

教育研修・中規模データ分析のための
機械学習ソフトウェア
JUSE-StatWorks/V5 機械学習編
の機能と使い方

(株) 日本科学技術研修所
数理事業部 犬伏秀生

発表内容

1. JUSE-StatWorks/V5機械学習編の概要

- ① StatWorks/V5機械学習編とは
- ② StatWorks/V5機械学習編の特徴
- ③ 搭載手法

2. デモンストレーション

- ① lasso回帰（正則化回帰）
- ② サポートベクターマシン（SVM）
（ランダムフォレスト）

3. 今後の予定

JUSE-StatWorks/V5機械学習編とは

- 「JUSE-StatWorks/V5機械学習編」は、弊社が開発・販売を行っている統計パッケージJUSE-StatWorks/V5シリーズに今後新規に追加する、機械学習を実行できるパッケージとなります。



“実務に役立つ”進化を遂げたデータ解析ツール

JUSE-StatWorks®/V5

既存パッケージ

総合編プレミアム

総合編 with MT

総合編 with SEM

総合編

品質管理編

QC七つ道具編

SEM因果分析編

品質工学編

新規追加

機械学習編

機械学習編の特徴

インターゲ ット	・ 製造業で品質管理に携わる実務者
主な使用用途	1. 機械学習の教育・研修 2. 中規模以下のデータの分析 3. 大規模データを予測・判別するためのモデルの検討
動作環境	・ クライアントPC（64bitOS）上で動作
機能の特徴	・ 画面構成や操作性はStatWorks/V5の既存製品と共通 ・ SQCで使用する多変量解析手法との対比で理解可能な機械学習手法を搭載 ・ 分析可能なデータサイズは1,000変数×100,000サンプル以内

搭載手法

No	グループ	解析手法
1	データクリーニング	データクリーニング
2	データ可視化	濃淡散布図・密度プロット・等高線図
3	情報要約	カーネル主成分分析
4	層別	混合ガウス分布
5	正則化回帰	リッジ回帰
6		lasso回帰
7		Elastic Net
8	分類・予測	サポートベクターマシン(SVM)
9		ランダム・フォレスト
10	外れ値検出	1クラスSVM
11	因果分析	glasso



データ概要 (lasso回帰)

	データ概要
目的	超電導物質の臨界温度を予測する
変数	【目的変数：1変数】 - 超電導物質の臨界温度（量的変数） 解析では常用対数（log10）を取った値を使用 【説明変数：150変数】 - 超電導質の特徴量（81変数：全て量的変数）， - 元素の構成比率（69変数：全て量的変数） ※オリジナルデータの変数の数は167（特徴量：81＋構成比：86）であるが、ここでは分散が0ではない150変数を使用
サンプル数	- 学習データ：200 - 検証用データ：58 ※オリジナルデータのサンプル数は21,263であるが、そこから学習データとして200サンプル、検証用データとして100サンプルをランダムに抽出。ただし、検証用データの外挿となるサンプルは除外
出典	[1]The UCI Machine Learning Repository https://archive.ics.uci.edu/ml/datasets/Superconductivity+Data

検証用データの予測結果

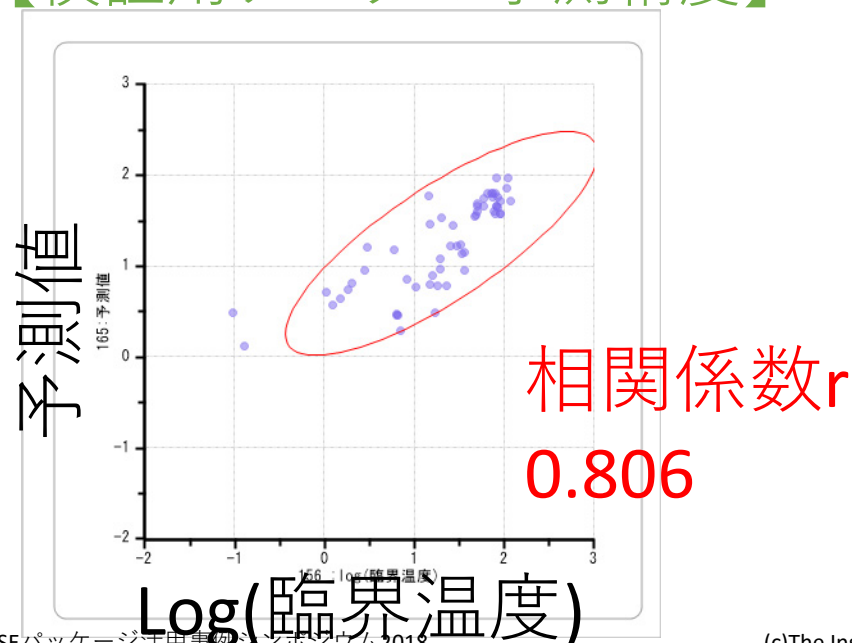
解析に使用した説明変数：150変数

lasso回帰

【回帰式】

取り込まれた変数：30変数
正則化パラメータ λ ：0.05248

【検証用データの予測精度】

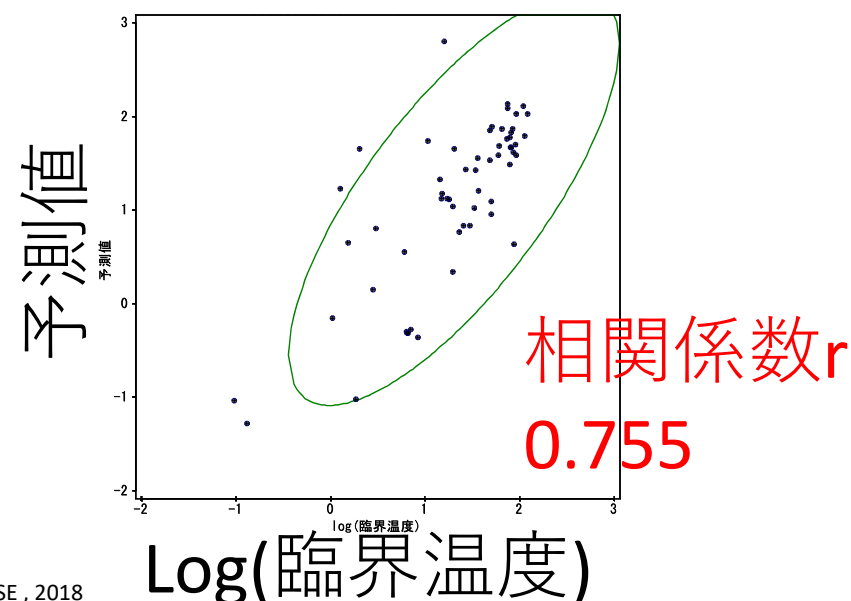


重回帰分析

【回帰式】

取り込まれた変数：105変数
変数選択方法：変数減増法

【検証用データの予測精度】



参考：21,163サンプルの結果

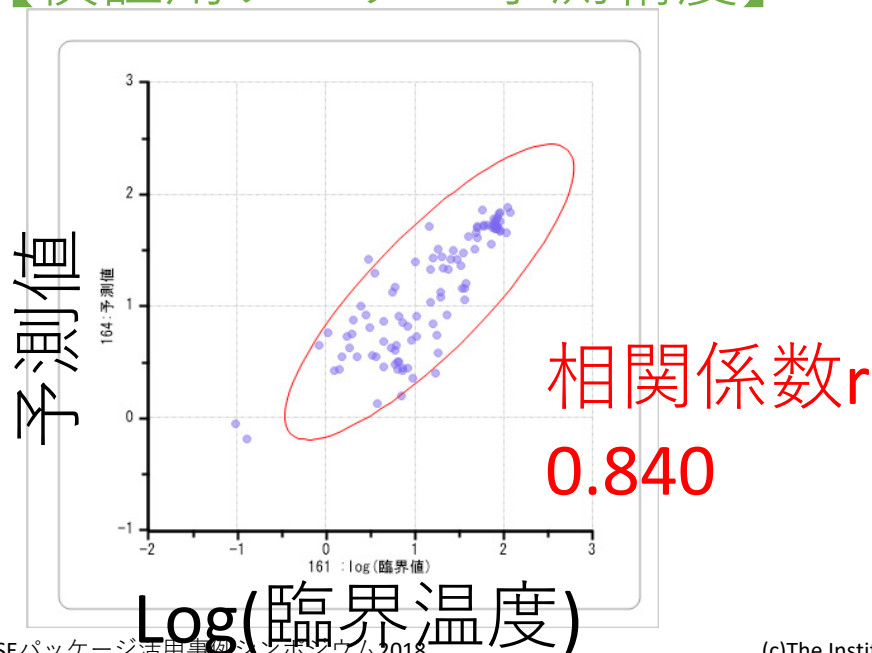
解析に使用した説明変数：158変数

lasso回帰

【回帰式】

取り込まれた変数：21変数
正則化パラメータ λ ：0.07943

【検証用データの予測精度】

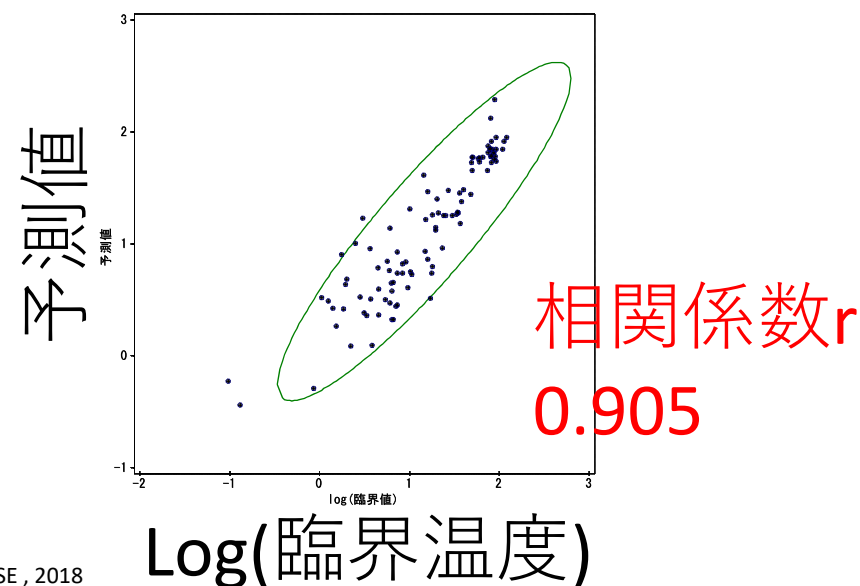


重回帰分析

【回帰式】

取り込まれた変数：118変数
変数選択方法：変数減増法

【検証用データの予測精度】



データ概要 (SVM)

	データ概要
目的	犯罪現場に落ちているガラスの種類を判別する
変数	【目的変数：1変数】 ・ ガラスの種類（質的変数） ≪カテゴリ≫車の窓，ヘッドランプ 【説明変数：9変数】 ・ 屈折率（量的変数）， ・ 元素の構成比（8変数：全て量的変数）
サンプル数	・ 学習データ：40 （車の窓：14，ヘッドランプ：26） ・ 検証用データ：6 （車の窓：3，ヘッドランプ：3） ※オリジナルデータでは目的変数は6カテゴリであるが，そこから上記2カテゴリを抽出。また，それぞれのカテゴリの中の3サンプルを検証用データとして使用
出典	[1]The UCI Machine Learning Repository https://archive.ics.uci.edu/ml/datasets/glass+identification

今後の予定

- StatWorks/V5機械学習編は，2019年春に発売開始予定です.
- 弊社主催の機械学習セミナーも2019年度上期から継続的に開催予定です.
- StatWorks/V5有償保守契約者様でご希望の方には2019年1月から機械学習編のモニター版を無償で貸出を行う予定です.
- StatWorks/V5機械学習編に関する最新情報は弊社HP上で随時公開します.

ご清聴ありがとうございました

本著作物は原作者の許可を得て、株式会社日本科学技術研修所（以下弊社）が掲載しています。本著作物の著作権については、制作した原作者に帰属します。

原作者および弊社の許可なく営利・非営利・イントラネットを問わず、本著作物の複製・転用・販売等を禁止します。

所属および役職等は、公開当時のものです。

■公開資料ページ

弊社ウェブページで各種資料をご覧ください <http://www.i-juse.co.jp/statistics/jirei/>

■お問い合わせ先

(株)日科技研 数理事業部 パッケージサポート係 <http://www.i-juse.co.jp/statistics/support/contact.html>