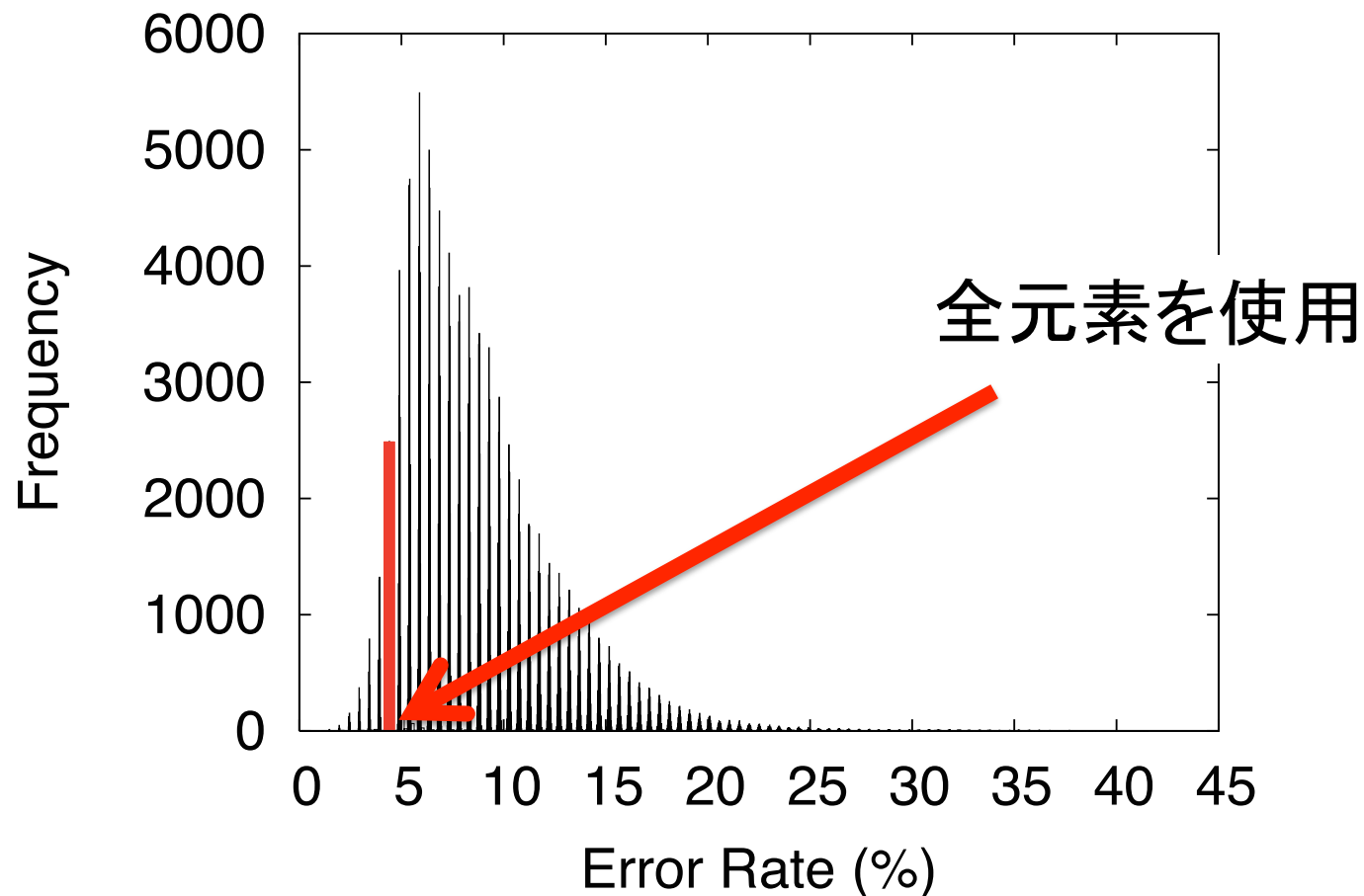
全組み合わせ (= $2^4 - 1$ 通り) を調べることで、
元素1,2,3のみが重要だということが分かった。

クロスバリデーションの結果

	訓練データ	CVE
(SiAl) & (Mg+Ca+Na+K)	91.39%	89.12%
全元素	100.00% 過学習	95.60%
Al, Ca, Ti, Mn, Cr, Sb, Cu, Zn, As, Cd, Pb	100.00%	100.00%

Kuwatani, Nagata, Okada, Watanabe, Ogawa, Komai and Tsuchiya, (2014)

ES-DoS (状態密度付き全状態探索法)



Kuwatani, Nagata, Okada, Watanabe, Ogawa, Komai and Tsuchiya, (2014)

Igarashi, Ichikawa, Nakanishi-Ohno, Takenaka, Kawabata, Eifuku,
Tamura, Nagata and Okada, submitted to *Proc. of HD3-2017*, 2017.

最先端の人工知能にもスパースモデリング
[Kim, Tadesse & Vannucci, Biometrika, 2006]

連立方程式とデータ駆動科学

連立方程式とその応用

鶴亀算 食塩水 寝坊
して追いかける問題

連立方程式への変換

加減法, 代入法

スパースモデリングPJ

実験・計測グループ
生物学, 地学

モデリング

機械学習

Igarashi, Nagata, Kuwatani, Omori, Nakanishi-Ohno and M. Okada, “Three Levels of Data-Driven Science”, *Journal of Physics: Conference Series*, 699, 012001, 2016.

(五十嵐, 竹中, 永田, 岡田, *応用統計学*, 2016)

MONOist IoT Forum 東京 講演レポート：

IoT時代のアナリティクスはエンジニアリングデータが主役になる

企業の関心はエンジニアリングデータ活用へ

- 生産装置の**センサーのデータ**や、**画像、動画**
- 顧客のプロファイルや、売上データ、在庫データなどのビジネスデータとは異なる



MathWorks Japan インダストリーマーケティング部
IA&Mインダストリーマーケティングマネジャーの遠山
巧氏

(Mathworks Japan, 遠山)

- 今後はセンサーや画像など、エンジニアリングデータも使っていきたいと考えている企業が多い

<http://monoist.atmarkit.co.jp/mn/articles/1701/12/news006.html>



IoTシステムを構築するために必要な5条件--エッジとサーバを全体最適化するために

大谷卓也 竹本佳充(MathWorks Japan)

IoTシステムを構築するために必要な5箇条

1. 精度の高い特徴量の抽出
2. 高度なアナリティクスの実装
3. コンピューティングリソースの全体最適化
4. エッジデバイスへのアルゴリズムの実装
5. アグリゲータでのコンピューティング環境

<https://japan.zdnet.com/article/35076390/>



IoTは新たな収益源を生み出すか-- ベンダー座談会 (1)

山田竜司 (編集部) 吉澤亨史 怒賀新也 (編集部) 2016/02/09 07:00

- **センサーデータ**を収集した際にどう上手くデータを前処理するか
- データサイエンティスト人材を見つけるのも大変ですが
- **関連分野のエキスパート**が、機械学習や最適化を簡単に使えるような製品をサポートし、業界全体を底上げする



(Mathworks Japan, 阿部)

<https://japan.techrepublic.com/article/35075285.htm>

まとめ

- スパースモデリング
 - スパース性でサンプリング定理を超える
 - スパース性でデータの本質を抽出
- データ駆動科学
 - データの本質を抽出する学問体系
 - 対象とする分野によらない普遍的体系
- エンジニアリングデータアナリティクス
 - Internet of Things, Cyber Physical System
 - エッジコンピューティング
 - 関連分野のエキスパートによるIT技術