

# 講義スライドのイメージ

第1回～第12回のスライド（全400枚）の中から数枚を抜粋

# MI解析OJTの目的

## 目的

- ・ご自身の研究テーマでMI解析を実践し、スキルを身に付けて頂く
- ・各部署でMIを教えられるエバンジェリストとなって頂く

## 本OJTの目標到達レベル

- ・ MIに必要な知識を理解すること
  - 解析フロー、統計、機械学習等の知識
- ・ MIを用いて解析を行うこと
  - ツールを活用（本質的な内容にフォーカス）
  - 機械学習モデルを作り、実験系の傾向の把握や、目標物性を得る為のパラメータ予測を行う



# 解析OJTの進め方

OJTは4つの要素で構成されています

【必須】

## 定例会（講義）

【場所】  
オンライン（Teams会議）

【日時】  
毎週X曜13:30~15:00  
(X/X~X/X)

- 【内容】
- ・ 解析ツール使い方
  - ・ 解析の基礎知識
  - ・ 現状報告
  - ・ 所連絡



【必須】

## 宿題

【〆切】  
毎週火曜12:00  
(X/X~X/X)

- 【内容】
- ・ 課題の整理
  - ・ データの整理、加工処理
  - ・ 解析ツールで解析
  - ・ 報告資料作成
  - ・ アンケート回答



【基本的に必須】

## 検証実験

【実施期間（目安）】  
追加実験：X/X~X/X  
検証実験：X/X~X/X

- 【内容】
- ・ 不足データ補充
  - ・ モデル精度検証

※追加実験は、モデルの精度に納得できない場合に行う。



【任意】

## 相談会、質問チャンネル

【場所】  
XX会議室  
&  
オンライン（Teams会議）

【日時】  
毎週月曜15:00~17:00  
(X/X~X/X)

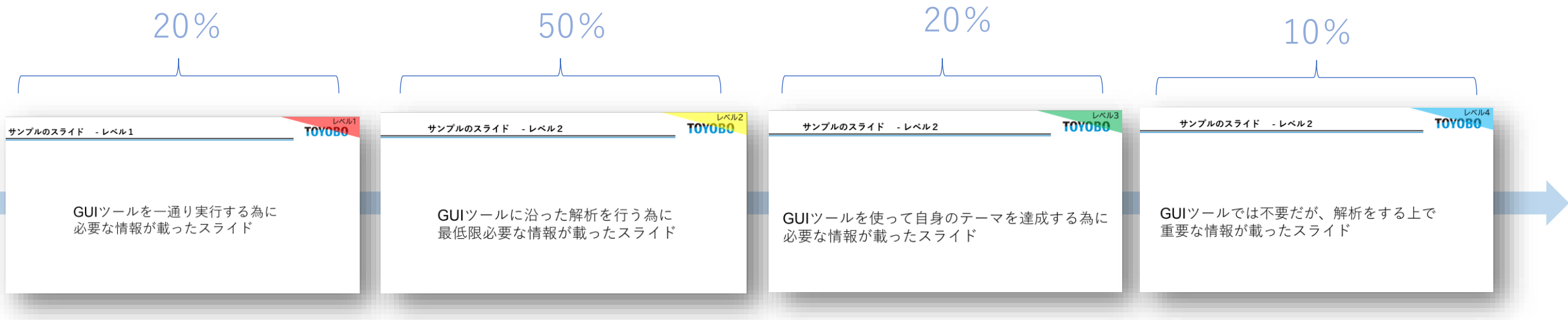
- 【内容】
- ・ 宿題の相談
  - ・ 講義内容の質問
  - ・ 方針の相談





- スライド右上の印の色が、理解して欲しいスライドの優先度です

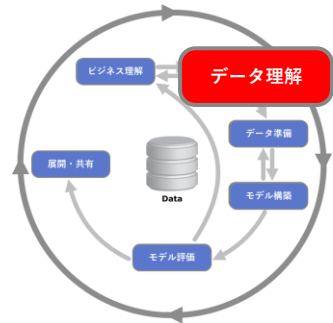
赤、黄、緑、青の順でスライドを理解して頂くと、分かり易いです。



- 3か月のOJTでお伝えする内容は膨大なので（合計400スライド）、優先度を付けました。
- ツールの使い方、解析のキホン、解析の実践的なコツ、詳細な理論…  
の順で理解して頂きたいと考えています。

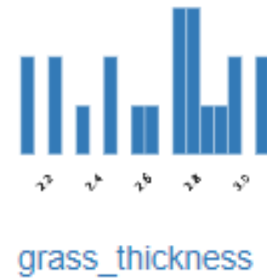
# • MI解析ツールの簡単な紹介

## • Step1では、データを入力し、データの基礎分析が行えます

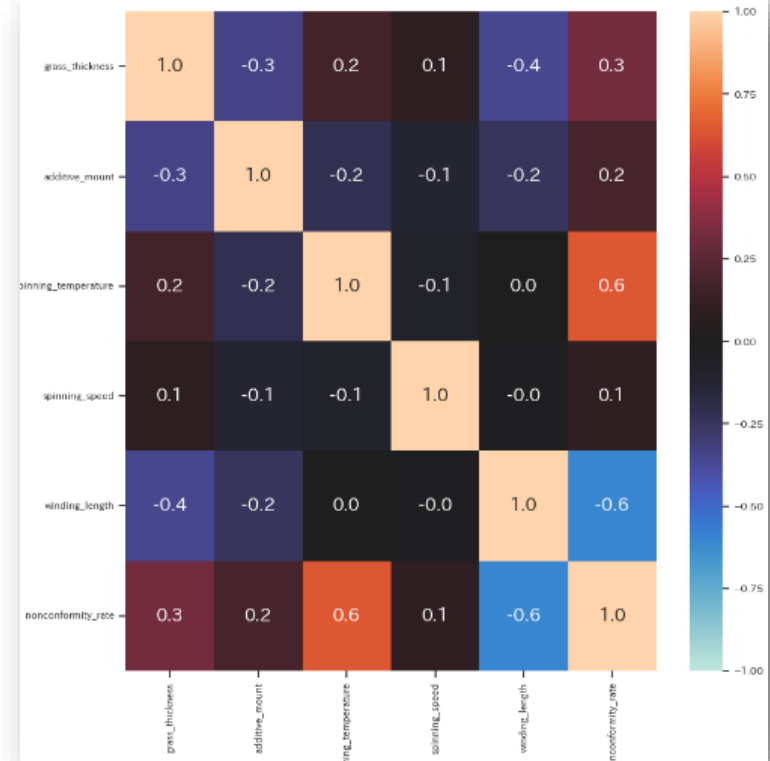


1 変数の分布（基本統計量、ヒストグラム）

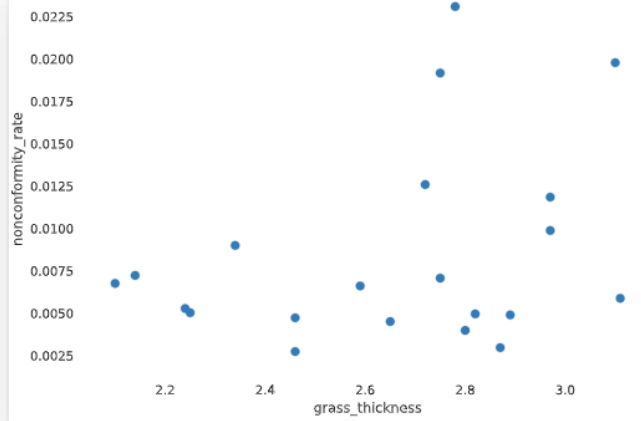
Distinct	18	Minimum	2.1
Distinct (%)	85.7%	Maximum	3.11
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	2.655238095	Memory size	296.0 B



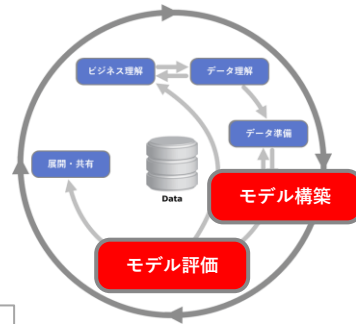
全変数の相関（ヒートマップ）



2 変数のプロット（散布図）



- MI解析ツールの簡単な紹介
- Step2では、入力データから機械学習モデルが作成できます



3ジャンルのモデルから選択

### モデルの指定

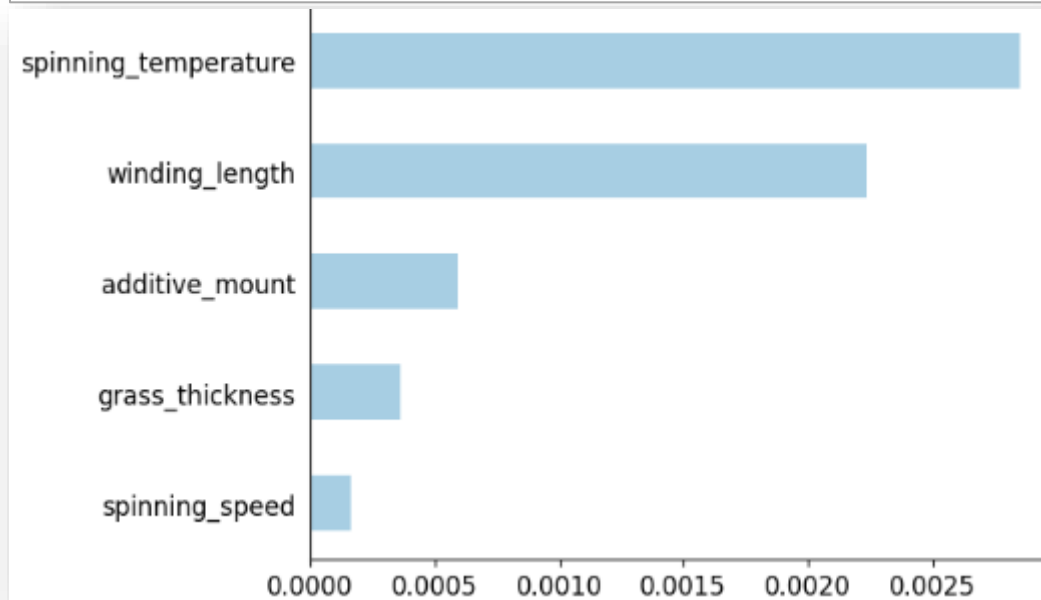
☒ Regression ☐ BinaryClassification ☐ MulticlassClassification

モデル構築

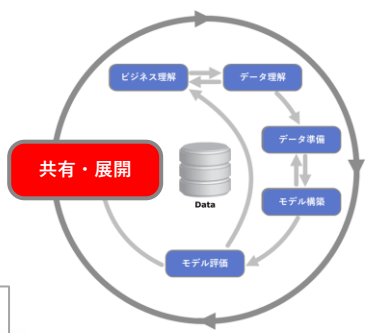
モデル精度を自動で測定

評価値	R2	MAE
	1.0	0.0

モデルの構造をグラフで可視化  
(寄与率の高い変数をランキング表示)



- MI解析ツールの簡単な紹介
- Step3では、機械学習モデルから推定値を算出できます



説明変数（因子）のとりうる範囲を設定

変数の値範囲を設定

個別設定

パラメータ:  下限値:  上限値:

変数の固定値を設定

パラメータ:  固定値:

推定値の目標範囲と、探索手法を設定

探索手法を設定

☒ Simulate ☐ Search\_Cluster ☐ Search\_Pareto

目的変数の目標値を設定

目的変数:  下限値:  上限値:

目標範囲を満たす為の、説明変数（因子）の推定値を算出（複数）

grass_thickness	additive_mount	spinning_temperature	spinning_speed	winding_length	nonconformity_rate
2.91	122.85	0.00	0.00	0.00	0.00
2.57	121.00	0.00	613.07	868.24	0.00
2.36	124.50	1060.38	0.00	2040.73	0.00
2.57	120.82	0.00	603.96	980.49	0.00



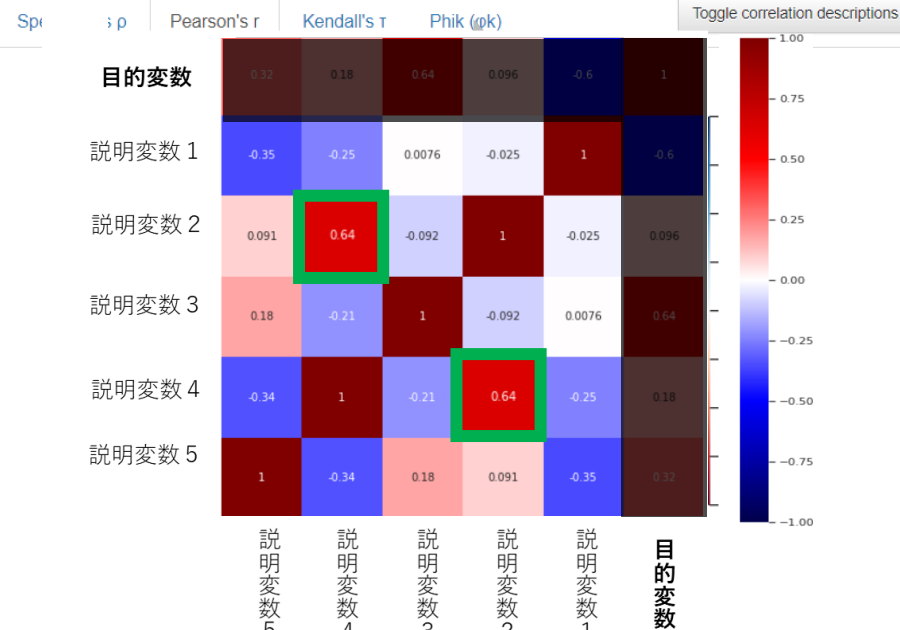
# データ理解の目的とそれに必要な作業: 相関関係

## 相関関係を確認する

説明変数同士の相関を確認する

説明変数間で相関が強すぎると**多重共線性\***が出る  
(相関が強い説明変数同士では微小変化に過敏に  
反応、予測が不安定になる)。

### Correlations



説明変数に多重共線性があると、その説明変数の係数値  
(目的変数に及ぼす影響力) が定まらない！

直感的な解釈：鶴亀算が連立方程式で解けるのは、足の本数が異なるから（独立）

$$\begin{aligned} x + y &= 20 \\ 4x + 2y &= 54 \end{aligned}$$

$x = 7$  かめは 7 匹



足が4本



足が2本

$$\begin{aligned} x + y &= 20 \\ 4x + 2y &= 54 \end{aligned}$$

$x, y$ の組み合わせは沢山ある。  
変数が独立していないと、  
値は一意に定まらない。



足が2本

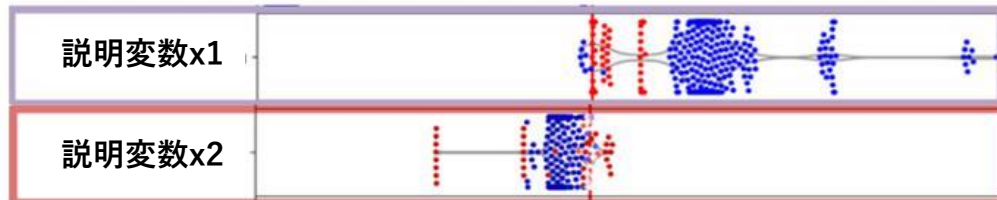


足が2本

# • 【参考スライド】 多重共線性の見つけ方 その②

- 複数回モデルを作り、それぞれのSHAP値(or FI値)を見比べて、毎回変動している説明変数を見つける

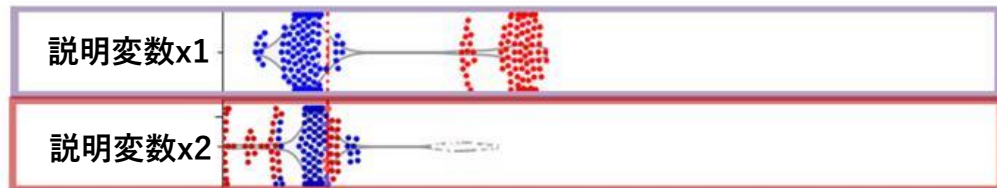
1回目のモデルのSHAP値



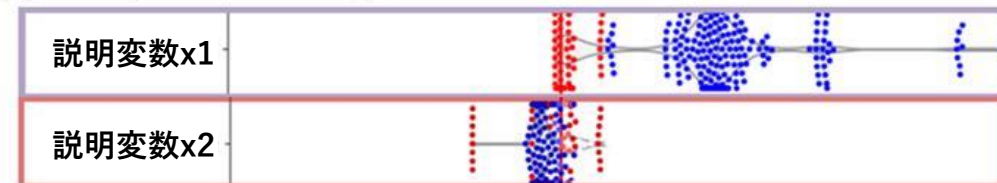
2回目のモデルのSHAP値



3回目のモデルのSHAP値



4回目のモデルのSHAP値

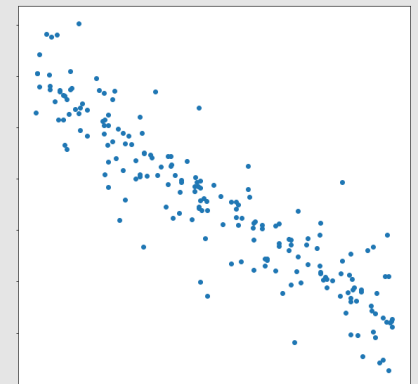


モデルを構築する毎に、 $x_1$ と $x_2$ の値が変動している

この不安定さは、多重共線性が原因と考えられる  
どちらか1つを削除するか、統合するかして、  
再度モデル構築すると良い

多重共線性にある変数の係数値はトレードオフの傾向にある

$x_2$ の係数値



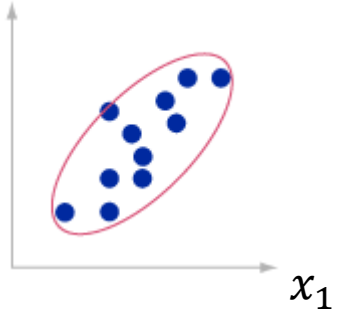
$x_1$ と $x_2$ が多重共線性にあるとして、  
 $x_1$ の係数が大きな値になった時、  
 $x_2$ は逆に小さくなる。  
なので全体（目的変数の予測値）としては、  
大きくズレる事はない。

$x_1$

- 【参考スライド】 多重共線性があると発生する弊害
- 多重共線性がある説明変数のペアは、係数「b」の値が不安定になる <sub>$x_2$</sub>

機械学習モデル（重回帰モデル）

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$



1回目	2回目	3回目	4回目	5回目
0.36	1.02	0.06	20.6	10.5

**x1の係数b1が定まらない！  
(b2も同様)**

係数bの値が不安定になる理由

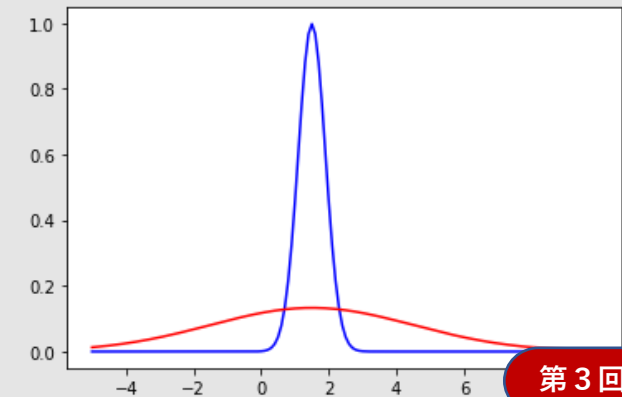
$$y = b_0 + b_1 x_1 + b_2 x_2 + e$$

$$b_1 \text{の分散} = \frac{\sigma^2}{\sum x_1^2 (1 - \underline{\underline{r_{12}}})}$$

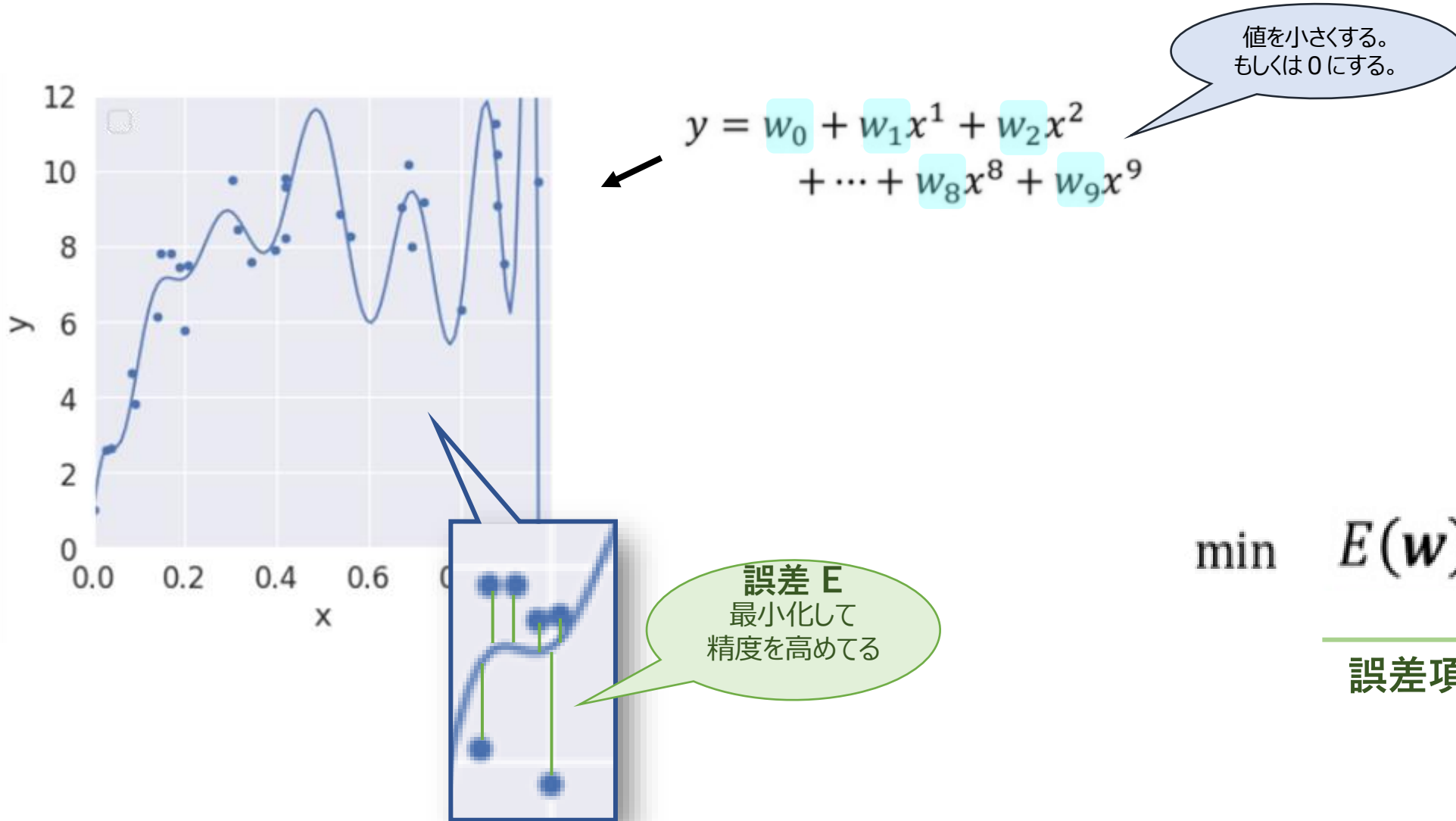
$x_1$ と $x_2$ の相関

分散が大きいと  
b1の取りうる  
範囲も広がる

$b_1$ の分散(多重共線性が**無い場合**と**ある場合**)



- 【参考スライド】 過学習回避の方法その②
- 過学習を抑制する手法の1例 - Lasso回帰

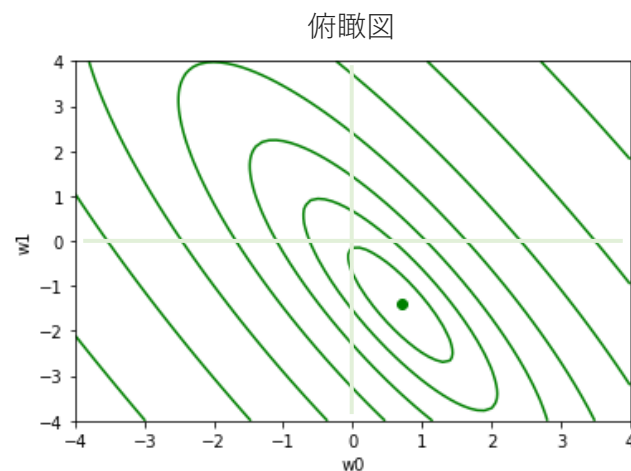
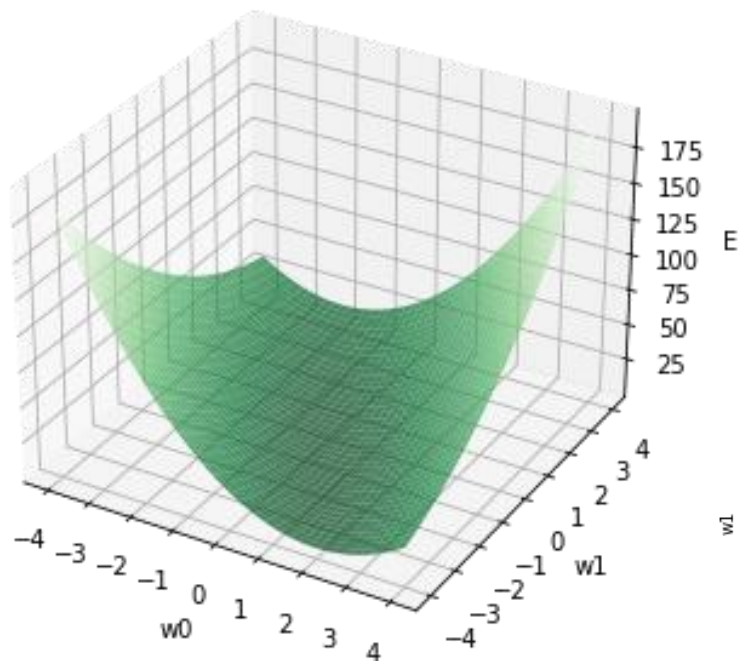


$$\min \underbrace{E(\mathbf{w})}_{\text{誤差項}} + \underbrace{\frac{\alpha}{2} \sum_{i=0}^M |w_i|^q}_{\text{正則化項}}$$

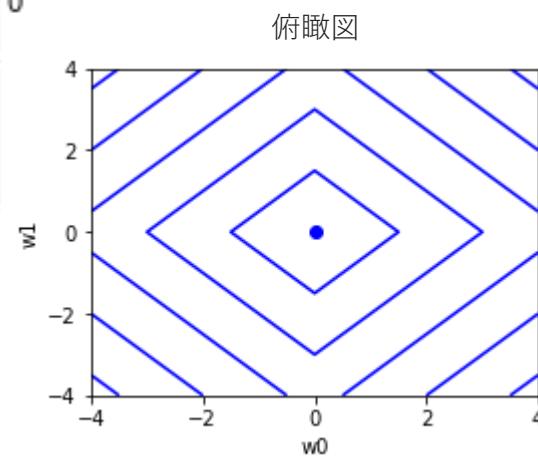
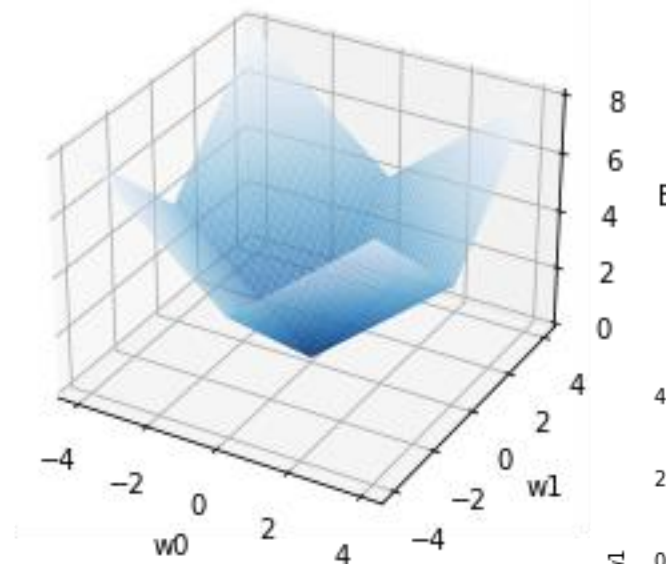
- 【参考スライド】 過学習回避の方法その②
- Lasso回帰

$$\min \underbrace{E(\mathbf{w})}_{\text{誤差項}} + \underbrace{\frac{\alpha}{2} \sum_{i=0}^M |w_i|^q}_{\text{正則化項}}$$

係数値の組合せ (w0,w1) に対する誤差Eの値



係数値の組合せ (w0,w1) に対する誤差Eの領域

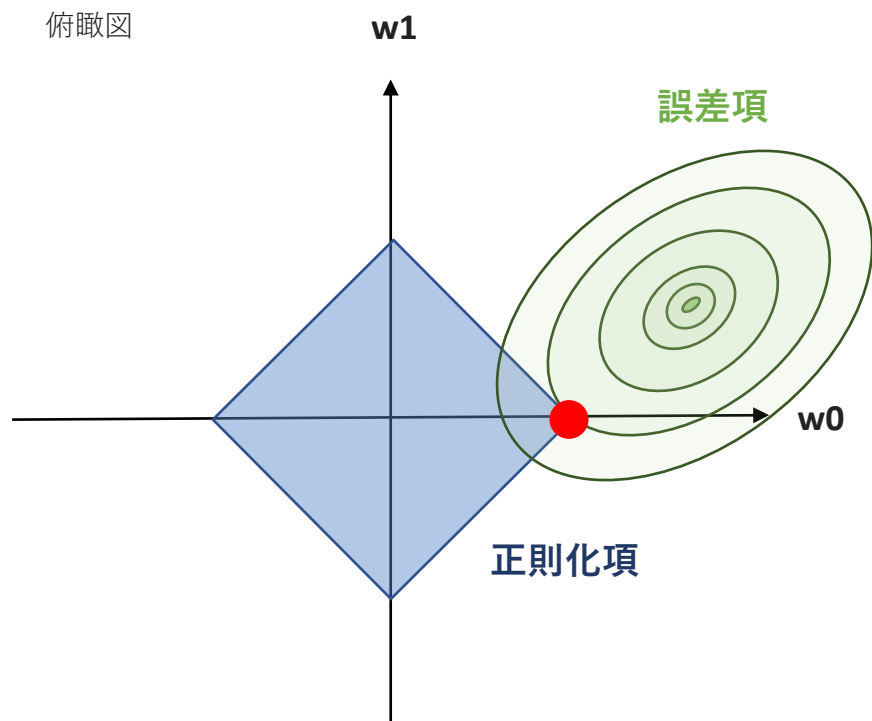


- 【参考スライド】 過学習回避の方法その②
- Lassoの他にもRidgeがある

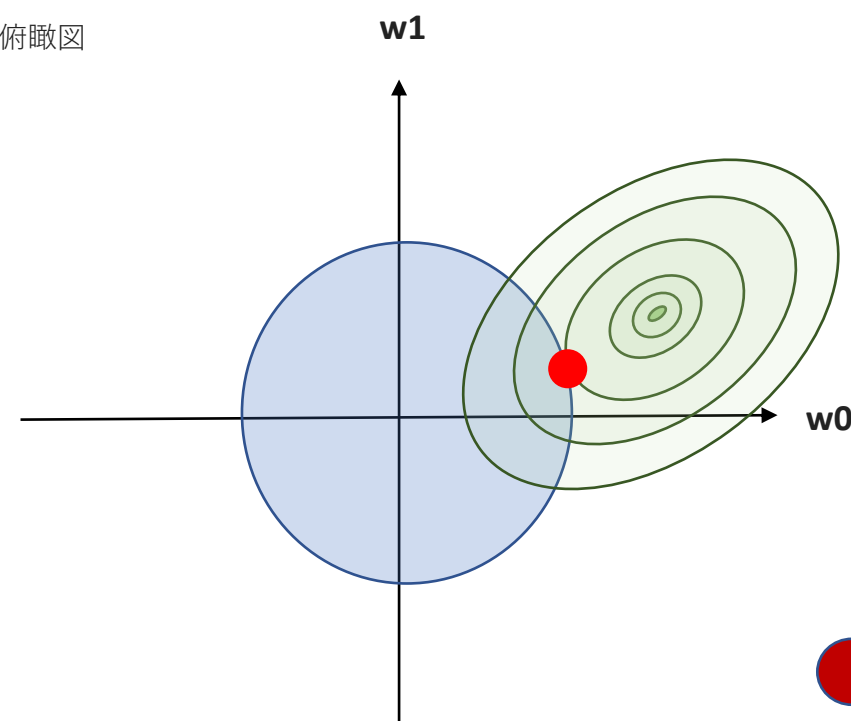
$$\min \underbrace{E(\mathbf{w})}_{\text{誤差項}} + \underbrace{\frac{\alpha}{2} \sum_{i=0}^M |w_i|^q}_{\text{正則化項}}$$

$$\min \underbrace{E(\mathbf{w})}_{\text{誤差項}} + \underbrace{\frac{\alpha}{2} \sum_{i=0}^M (w_i)^2}_{\text{正則化項}}$$

俯瞰図



俯瞰図



- MI解析OJTの成果（中澤担当分、一部抜粋）

目的	解析パターン	結果
割れ難く痕が付きにくい、折り曲げ用フィルムの製造条件を検討	組成最適化+製造プロセス検討	新知見を獲得、検討継続
環境にやさしい素材を使い(非強化)、長期耐熱に優れたプラスチック材料の開発	組成最適化	チャンピオンデータを獲得。検討継続
某装置の性能予測を行うモデルの作成。	製造プロセス検討(サロゲートモデル)	予期していた知見を獲得、課題解決の見通しが立った。部分的に制度が低い為、データ蓄積を継続。