

全状態探索型スパースモデリング

岡田真人

東京大学 大学院新領域創成科学研究科

自己紹介(理論物理学)

- 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)
 - 量産工場技術者: 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(福島研) (1989 - 1996)
 - ニューラルネットワーク(ディープラーニング)
- 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)
 - 情報統計力学
 - 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(強相関電子系, 表面, 地球惑星科学) (2004/07 –)
- (兼業) 物質・材料研究機構(NIMS) (2015/8/1 –)

本発表の主張

- データ駆動科学における特徴量選択や記述子抽出では、全状態探索型のスパースモデリングを用いるのが適切である.

内容

- 機械学習と特徴量選択
- 全状態探索法とWeight Diagram
- 特徴量選択の信頼度評価
- 拡張された特徴量空間における特徴量選択

線形回帰モデル

- 目的変数: y
- 準備した特徴量: (x_1, x_2, \dots, x_p)

関係式 $y = g(x_1, x_2, \dots, x_p)$ に対して, 線形和による近似を考える

$$\begin{aligned} y &= g(x_1, x_2, \dots, x_p) \\ &\approx w_1 x_1 + w_2 x_2 + \dots + w_p x_p \end{aligned}$$

ただし w_i は回帰係数

予測モデルの汎化性能

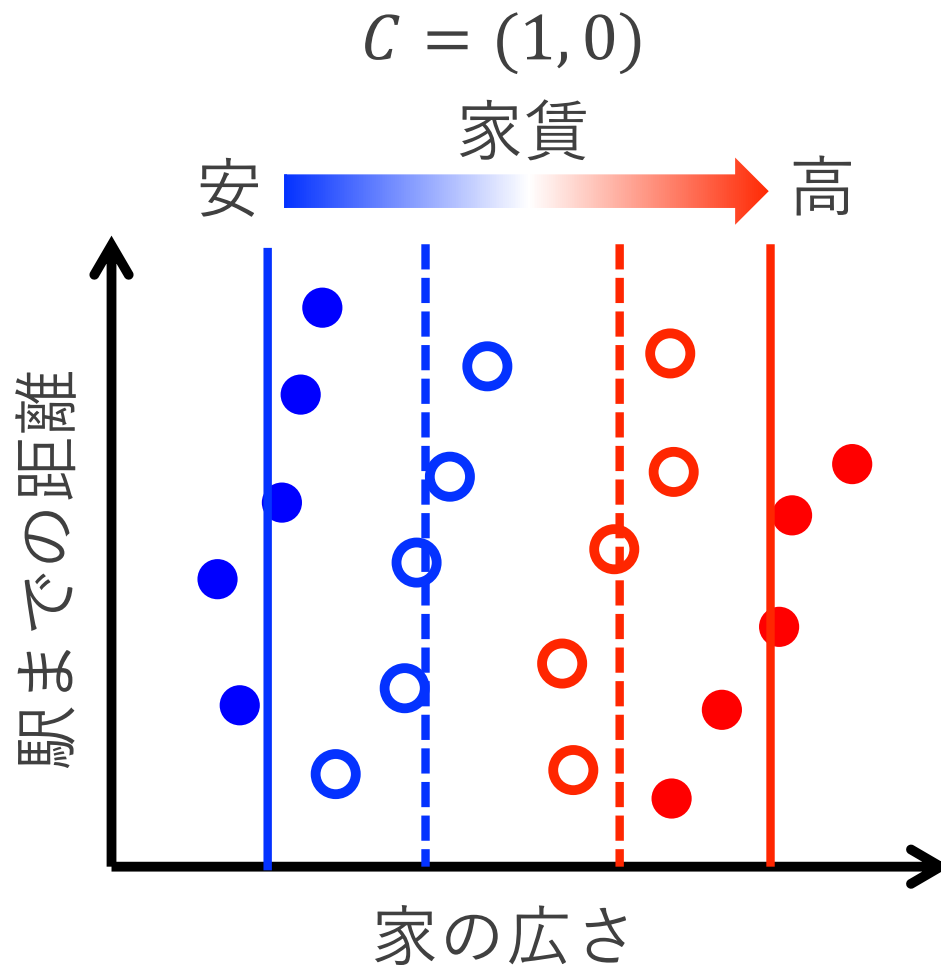
機械学習モデルに求められる性質

→ 未知データを上手く予測すること(汎化性能)

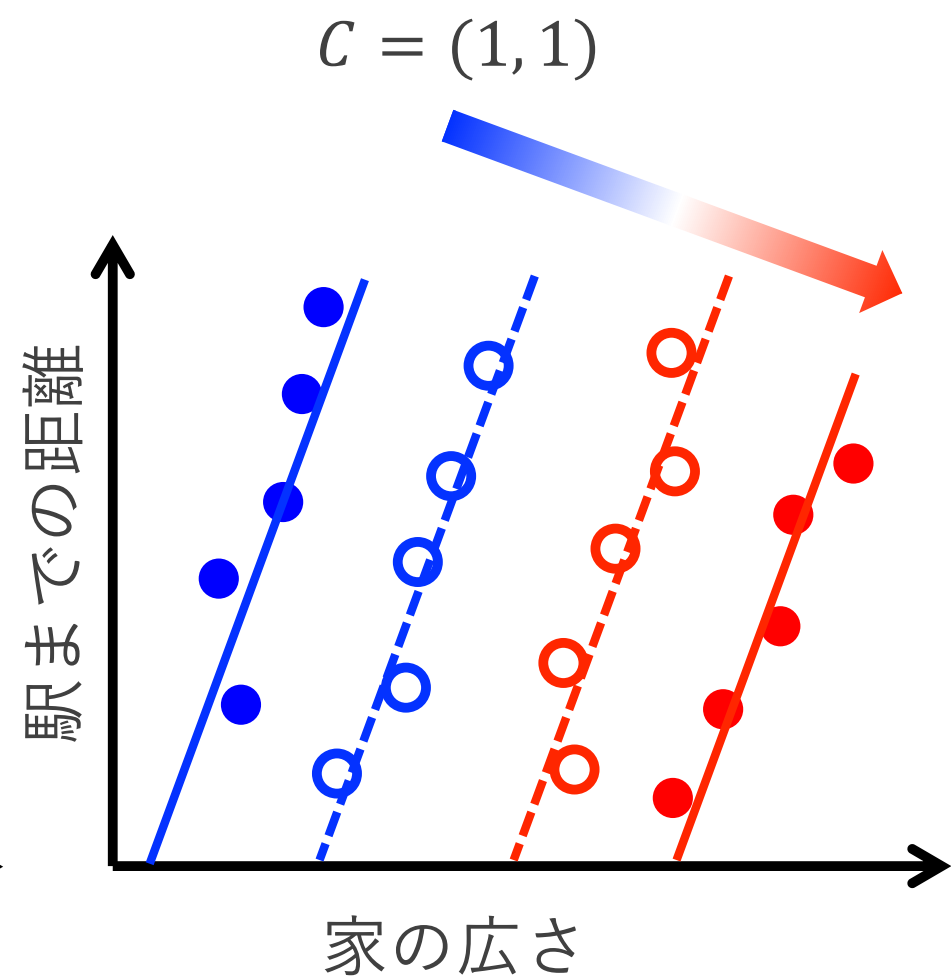
$$y \approx w_1x_1 + w_2x_2 + \cdots + w_px_p$$

汎化性能を高めるためには、必要な特徴量を見極めることが重要となる

特徴量選択の効果①



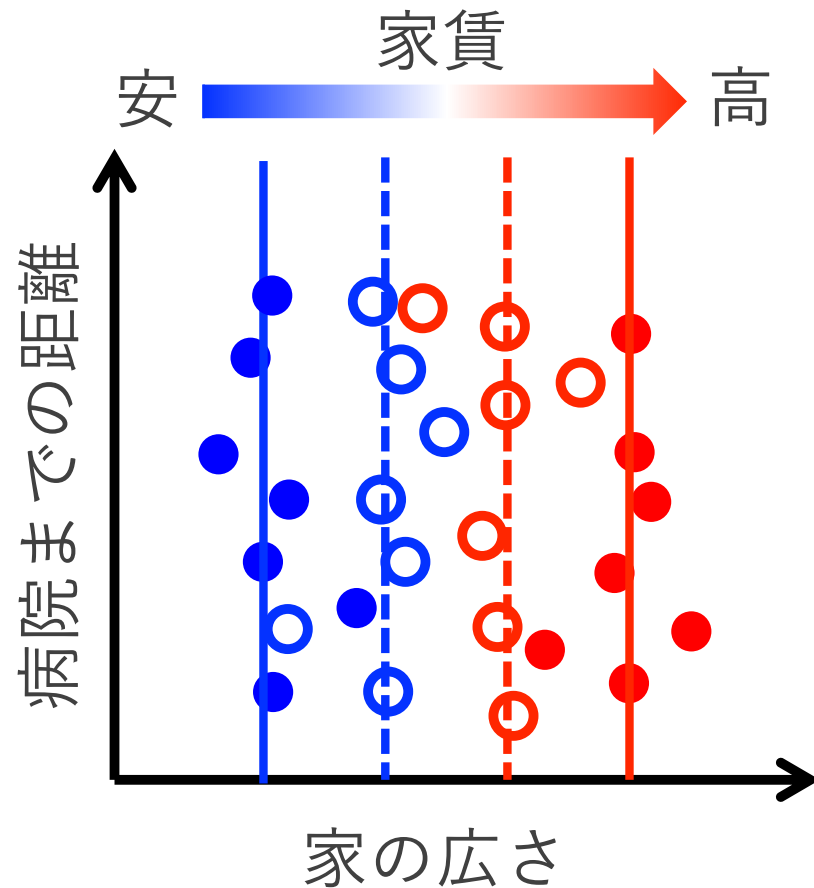
$$Error(1, 0) > Error(1, 1)$$



- 両方とも予測に必要

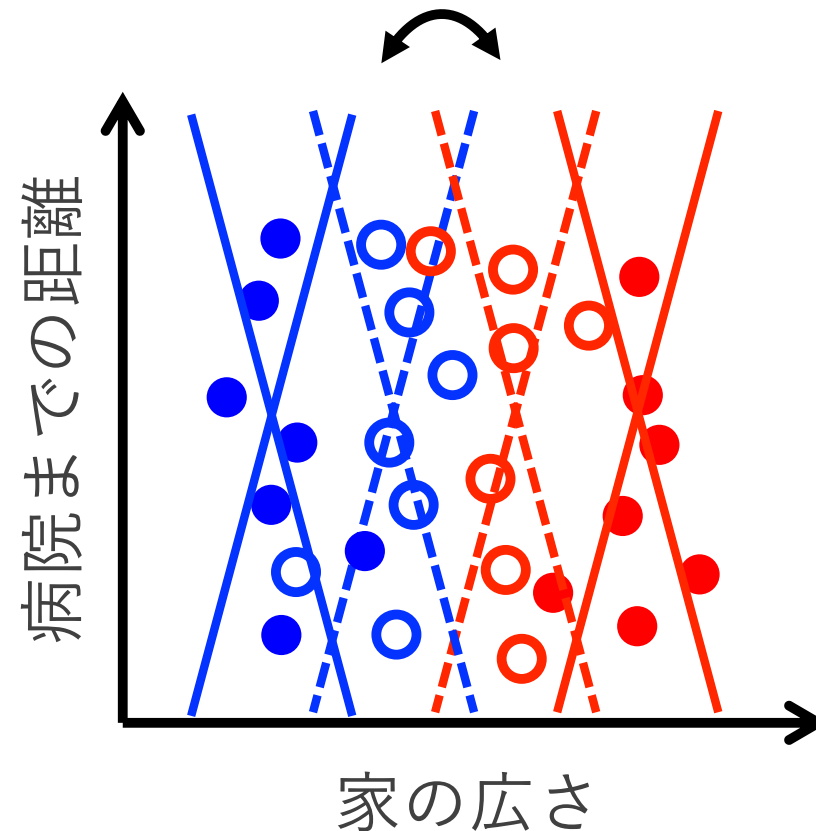
特徴量選択の効果②

家の広さのみ $C = (1, 0)$



$$Error(1, 0) < Error(1, 1)$$

家の広さと病院までの距離
 $C = (1, 1)$



- 家の広さ : 必要
- 病院までの距離 : 不要

物理学における特徴量選択の重要性

- 物理的知見の抽出

- 選ばれた特徴量と目的変数の関係性を考察することで、物理現象への知見が得られる

- Ghiringhelli *et al.*, 2015, Igarashi *et al.*, 2018

- 特徴量の準備の省力化

- 選ばれた特徴量のみ準備すればよいいため、実験の計測時間や回数及び数値シミュレーションの回数を削減することが可能

- Nakanishi-Ohno *et al.*, 2016

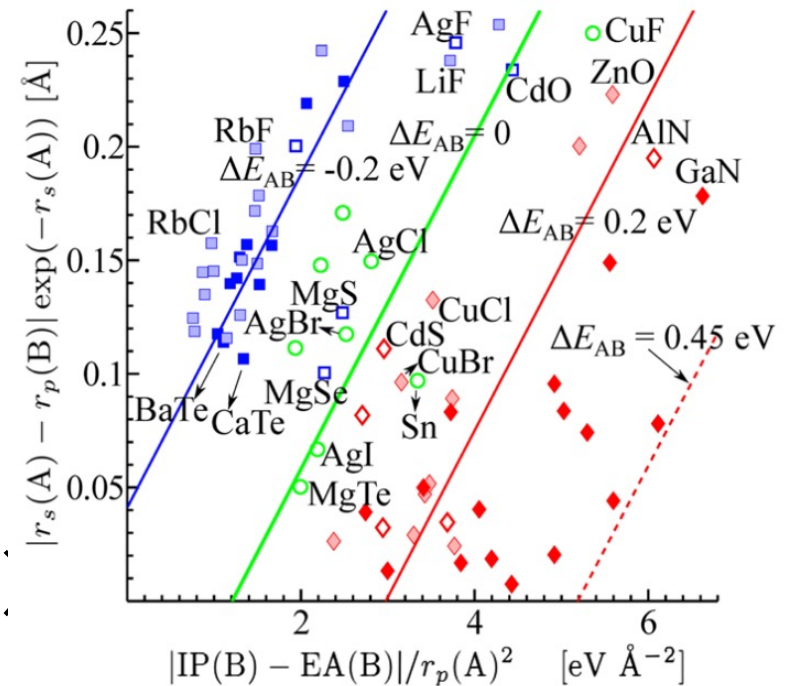
材料科学における特徴量選択の応用

1. 電池材料に関する化学反応の安定性予測

Sodeyama *et al*, 2018

2. 半導体化合物の結晶構造予測

Ghiringhelli *et al.*, 2015 (右図)



2つの特徴量からなる線形
モデルによる結晶構造予測

内容

- 機械学習と特徴量選択
- 全状態探索法とWeight Diagram
- 特徴量選択の信頼度評価
- 拡張された特徴量空間における特徴量選択

データの表記

- 目的変数: $y \in \mathbb{R}$
- 特徴量: $\boldsymbol{x} = (x_1, x_2, \dots, x_p)^T \in \mathbb{R}^p$

N 個の観測データ

$$\boldsymbol{y} = (y_1, \dots, y_N)^T \in \mathbb{R}^N$$

$$X = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)^T \in \mathbb{R}^{N \times p}$$

(補足) 目的変数 y と各特徴量 $x^{(i)}$ は正規化等の前処理済みとする

最小二乗法による線形回帰

真のモデルを線形関数でモデル化:

$$\begin{aligned} f(\mathbf{x}; w_0, \mathbf{w}) &= w_0 + w_1 x_1 + \cdots + w_p x_p \\ &= w_0 + \mathbf{w}^T \mathbf{x} \end{aligned}$$

– 切片: $w_0 \in \mathbb{R}$

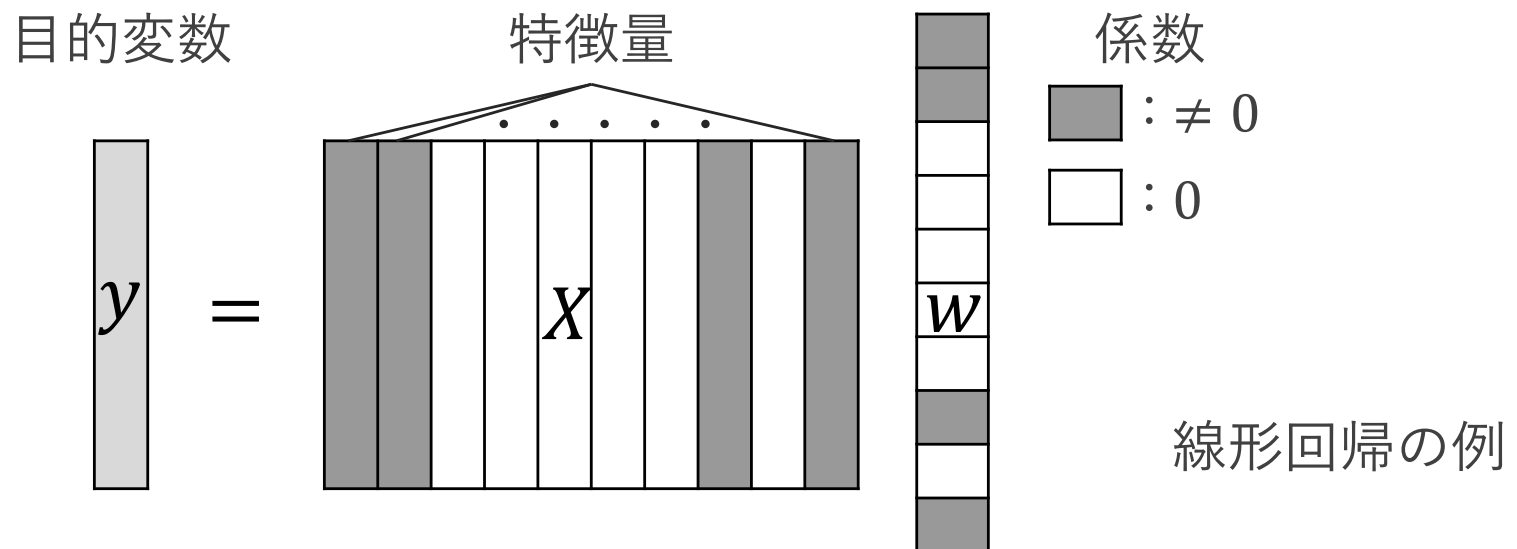
– 係数: $\mathbf{w} \in \mathbb{R}^p$

平均二乗誤差の最小化により, 係数及び切片の値を求める

$$E(w_0, \mathbf{w}) = \frac{1}{N} \sum_{n=1}^N (y_n - f(\mathbf{x}_n; w_0, \mathbf{w}))^2$$

特徴量選択

目的変数を説明するために重要な特徴量のみを選び出すことを目的とした，統計的機械学習の一分野



代表的な手法

- Lasso (Tibshirani, 1996)
- 全状態探索 (Exhaustive search) (Igarashi et al., 2018)

変数選択に関する二つの戦略

原則：変数選択の問題の計算量は指数爆発する
(Cover and Van Campenhout, 1977)

＜変数選択に対する二つの戦略＞

1. 凸最適化や変分ベイズにもとづく緩和型アプローチ
通常に用いられているスパース推定のアルゴリズム
凸最適化: Lasso(1996, L1正則化)
変分ベイズ: ARD(2008, 関連自動度決定)
2. 全状態を効率的に探索するサンプリングアプローチ
サンプリング: MCMC(1993, マルコフ連鎖モンテカルロ法)
REMC(2006, レプリカ交換モンテカルロ法)

サンプリングアプローチの方が歴史は長い

インジケータベクトルによる モデルの指定

インジケータベクトル c によって部分モデルを定義
 $c = (1, 1, 0, \dots, 0, 1, 0, 1) \in \{0, 1\}^p$

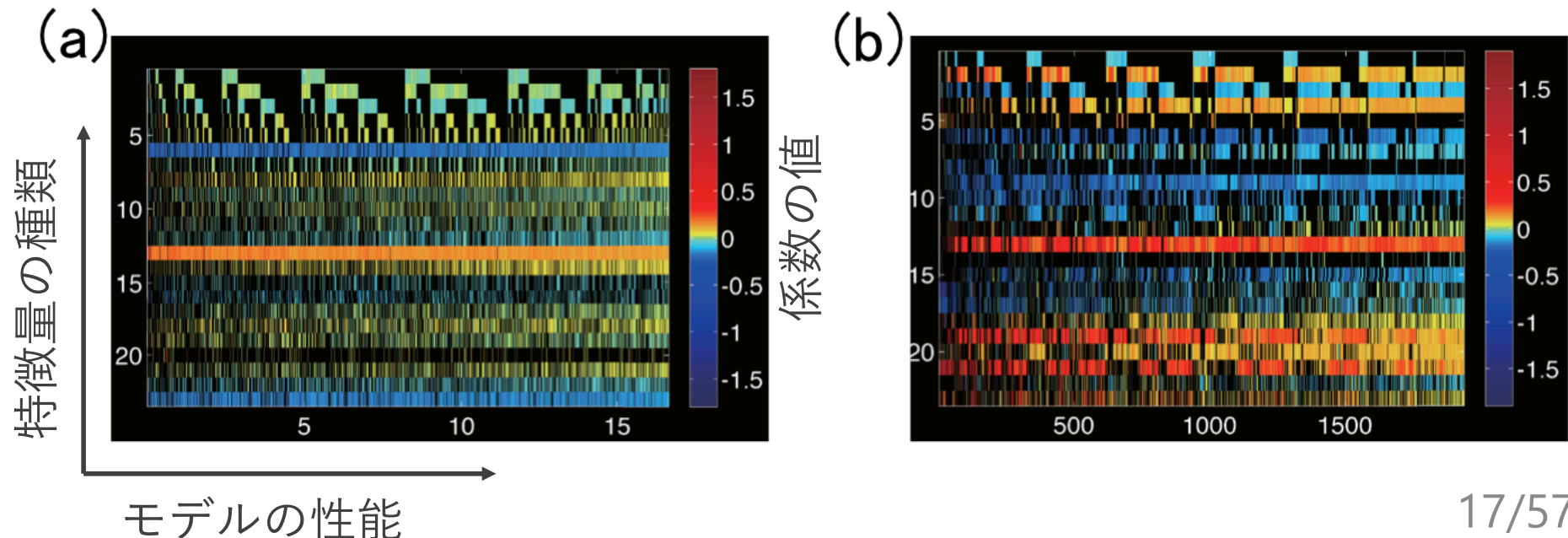
- i 番目の要素が特徴量 i に対応
 - 1: モデルに含まれる
 - 0: モデルに含まれない
- インジケータベクトルは 2^p 状態を取る
 - その内1つは空のモデル

厳密な特徴量選択を行うにはどの手法であっても
計算量が指数関数的に増加する(Cover and Van
Campenhout, 1977)

全状態探索とWeight Diagram

全状態探索: c の組み合わせを全て調べ, 性能の良いモデルを探索する枠組み

探索結果から性能上位のモデルの係数をプロットする (Weight diagram) ことで, 特徴量選択の揺らぎを定性的に分析する (Y. Igarashi *et al.* 2018 J. Phys.)



p (変数の数)が20程度まであれば
厳密な全状態探索が適切 (1/3)



市川寛子
東京理科大



桑谷 立
JAMSTEC



五十嵐康彦
筑波大

p (変数の数)が20程度まであれば 厳密な全状態探索が適切 (2/3)

- 脳科学/行動科学
 - NIRS: ASDとADHDの鑑別
 - Ichikawa *et al.*, 2014 Front Hum Neurosci
 - 歩行動作からASD症状の強さの回帰
 - Shigeta *et al.*, 2018 Adv Biomed Eng
- 地球科学
 - 高精度な津波堆積物の判別を実現
 - Kuwatani *et al.*, 2014 Scientific Reports
 - 含水マグマの熱力学モデルを構築
 - Ueki *et al.*, 2020 Physics of the Earth and Planetary Interiors,
 - 岩石流体相互作用における反応経路を全探索によって決定
 - Oyanagi *et al.*, 2021 The European Physical Journal
 - その他 3本

p (変数の数)が20程度まであれば 厳密な全状態探索が適切 (3/3)

- 物質材料科学
 - リチウムイオン電池の電解液材料探索への応用
Sodeyama *et al.*, 2018 Phys. Chem. Chem. Phys, その他3本
 - 高収率なナノシート合成開発への応用
Nakada *et al.*, 2019 Adv. Theory & Sim. その他4本
 - リチウムイオン二次電池用 有機負極活物質の探索
Nakada *et al.*, 2019 Adv. Theory & Sim. その他3本
 - 熱応答性高分子の機能予測モデルの開発
Hiruta *et al.*, under review
- 核融合
 - トカマク型核融合炉JT-60Uにおける高ベータ障害予測モデル構築
Yokoyama *et al.*, 2019 Fusion Engineering and Design その他1本
 - 大型ヘリカル装置における放射性崩壊回避のためのデータ駆動型制御
Yokoyama *et al.*, 2022 Plasma and Fusion Research

内容

- 機械学習と特徴量選択
- 全状態探索法とWeight Diagram
- 特徴量選択の信頼度評価
- 拡張された特徴量空間における特徴量選択

共著者



大日方孝輝
東京大学



中山智文
カラクリ株式会社



石川敦之
NIMS



袖山慶太郎
NIMS



永田賢二
NIMS

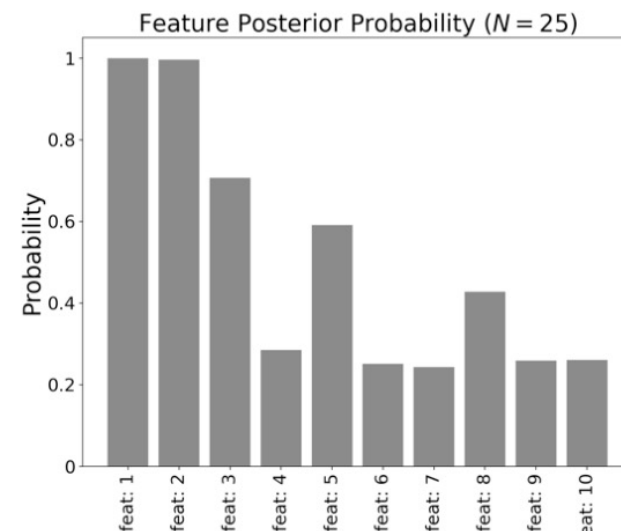
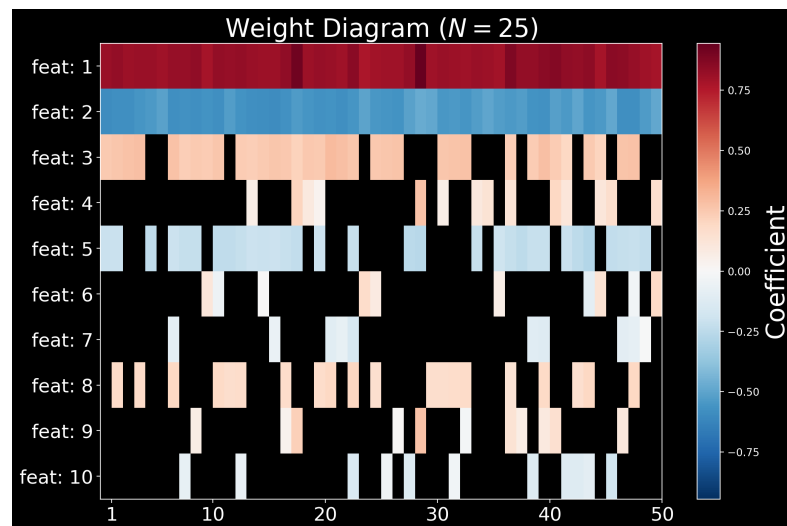


五十嵐康彦
筑波大

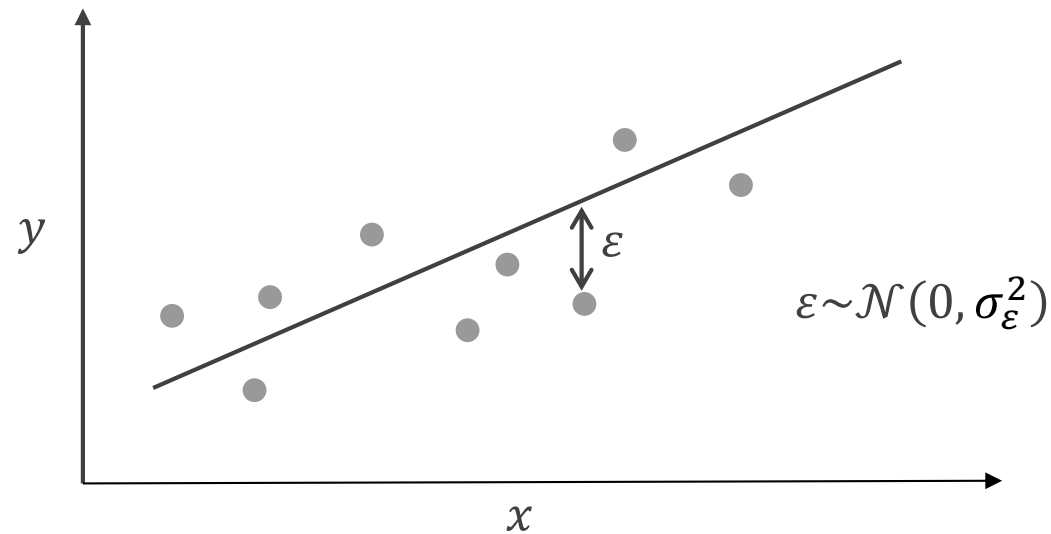
BMAによる特徴量選択の信頼度評価

全状態探索の結果をベイズモデル平均化
(BMA) の枠組みで周辺化し, 得られた周辺事
後確率によって特徴量選択の信頼度を評価

→ 定量的な信頼度評価を伴った特徴量選択へ



観測過程



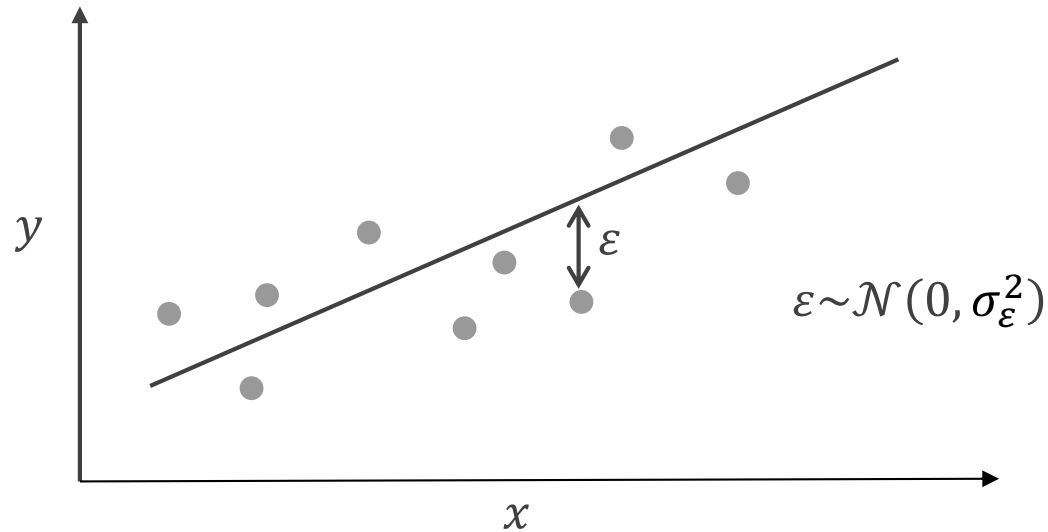
観測ノイズ ε :

$$y = f(\mathbf{x}; w_0, \mathbf{w}) + \varepsilon$$

確率分布としての表現:

$$p(y|\mathbf{x}, w_0, \mathbf{w}, \sigma_\varepsilon) = \mathcal{N}(y|w_0 + \mathbf{x}^T \mathbf{w}, \sigma_\varepsilon^2)$$

観測過程



N 個の独立なデータを観測:

$$\begin{aligned} p(\mathbf{y} | X, w_0, \mathbf{w}, \sigma_\varepsilon) &= \prod_{n=1}^N p(y_n | x_n, w_0, \mathbf{w}, \sigma_\varepsilon) \\ &= \mathcal{N}(\mathbf{y} | w_0 \mathbf{1} + X\mathbf{w}, \sigma_\varepsilon^2 I_N) \end{aligned}$$

パラメータの事前分布

\mathbf{w} と w_0 に対する事前分布:

$$p(w_i) = \mathcal{N}(w_i | 0, \sigma_w^2)$$

$$p(w_0) = \mathcal{N}(w_0 | 0, \sigma_y^2)$$

同時分布:

$$p(\mathbf{y}, w_0, \mathbf{w} | X, \sigma_\varepsilon, \sigma_w)$$

$$= \underbrace{\mathcal{N}(\mathbf{y} | w_0 \mathbf{1} + X\mathbf{w}, \sigma_\varepsilon^2 I_N)}_{\text{観測過程}} \underbrace{\mathcal{N}(w_0 | 0, \sigma_y^2) \mathcal{N}(\mathbf{w} | 0, \sigma_w^2 I_p)}_{\text{事前知識}}$$

上の同時分布は \mathbf{w} と w_0 に関する事後分布に比例した値となっており, 最大事後確率推定はRidge回帰と等価となる

最小二乗法との関係

\mathbf{w} に関する最大事後確率推定はRidge回帰と等価となる

$$p(w_0, \mathbf{w} | X, \mathbf{y}, \sigma_\varepsilon, \sigma_w)$$

$$\propto p(\mathbf{y}, w_0, \mathbf{w} | X, \sigma_\varepsilon, \sigma_w)$$

$$= \mathcal{N}(\mathbf{y} | w_0 \mathbf{1} + X\mathbf{w}, \sigma_\varepsilon^2 I_N) \mathcal{N}(w_0 | 0, \sigma_y^2) \mathcal{N}(\mathbf{w} | 0, \sigma_w^2 I_p)$$

$$\propto \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \|\mathbf{y} - (w_0 \mathbf{1} + X\mathbf{w})\|^2 - \frac{1}{2\sigma_w^2} \|\mathbf{w}\|^2 - \frac{1}{2\sigma_y^2} w_0^2 \right]$$

$$\propto \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \|\mathbf{y} - X\mathbf{w}\|^2 - \frac{1}{2\sigma_w^2} \|\mathbf{w}\|^2 \right] \exp \left[-\frac{1}{2} \left(\frac{N}{\sigma_\varepsilon^2} + \frac{1}{\sigma_y^2} \right) w_0^2 \right]$$

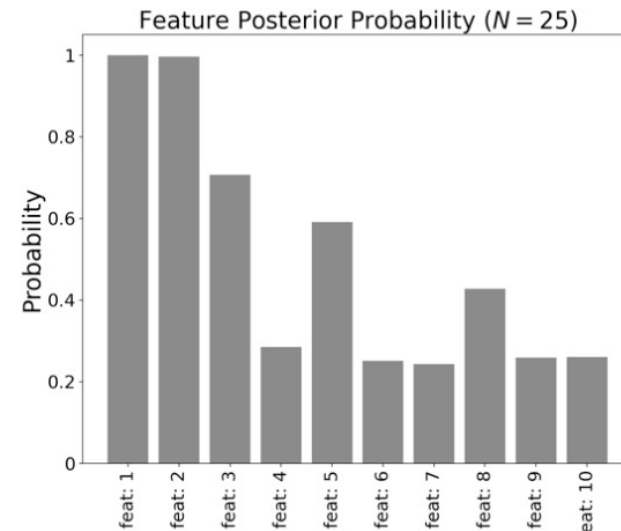
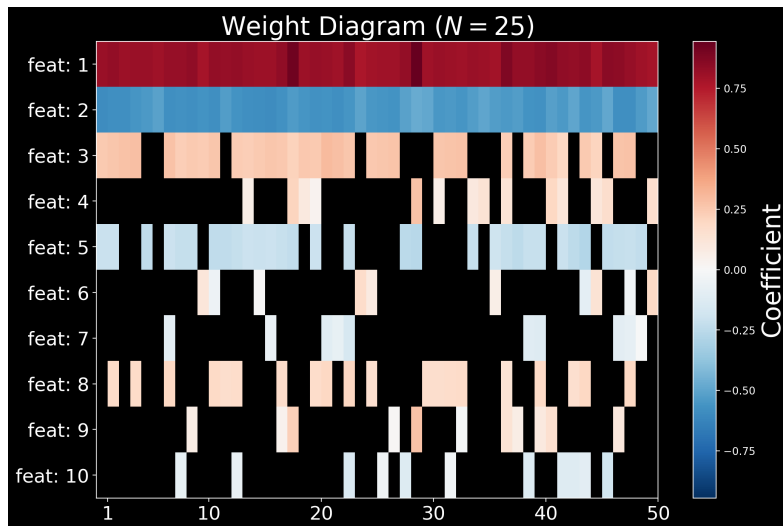
誤差項

L2正則化項

特徴量の周辺事後確率

i 番目の特徴量がモデルに含まれる確率（特徴量 i の周辺事後確率）を考える

$$\begin{aligned} p(c_i = 1|X, \mathbf{y}) &= \sum_{\mathbf{c}_{\setminus i}} p(c_i = 1|\mathbf{c}_{\setminus i}, X, \mathbf{y}, \alpha) p(\mathbf{c}_{\setminus i}|X, \mathbf{y}) \\ &= \sum_{\{\mathbf{c} | c_i = 1\}} p(\mathbf{c}|X, \mathbf{y}) \end{aligned}$$



ベイズモデル平均化 (BMA)

BMAでは各モデルの不確実さを考慮したモデル平均によって推論を行う

– David Madigan and Adrian E. Raftery, 1994

$$p(\tau|X, \mathbf{y}) = \sum_k p(\tau|M_k, X, \mathbf{y})p(M_k|X, \mathbf{y})$$

– M_k : 考えているモデル

- M_k はインジケータベクトル c で指定される

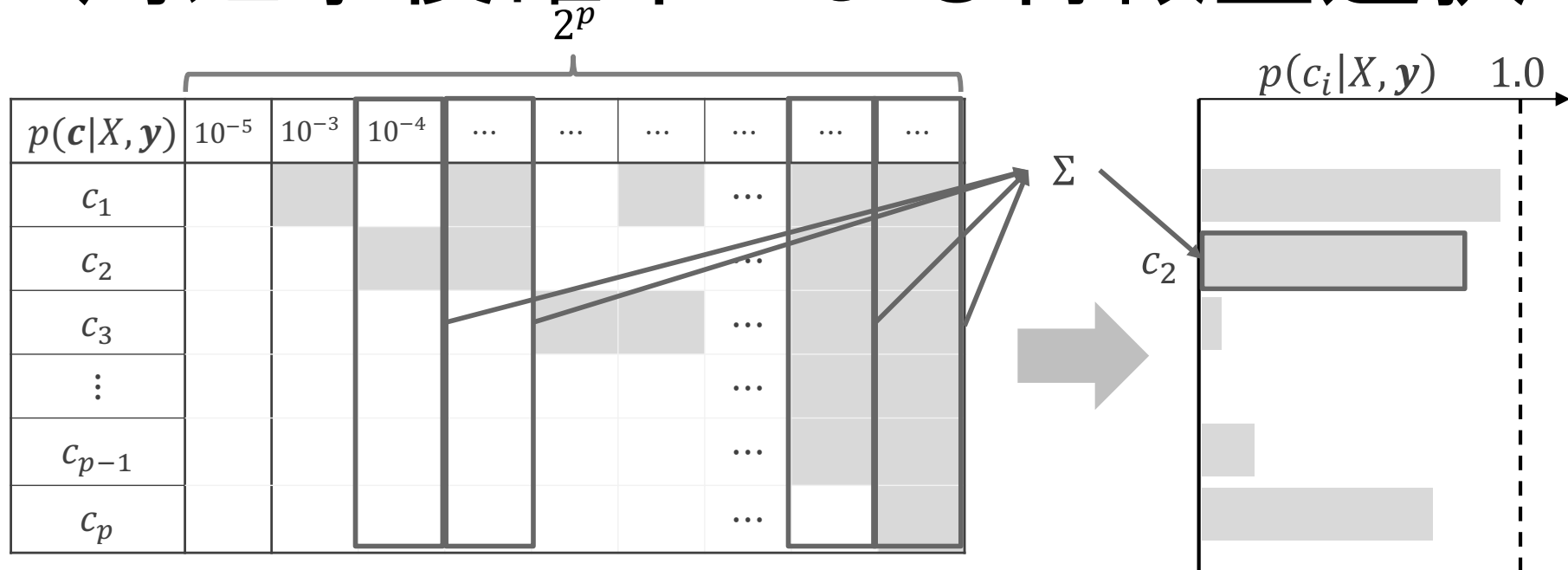
– τ : 推定対象の変数

- 新規データ x_{new} に対する予測 y_{new}

- 特定の係数値 w_i

- i 番目の特徴量が含まれるか否か ($c_i = 1$)

周辺事後確率による特徴量選択

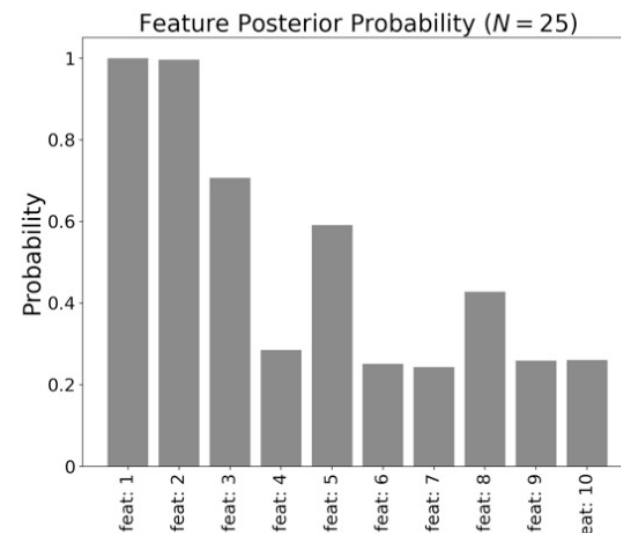
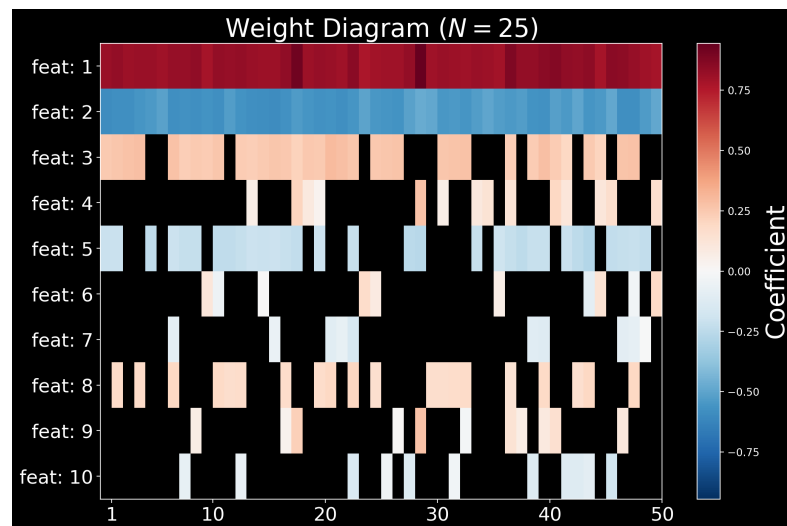


- $p(c_i = 1|X, \mathbf{y})$ によって特徴量選択を行う
 - 特定の条件下において, 上記の確率に対する閾値0.5が予測における最適な特徴量選択基準となる(Barbieri et al., 2018)
- 選択基準 $p(c_i = 1|X, \mathbf{y})$ 自体が選択の確信度となる
 - 精度 (Precision) に対応する値となる (Viallefont et al., 2001)

BMAによる特徴量選択の信頼度評価

全状態探索の結果をベイズモデル平均化
(BMA) の枠組みで周辺化し, 得られた周辺事
後確率によって特徴量選択の
信頼度を評価

→ 定量的な信頼度評価を伴った特徴量選択へ



内容

- 機械学習と特徴量選択
- 全状態探索法とWeight Diagram
- 特徴量選択の信頼度評価
- 拡張された特徴量空間における特徴量選択

- この節の内容は論文投稿後にアップします。