

MATLAB EXPO

機械学習：
“結果”を出した適用例と最新機能紹介

吉野 紘和



どのように機械学習を始めたら良いのでしょうか？

🔍 get started with machine learning|

About 611,000,000 results (0.63 seconds)

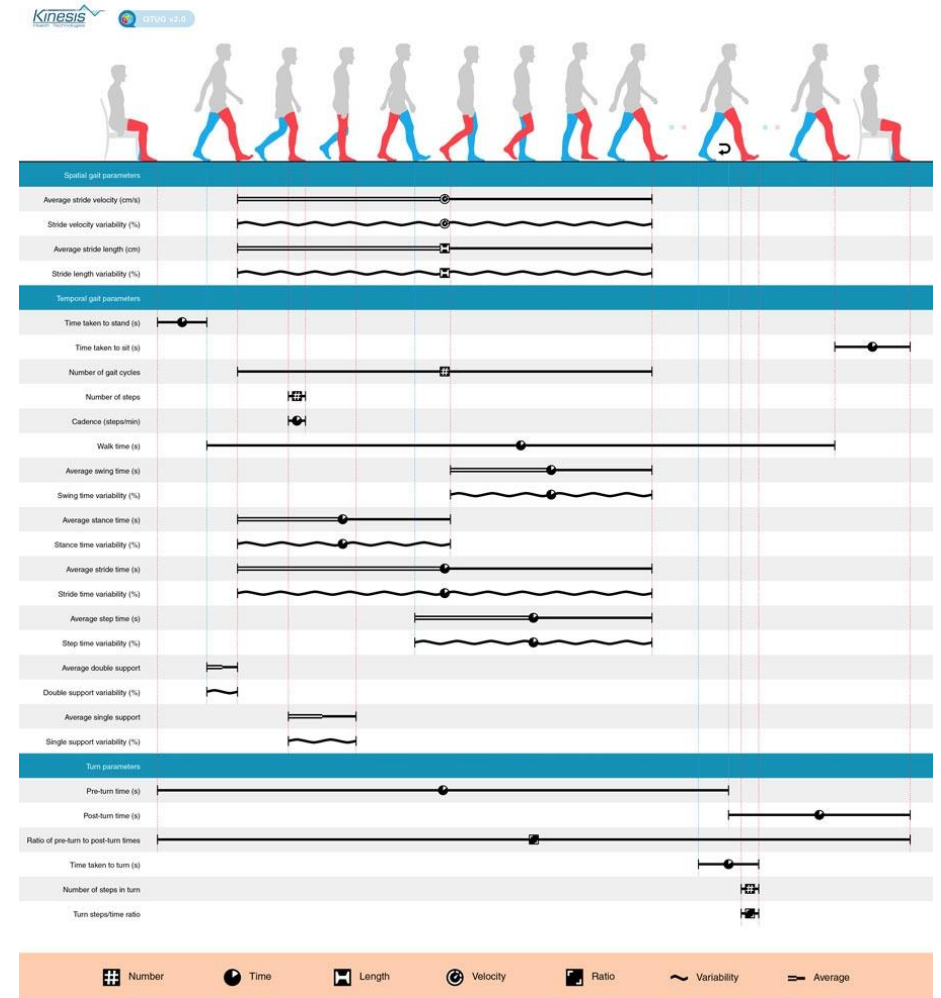
機械学習の成功例

Kinesis Health Technologies

機械学習を用いて患者の転倒リスクを予測



[詳細はこちら](#)



機械學習

+

X

機械学習

+

業界のノウハウ

アプリケーションの専門知識

個人のノウハウ

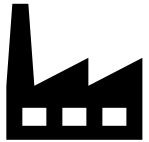
“結果”を出した機械学習のアプリケーション例



フリートデータ解析



電力予測



製造アナリティクス

新機能

- MATLAB® アプリ
- AutoML
- 機械学習を用いた信号処理
- C/C++ コード生成

“結果”を出した機械学習のアプリケーション例



フリートデータ解析



電力予測

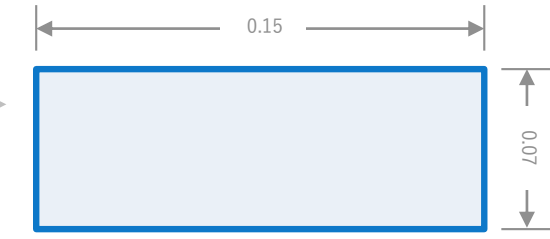


製造アナリティクス

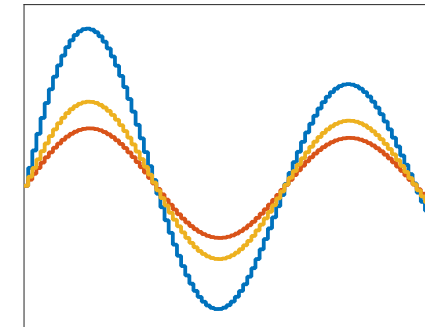
フリートデータ解析

機械学習モデル学習

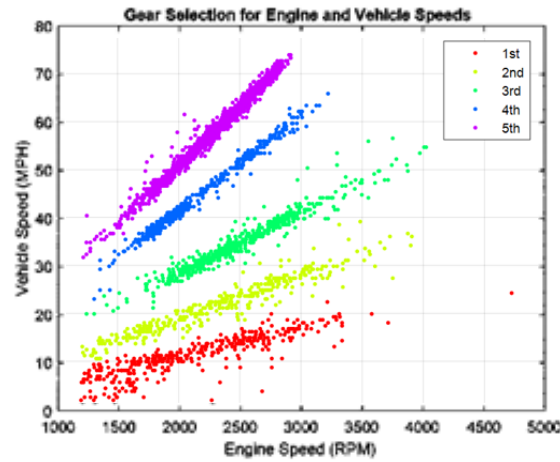
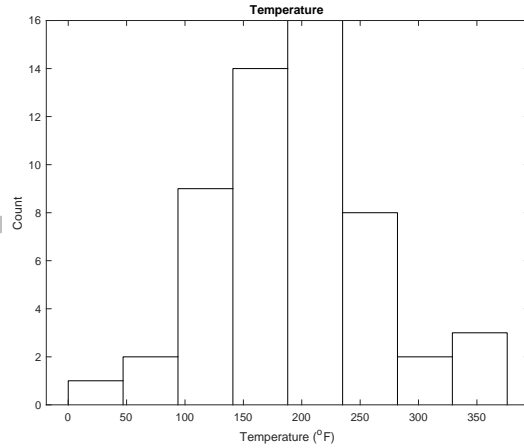
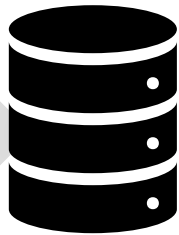
設計決定



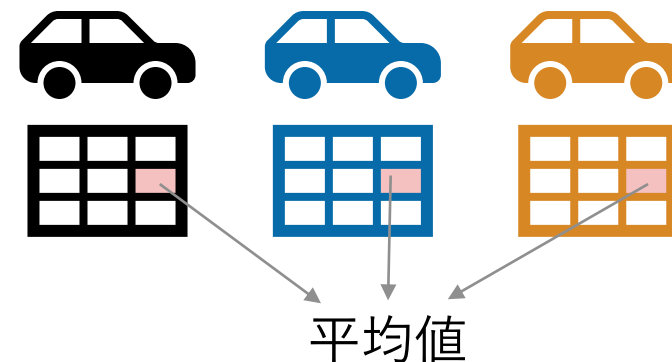
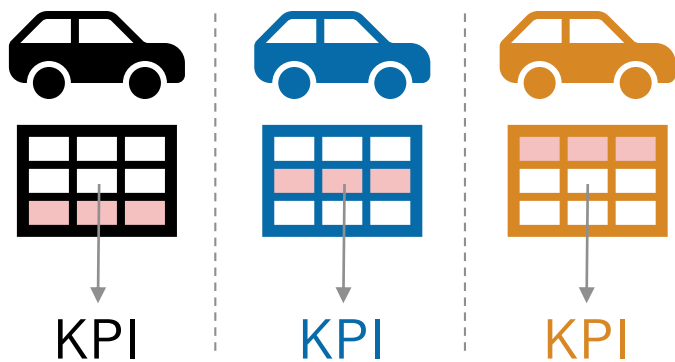
試験計画



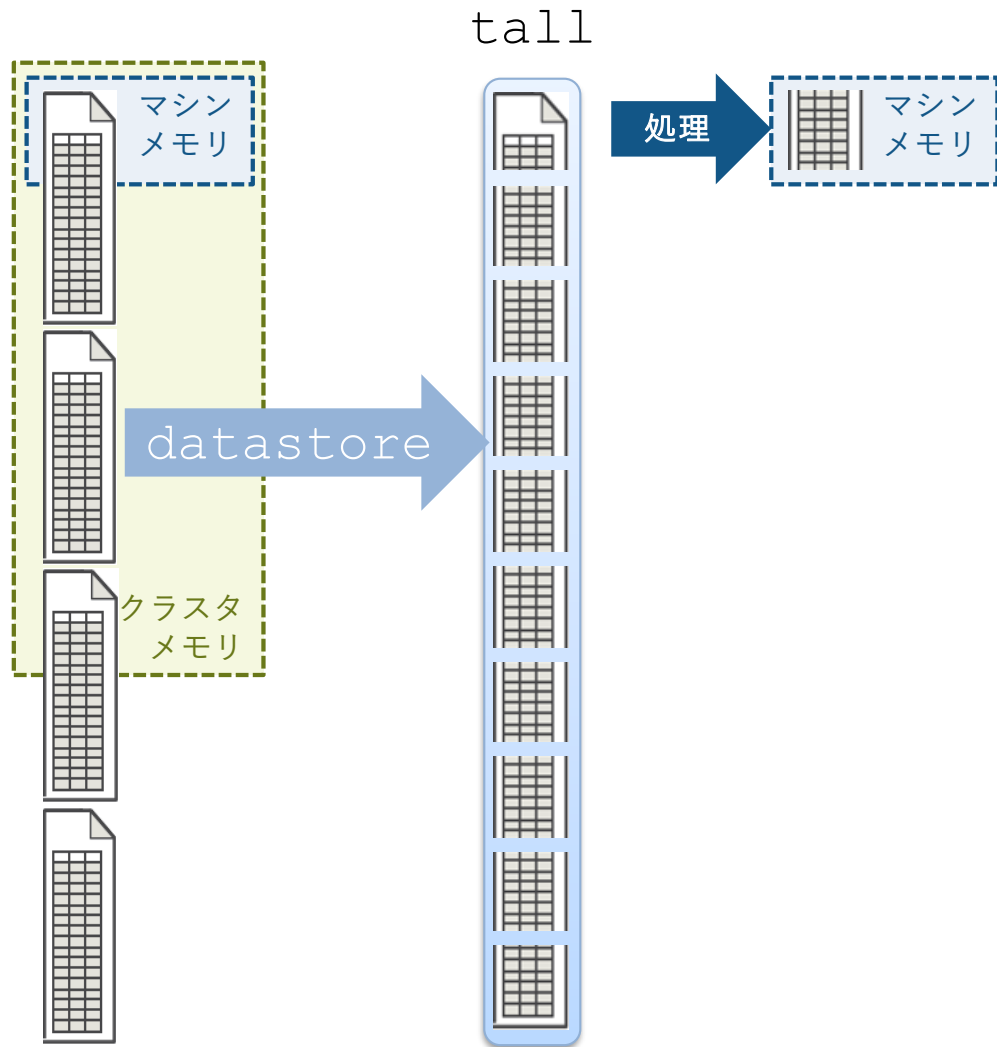
フリート



問題のタイプ?



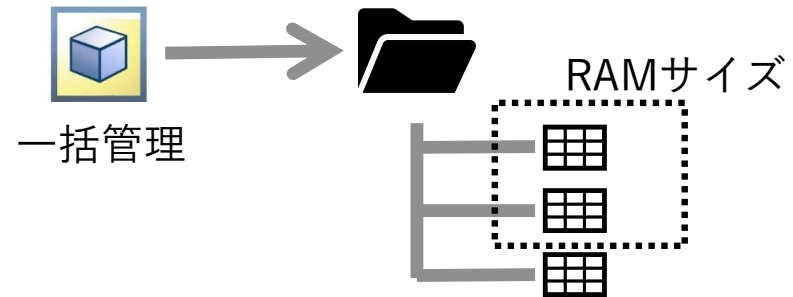
メモリに収まらないデータ取り扱いのためのテクニック



datastore型変数

メモリに収まらないデータの集まりのリポジトリ

```
ds = datastore('*.*csv')
```



tall配列

datastoreに格納されたデータを操作する

```
t = tall(ds)
```

datastoreで大きなデータ集合を扱う

必要な項目（センサー等）の選択

```
varnames = {'Time', 'EngineSpeedRPM'}
```

選択項目についてのdatastore型変数の作成

```
ds = datastore('EngineData*.csv', ...  
              'SelectedVariableNames', varnames)
```

“メモリに乘る”サイズのデータの読み込み

```
data = read(ds)
```



利用可能な datastore 一覧	
一般	datastore
	spreadsheetDatastore
	tabularTextDatastore
	fileDatastore
データベース	databaseDatastore
画像	imageDatastore
	denoisingImageDatastore
	randomPatchExtractionDatastore
	pixelLabelDatastore
	augmentedImageDatastore
音声	audioDatastore
予知保全	fileEnsembleDatastore
	simulationEnsembleDatastore
Simulink	SimulationDatastore
自動車	mdfDatastore
カスタム	subclass matlab.io.Datastore
変換	transform 存在するdatastoreの変換

“全体”の計算はtallで実行

1個のファイル

データにアクセス

```
data = readtable("EngineData1.csv");  
data = table2timetable(data);
```

データの前処理

興味のあるデータの選択

```
data = data(:, "EngineSpeedRPM");
```

欠損データ処理

```
data = fillmissing(data, "linear");
```

統計量計算

```
m = mean(data.EngineSpeedRPM);  
s = std(data.EngineSpeedRPM);
```

1,000個のファイル

データにアクセス

```
data = datastore("EngineData*.csv");  
data = tall(data);  
data = table2timetable(data);
```

データの前処理

興味のあるデータの選択

```
data = data(:, "EngineSpeedRPM");
```

欠損データ処理

```
data = fillmissing(data, "linear");
```

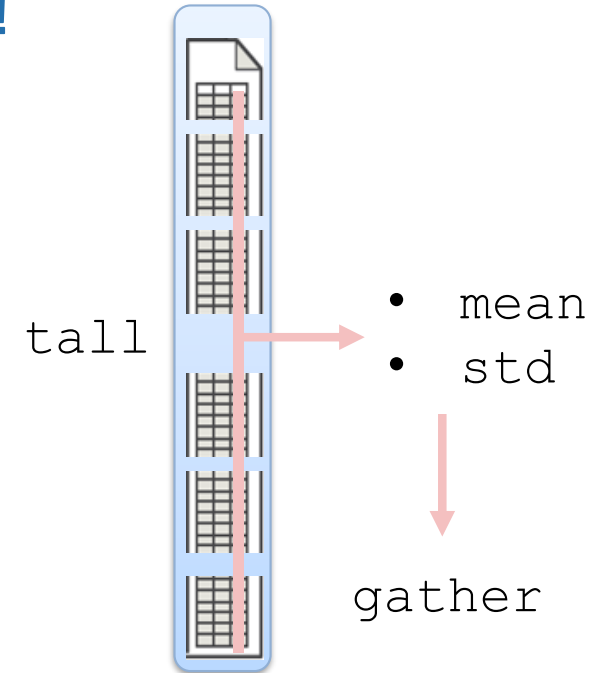
統計量計算

```
m = mean(data.EngineSpeedRPM);  
s = std(data.EngineSpeedRPM);
```

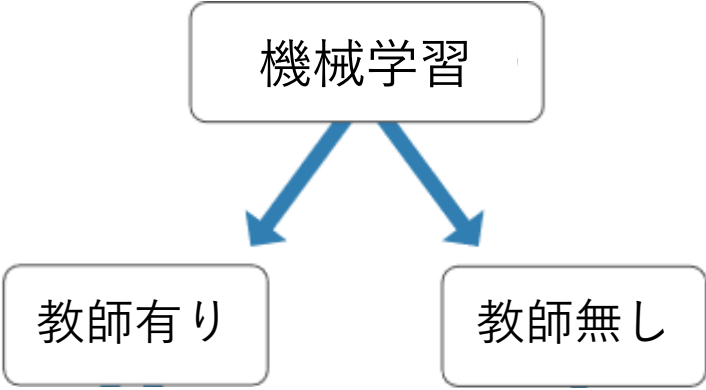
- 平均
- 分散

```
[m,s] = gather(m,s);
```

2行付け加えるだけ!



教師無し学習手法でフリーデータを調査



オペレーションモードクラスタリングのための教師無し学習

生データをプロット

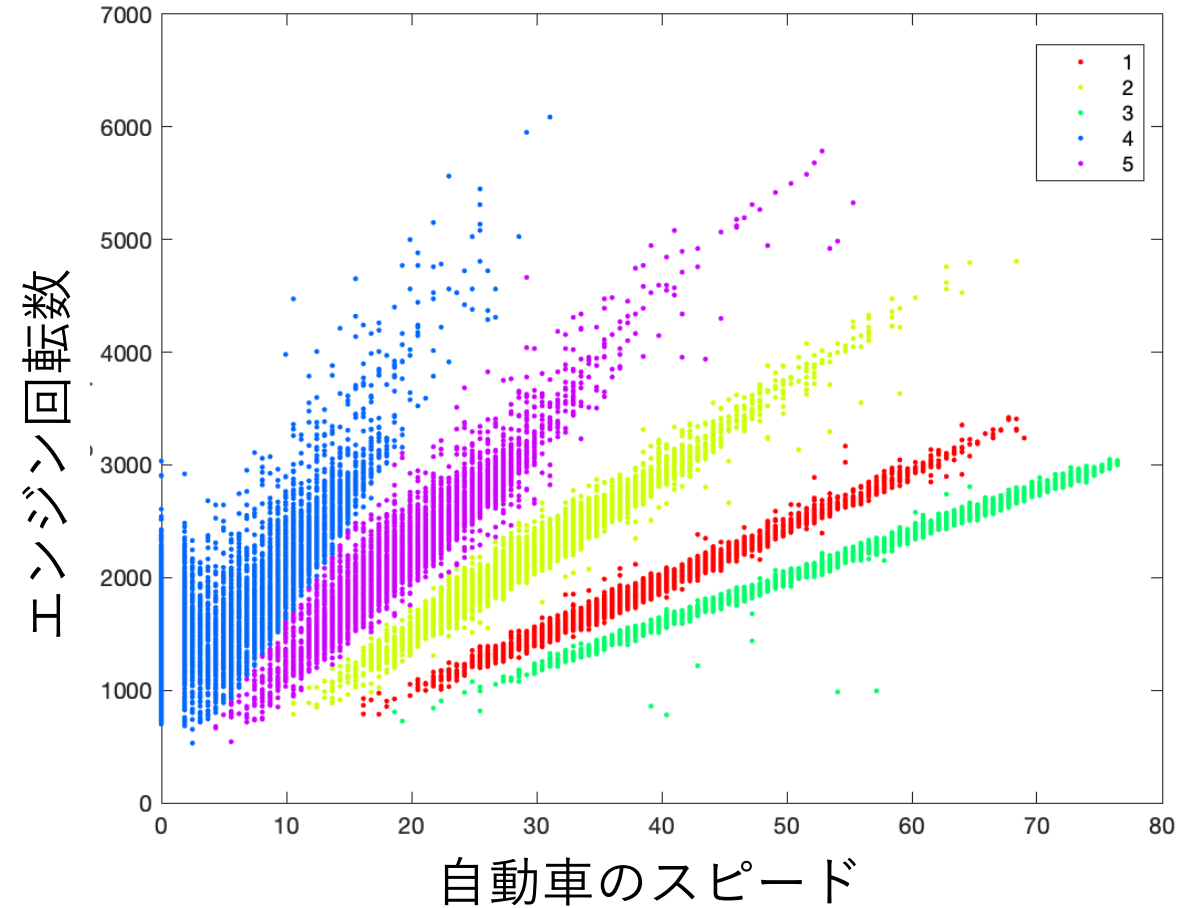
```
figure;  
plot(t.Speed_OBD_,t.EngineRPM, '.k')  
xlabel('Vehicle Speed');  
ylabel('Engine Speed');
```

K-meansアルゴリズムでクラスタリング

```
X = [t.Speed_OBD_,t.EngineRPM];  
IDX = kmeans(X,5,"Distance","cosine");
```

クラスタリングの結果をプロット

```
gscatter(t.Speed_OBD_,t.EngineRPM,IDX);  
xlabel('Vehicle Speed');  
ylabel('Engine Speed');
```



オペレーションモードクラスタリングのための教師無し学習 (Cont.)

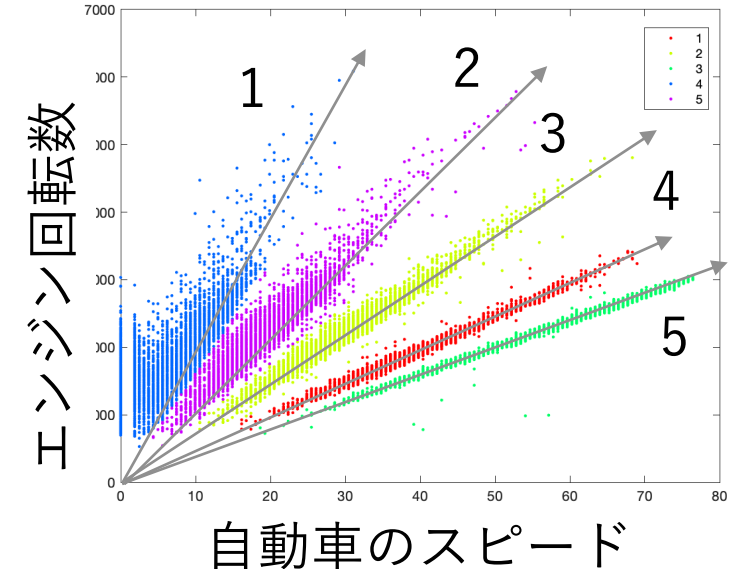
機械学習

+

X

```
D = [t.Speed_OBD_ , t.EngineRPM];  
IDX = kmeans(D, 5, "Distance", "cosine");
```

- **5** : ギアの数がある5つであるという装置のノウハウ
- **“cosine”** : 自動車のスピードとエンジン回転数は線形の関係であるという領域のノウハウ



フリート解析 - 実装

自動車データ,
運転手のプロフィール



“コールドストレージ”

過去のデータ:

- バッチ処理
- クラスタ上の大規模データ
- 長期トレンドの探索
- モデル構築

“ホットストレージ”

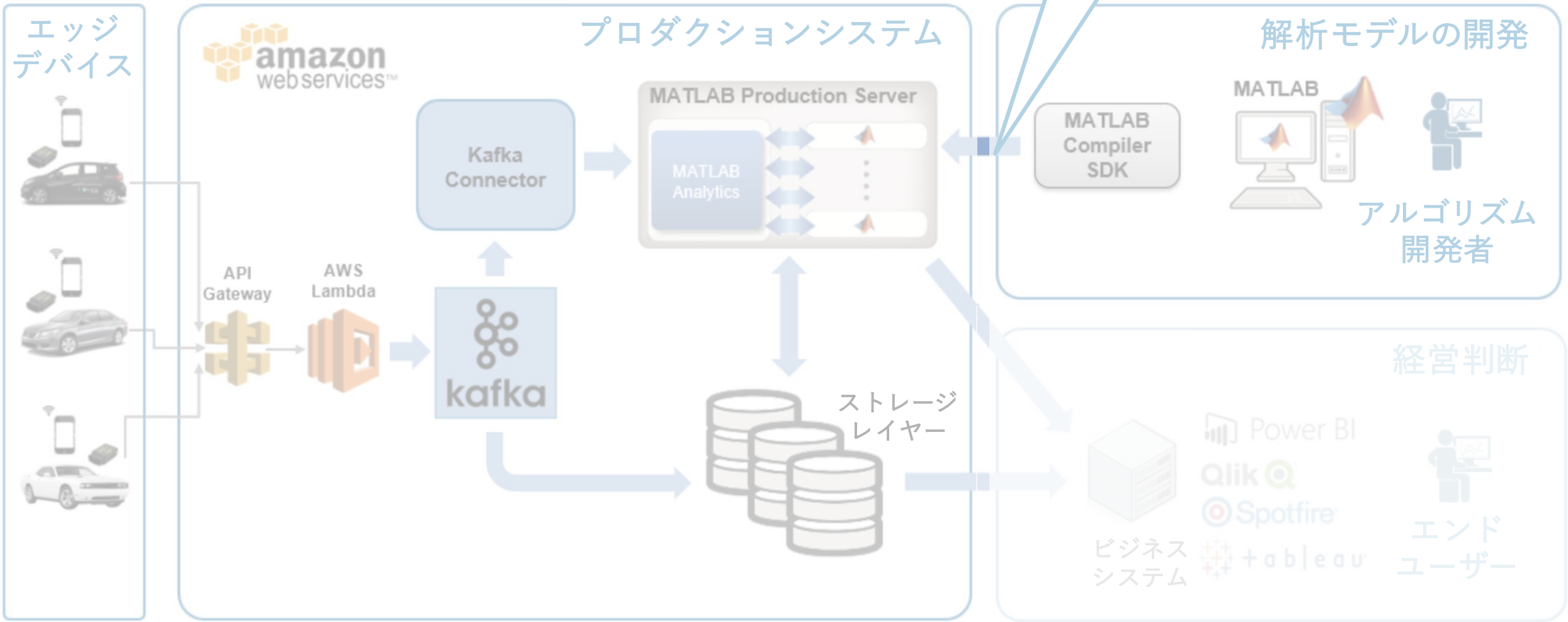
ストリーミングデータ:

- 準リアルタイム
- 新データ用にモデルのテストと実装
- ストリーミング処理



フリート解析 - ストリーミングアーキテクチャ

数クリック+
ドラッグ&
ドロップ



実際のフリート解析: Volkswagen Data Lab

個人向けの自動車の特徴とサービスを“仕立てる”

要素技術の開発

- 運転者と車両の安全
- 運転者へのコーチング
- 運転者固有の自動車保険

データソース

- CANバスデータと運転記録

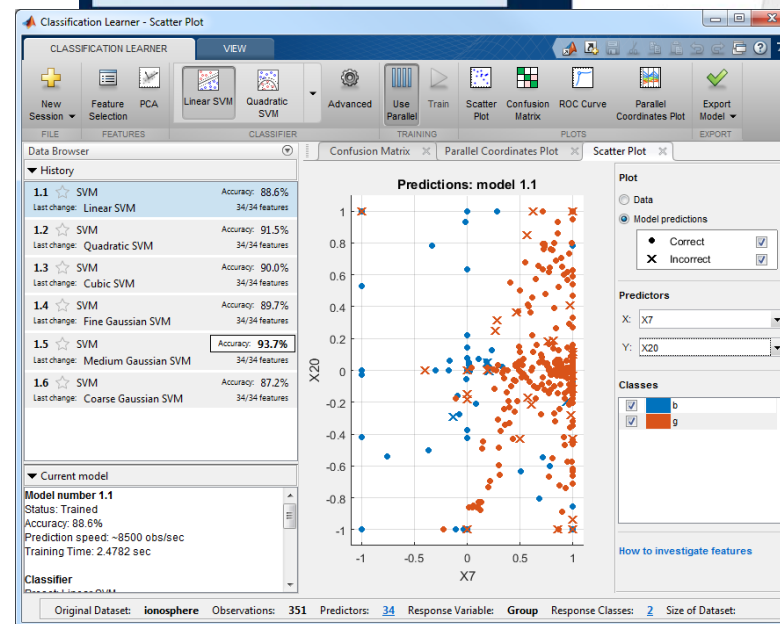
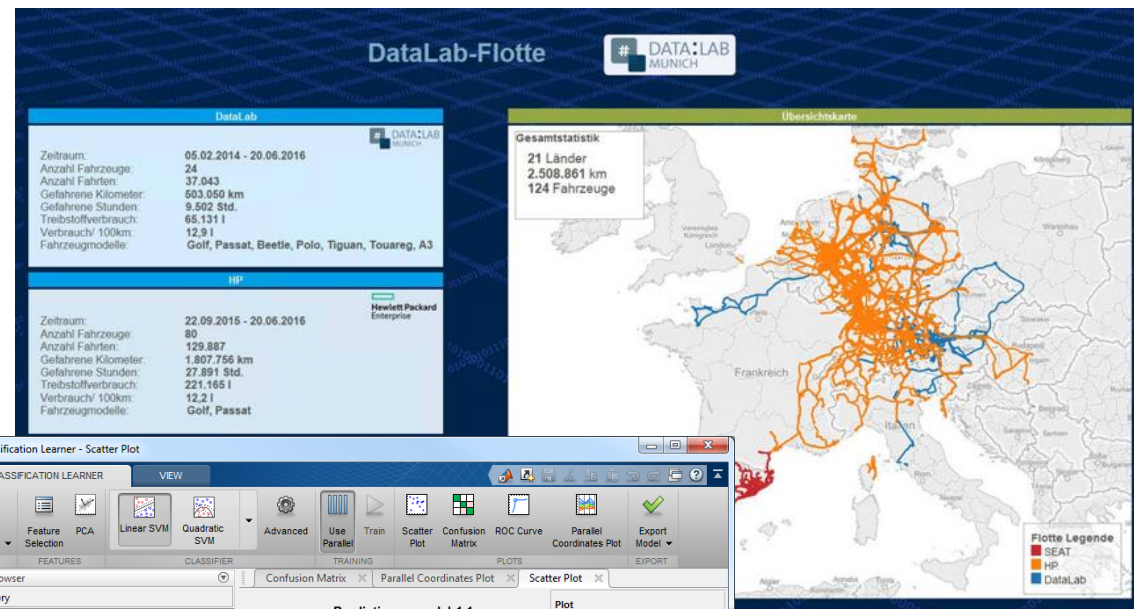
結果

- Telematic fingerprint (誰が自動車を運転したか)のためのPOCモデル
- “従量制ドライブ”の基本コンセプトの確立

Source: [“Connected Car – Fahrererkennung mit MATLAB”](#)

Julia Fumbarev, Volkswagen Data Lab

MATLAB EXPO Germany, June 27, 2017, Munich Germany



学習分類器アプリ

フリート解析

装置のノウハウ

設計仕様

オペレーションモード

オペレーション条件

機械学習

統計解析

教師無し学習

“結果”を出した機械学習のアプリケーション例



フリートデータ解析

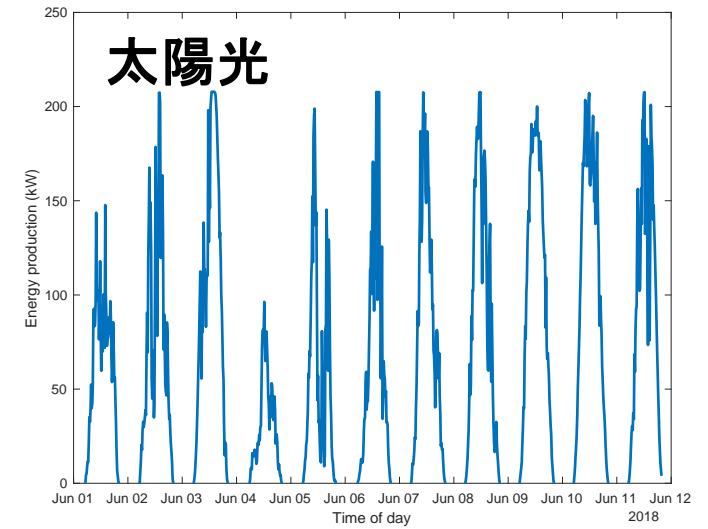
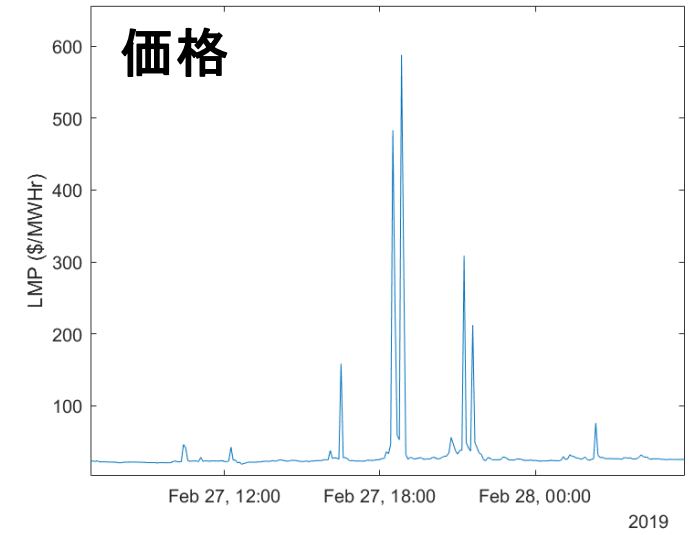
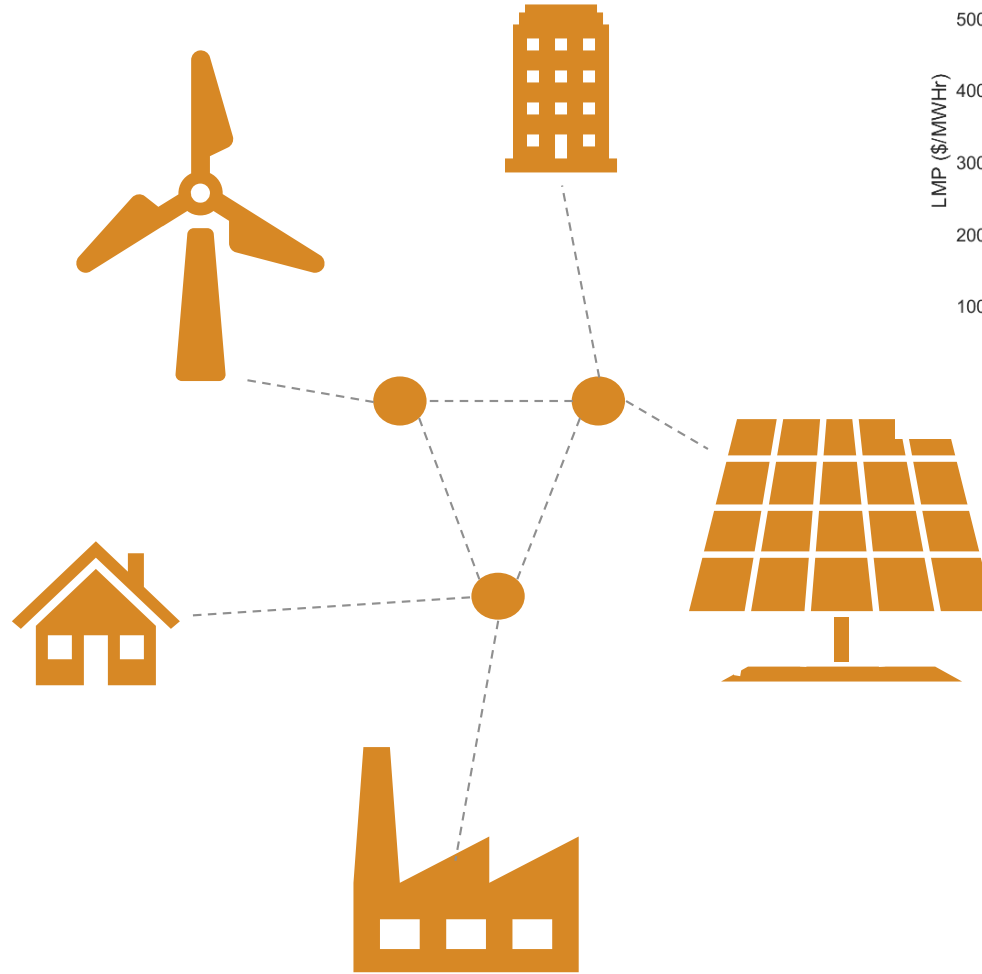
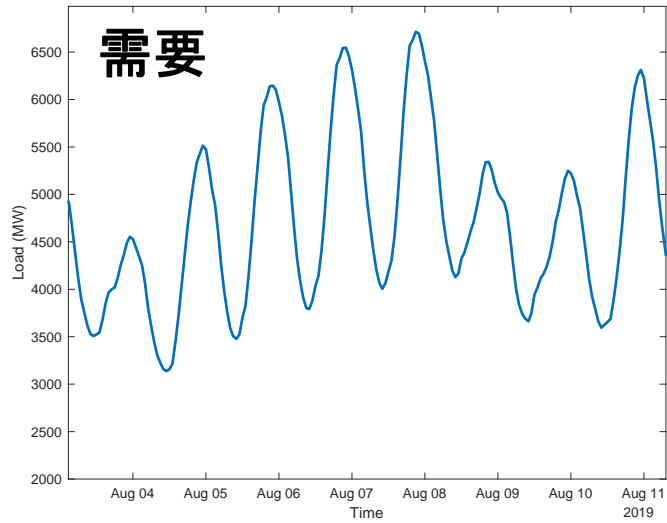
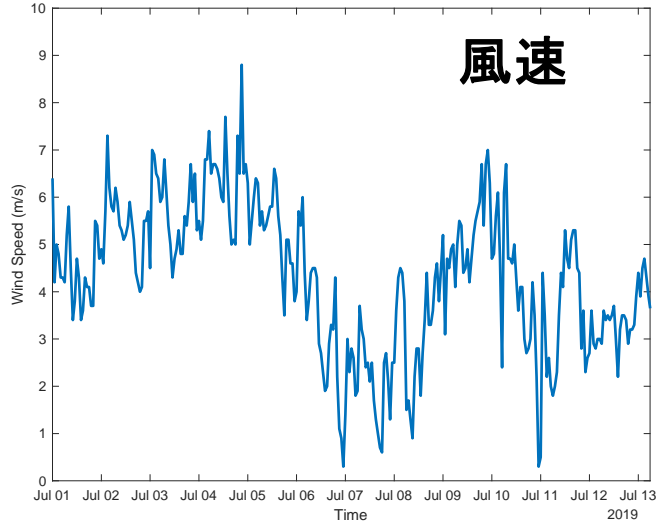


電力予測



製造アナリティクス

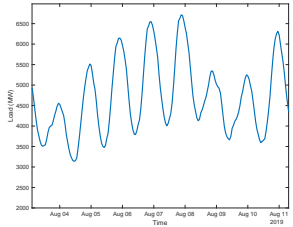
電力予測の必要性



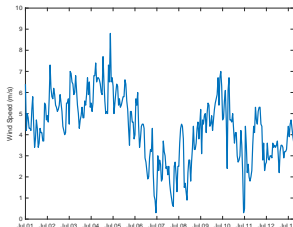
電力予測の流れ

過去のデータ

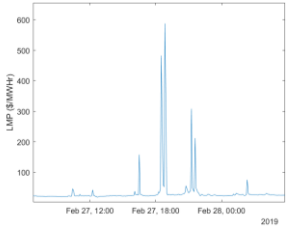
電力需要



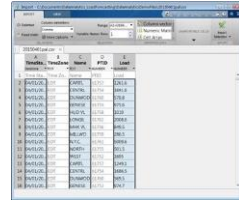
天候



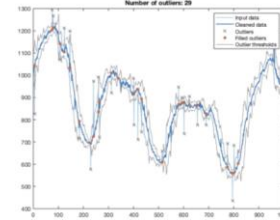
電力価格



結合



前処理

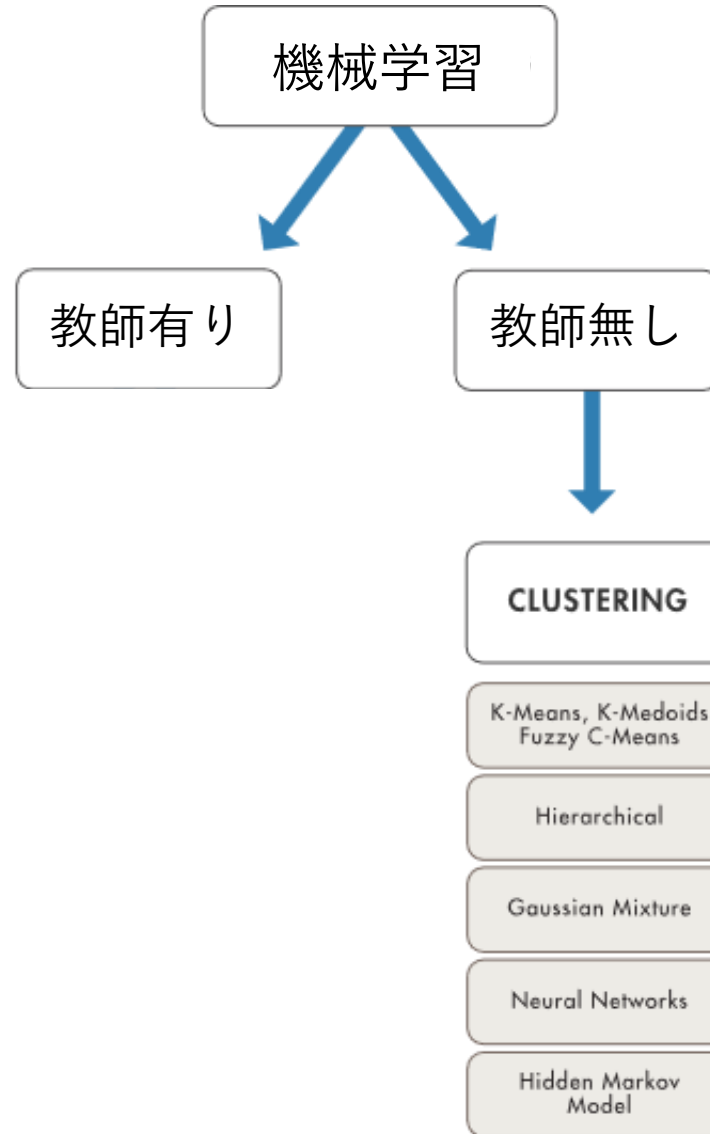


特徴量

load wind
temp 24hr
day 1week
month

機械学習

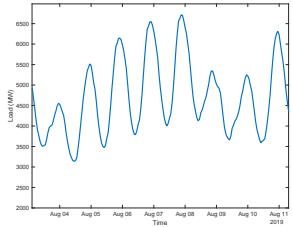
回帰手法で予測モデルを構築する



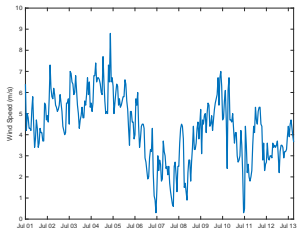
電力予測モデルの利用

新しいデータ

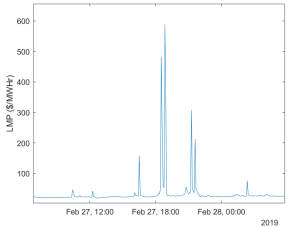
電力需要



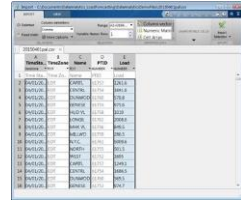
天候



電力価格



結合

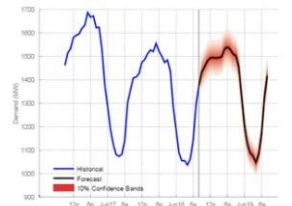


特徴量

load wind
temp 24hr
day 1week
month

学習済み
機械学習モデル

予測



時系列データの結合は synchronize 関数

結合

TimeSta	TimeZone	Name	PTID	Load
1				
2	04/01/2011	CAPITL	01757	1201.6
3	04/01/2011	CENTRE	01754	1691.6
4	04/01/2011	SUNWOOD	01760	579.8
5	04/01/2011	GENESE	01753	975.6
6	04/01/2011	HUD VS.	01758	1019
7	04/01/2011	LONGL	01762	2008.6
8	04/01/2011	MHK VS.	01756	849.5
9	04/01/2011	MILLAWD	01759	349.5
10	04/01/2011	NYC	01761	3009.6
11	04/01/2011	NORTH	01755	301.5
12	04/01/2011	WEST	01752	1695
13	04/01/2011	CAPITL	01757	1249.1
14	04/01/2011	CENTRE	01754	1695.5
15	04/01/2011	SUNWOOD	01760	569.5
16	04/01/2011	GENESE	01753	974.7

t1

Time
2015-11-15 00:00:24
2015-11-15 01:13:35
2015-11-15 01:30:24
2015-11-15 02:26:47
2015-11-15 03:00:24

Humidity_indoors	AirQuality
36	80
36	80
NaN	NaN
37	79
NaN	NaN

t2

Humidity_outdoors	TemperatureF	PressureHg
49	51.3	29.61
NaN	NaN	NaN
48.9	51.5	29.61
NaN	NaN	NaN
48.9	51.5	29.61

t3 = synchronize(t1, t2)



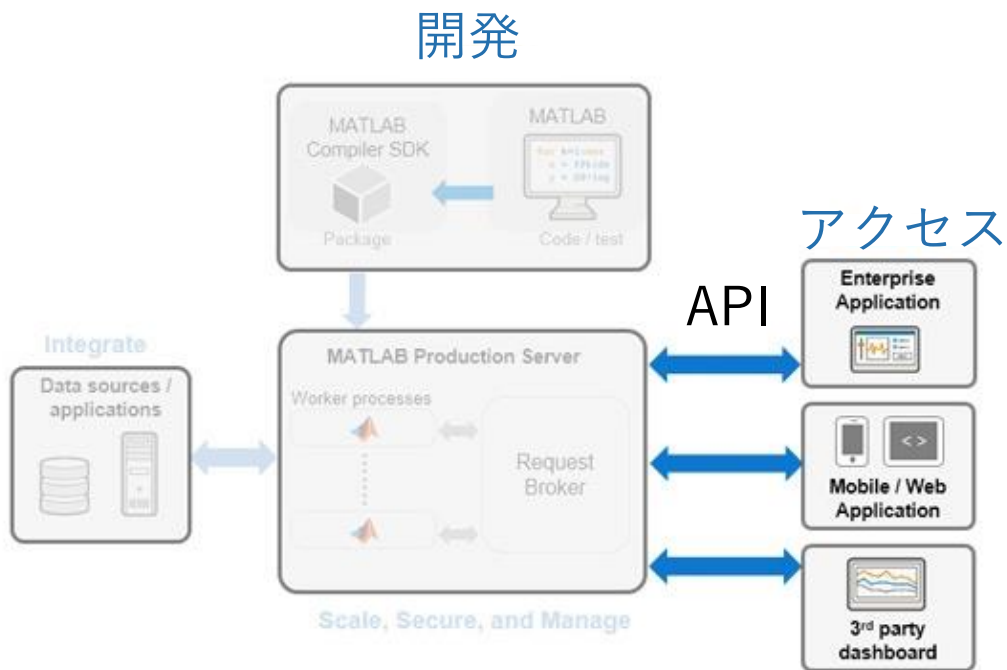
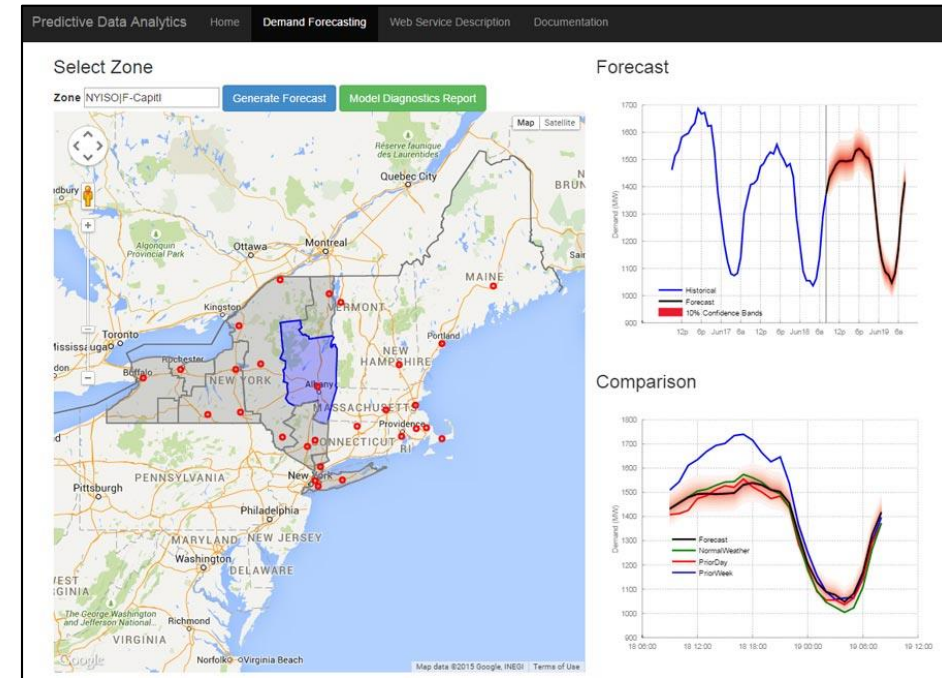
t3

Time
2015-11-15 00:00:24
2015-11-15 01:13:35
2015-11-15 01:30:24
2015-11-15 02:26:47
2015-11-15 03:00:24

Humidity_indoors	AirQuality	Humidity_outdoors	TemperatureF	PressureHg
36	80	49	51.3	29.61
36	80	48.919	51.463	29.61
36.23	79.77	48.9	51.5	29.61
37	79	48.9	51.5	29.61
37	80.378	48.9	51.5	29.61

電力予測モデルの展開

オペレーター & トレーダー用 ダッシュボード



アプリ開発者用API

予測を最適化問題と組み合わせる

“投資のリターンを最大化するために、いつ発電機を稼働させればいいだろうか?”

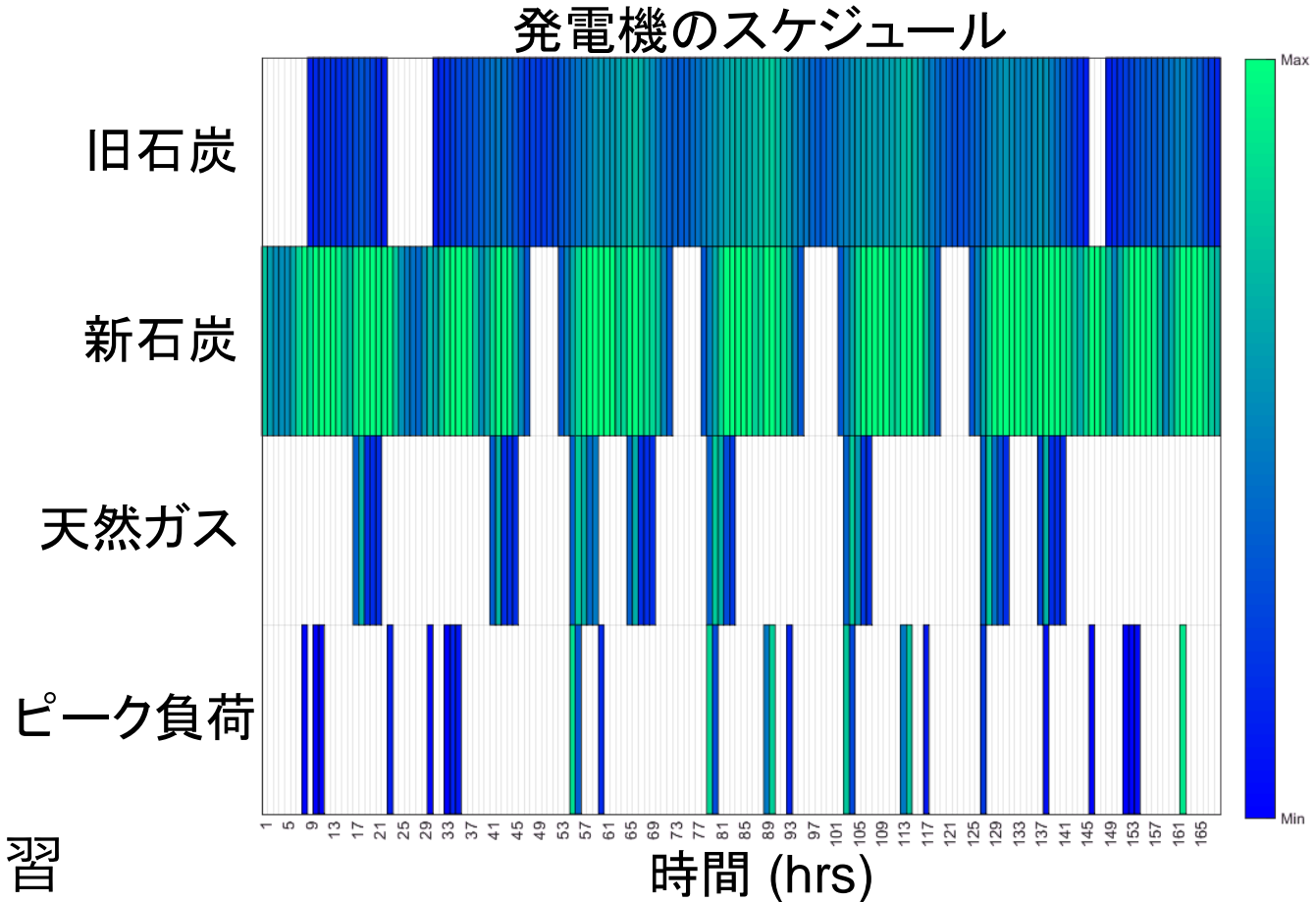
最適化問題:

最小化:

発電コスト

制約条件:

- 1) 需要予測と適合する
- 2) オペレーション上の制約
- 3) その他



Optimization Toolbox™ × 機械学習

最適スケジューリング問題

0-1整数計画問題 (線形計画法の解を整数に限定)

ヘルプセンター

サポートを検索する

ドキュメンテーション 例 関数 ビデオ MATLAB Answers

最新のリリースでは、このページがまだ翻訳されていません。このページの最新版は英語でご覧になれます。

発電機の最適な運転スケジュール: 問題ベース

この例では、収益からコストを引いた後の利益を最大化するよう、2台のガス火力発電機の最適な運転スケジュールを設定する方法を示します。必ずしも現実的な例ではありませんが、判定のタイミングに応じて変動するコストを考慮する方法が説明されています。

この問題に対するソルバーベースのアプローチについては、発電機の最適な運転スケジュール: ソルバーベースを参照してください。

問題の定義

電力市場の価格は、一日の間にその時間帯によって変動します。発電機を運転する場合、価格が高い時間帯に稼働するようスケジュールを設定することで、この価格の変動を有利に利用することができます。たとえば、2台の発電機を管理していると仮定します。各発電機にオフ、低、高の3つの稼働レベルがあり、そのレベルごとに各発電機の燃料消費と電力生産量が決まっています。発電機がオフのときの燃料消費量は0になります。

1日につき30分ごとの間隔(つまり24時間に48回)で、各発電機に稼働レベルを割り当てることができます。過去の記録から、この各時間帯におけるメガワット時 (MWh) あたりの収益がわかっています。この例のデータは Australian Energy Market Operator 提供による <https://www.nemweb.com.au/REPORTS/CURRENT/2013年中期レポートから引用したもので、同社の利用条件> https://www.aemo.com.au/Privacy_and_Legal_Notices/Copyright_Permissions_Notice のもとに使用されています。

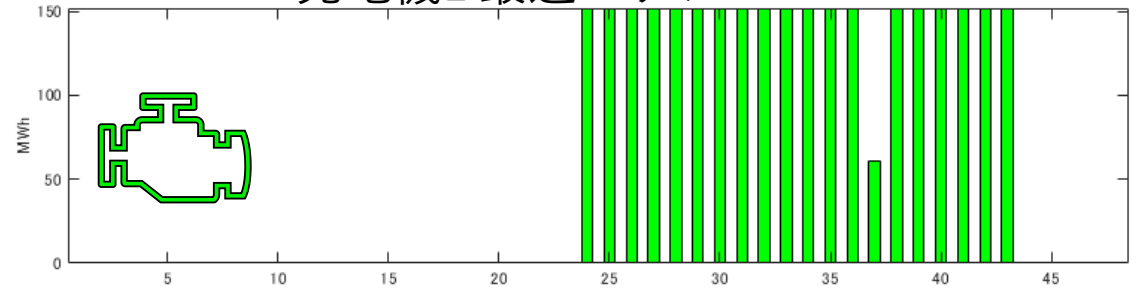
```
load dispatchPrice; % Get poolPrice, which is the revenue per MWh
bar(poolPrice,.5)
xlim([.5,48.5])
xlabel('Price per MWh at each period')
```

Open in MATLAB Online

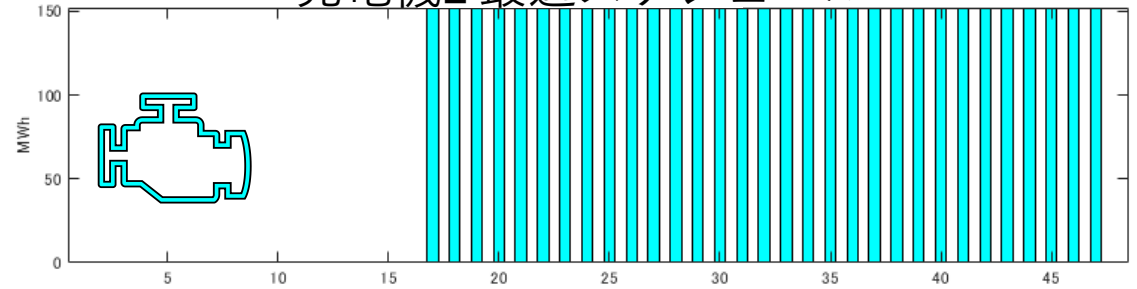
View MATLAB Command

[例題はこちら](#)

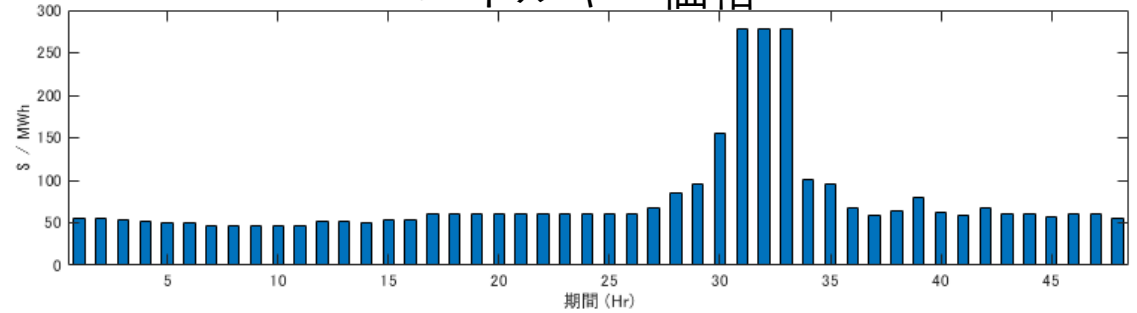
発電機1 最適スケジュール



発電機2 最適スケジュール



エネルギー価格



数値最適化

アプローチは2種類

- ソルバーベース
 - 従来の方法
- 問題ベース
 - ソルバーは問題に合わせて選択される
 - 直感的に記述可能

`optimproblem`
`optimvar`
`solve`

Optimization Toolbox™ Global Optimization Toolbox

制約条件 のタイプ	目的関数のタイプ					
	Linear	Quadratic	Least Squares	General Smooth	General Nonsmooth	Multiobjective
None		<code>quadprog</code>	<code>lsqcurvefit</code> <code>lsqnonlin</code> <code>mldivide</code>	<code>fminsearch*</code> <code>fminunc</code>	<code>fminsearch*</code> <code>patternsearch</code> <code>ga</code> <code>particleswarm</code> <code>simulannealbnd</code>	<code>fgoalattain</code> <code>fminimax</code> <code>paretosearch</code> <code>gamultiobj</code>
Bound	<code>linprog</code>	<code>quadprog</code>	<code>lsqcurvefit</code> <code>lsqnonlin</code> <code>lsqnonneg</code> <code>lsqlin</code>	<code>fmincon</code>	<code>surrogateopt</code> <code>patternsearch</code> <code>ga</code> <code>fminbnd*</code> <code>particleswarm</code> <code>simulannealbnd</code>	<code>fgoalattain</code> <code>fminimax</code> <code>paretosearch</code> <code>gamultiobj</code>
Linear	<code>linprog</code>	<code>quadprog</code>	<code>lsqlin</code>	<code>fmincon</code>	<code>patternsearch</code> <code>ga</code>	<code>fgoalattain</code> <code>fminimax</code> <code>paretosearch</code> <code>gamultiobj</code>
General Smooth	<code>fmincon</code>	<code>fmincon</code>	<code>fmincon</code>	<code>fmincon</code>	<code>patternsearch</code> <code>ga</code> <code>surrogateopt</code>	<code>fgoalattain</code> <code>fminimax</code> <code>paretosearch</code> <code>gamultiobj</code>
General Nonsmooth	<code>patternsearch</code> <code>ga</code> <code>surrogateopt</code>	<code>patternsearch</code> <code>ga</code> <code>surrogateopt</code>	<code>patternsearch</code> <code>ga</code> <code>surrogateopt</code>	<code>patternsearch</code> <code>ga</code> <code>surrogateopt</code>	<code>patternsearch</code> <code>ga</code> <code>surrogateopt</code>	<code>paretosearch</code> <code>gamultiobj</code>
Integer	<code>intlinprog</code>				<code>ga</code> <code>surrogateopt</code>	

線形、非線形、局所解、大域解、線形制約、非線形制約
様々なソルバーが利用可能

実際の電力予測: Naturgy Energy Group S.A.

課題

利用可能供給量と需要ピークを予測し、エネルギー取引のマージンを最大化する

解決案

過去のデータ、天気予報と規制ルール統合と、モデルの構築と最適化をMATLAB®を使って行う

結果

- レスポンスタイムが数か月短縮
- 2倍の生産性
- プログラムの保守が簡略化

機械学習

+

X



Portomouros 水力発電ダム

“シフトしていく発電制約と変化する需要にいち早く対応しなかったため、我々は閉じていたり独占された解決策に依存することはできません。MathWorksのツールを用いて、より正確な結果を得ることができました – さらには、変化するニーズに応じて自分たちのモデルの開発、更新、そして最適化を柔軟に行うことができるようになりました”

- Angel Caballero, Gas Natural Fenosa

[Link to user story](#)

フリート解析

装置のノウハウ

設計仕様
オペレーションモード
オペレーション条件

機械学習

統計解析
教師無し学習

電力予測

電力網のノウハウ

季節性
天候の影響
発電機の特性

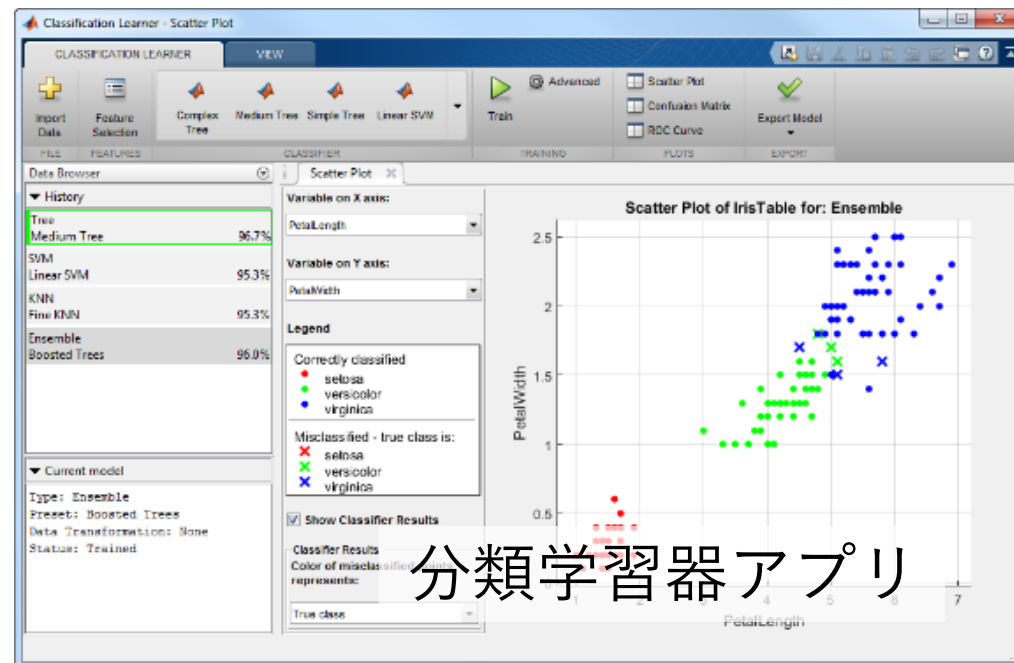
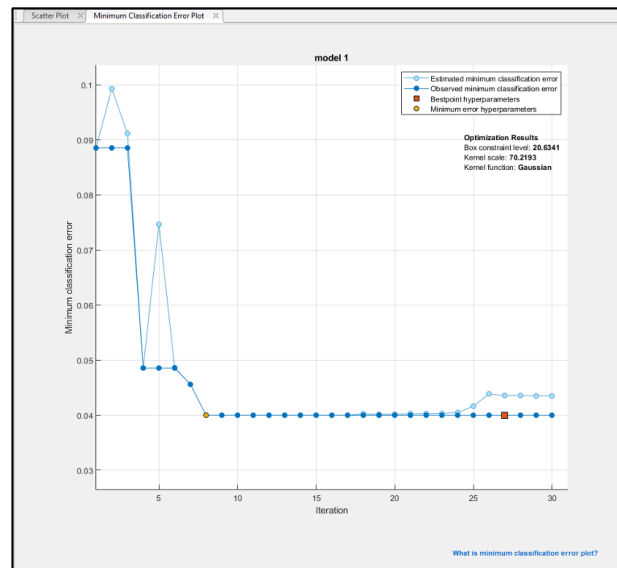
機械学習

時系列モデリング
回帰

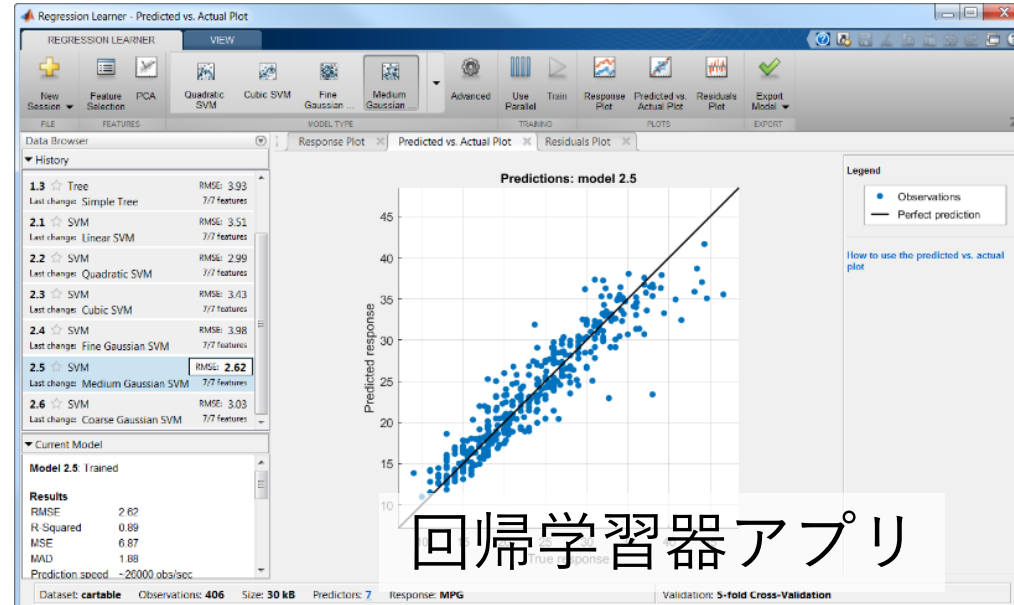
機械学習アプリ

- 複数のモデルを試す
- 結果の比較
- 詳細について心配せずに、合理的なモデルにたどり着く

アプリ内で
ハイパーパラメータ
の最適化を実行
(ベイズ最適化)



分類学習器アプリ



回帰学習器アプリ

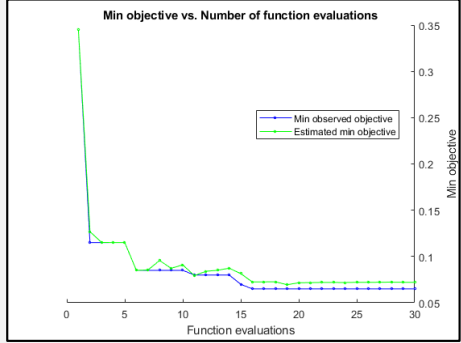
AutoML

- 複数の機械学習モデルを構築
- 専門家でなくても、“良いモデル”を見つけることができる

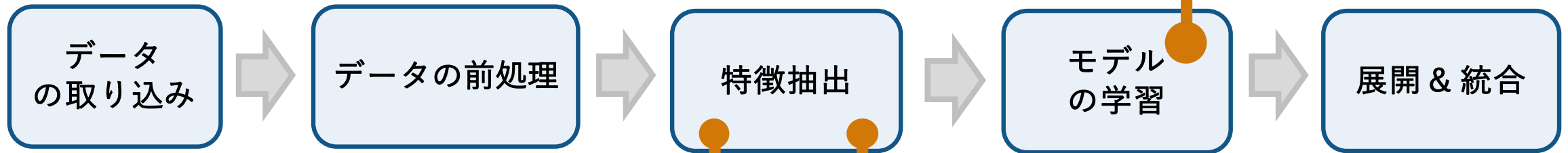
モデル選択

R2020a
fitcauto

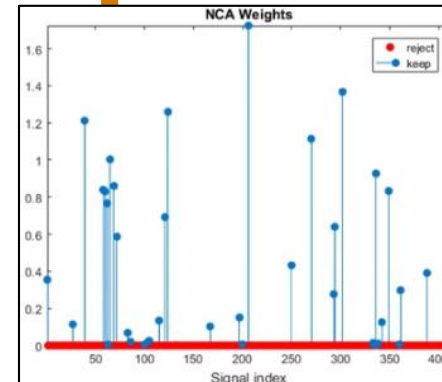
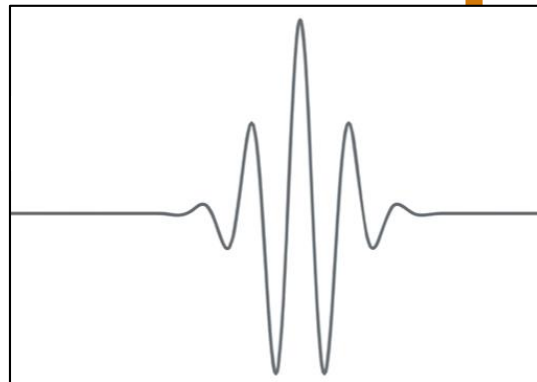
(自動)ハイパーパラメータの最適化



決定木?
SVM?
KNN?
アンサンブル?
…?



ウェーブレット
散乱
(自動特徴抽出)



(自動)特徴選択

AutoML “実行中”

% Step1: ウェーブレット散乱を適用し、自動特徴抽出

```
sf = waveletScattering('SignalLength',N, 'SamplingFrequency',50);  
Wfeatures = featureMatrix(sf,thisSignal(1:N),'Transform','Log');  
% do this across signals <thisSignal> and accumulate <allFeatures> with labels
```

% Step2: 自動でTop N の特徴量をランキングから選択, e.g., MRMR

```
[mrmrFeatures , scores] = fscmrmr(allFeatures, 'class');  
trainFeatures = allFeatures(:, [mrmrFeatures(1:numPredictorsToUse);true]);
```

% Step3: 自動で最適なモデルを100回の反復から選択

```
modelAuto = fitcauto(trainFeatures,'class', 'Learners','all',  
'MaxObjectiveEvaluations',100);
```

AutoML “実行中”

% Step1: ウェーブレット散乱を適用し、自動特徴抽出

```
sf = waveletScattering('SignalLength',N, 'SamplingFrequency',50);  
Wfeatures = featureMatrix(sf,thisSignal(1:N),'Transform','Log');  
% do this across signals <thisSignal> and accumulate <allFeatures> with labels
```

% Step2: 自動でTop N の特徴量をランキングから選択, e.g., MRMR

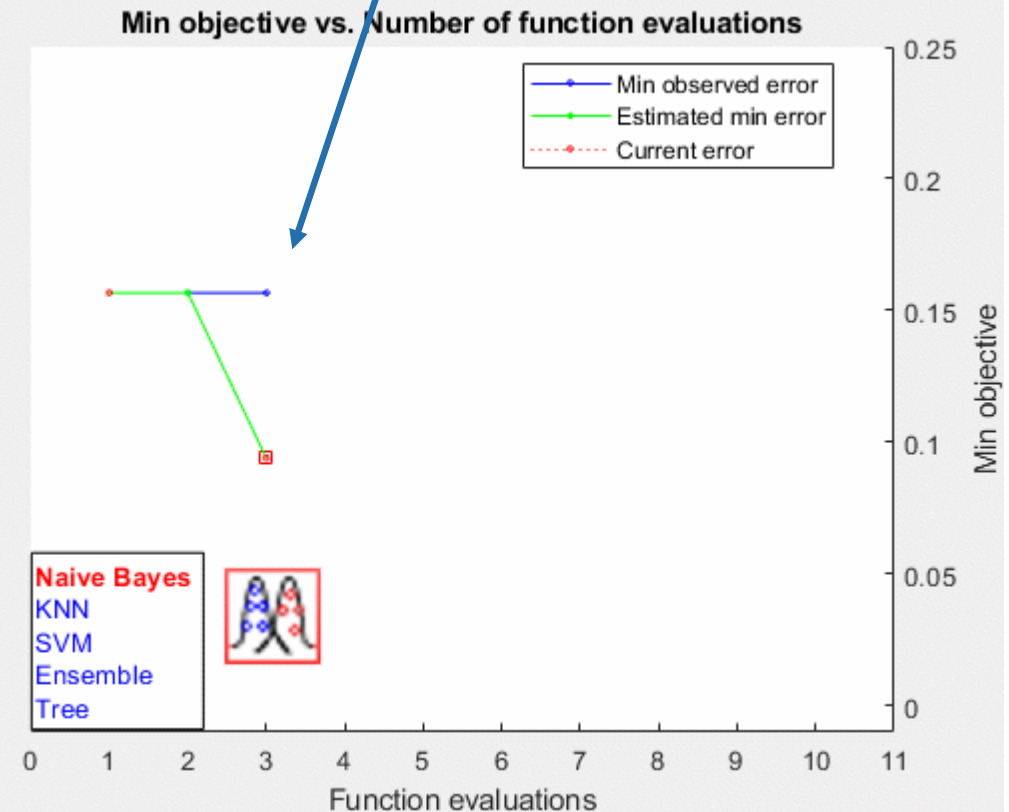
```
[mrmrFeatures , scores] = fscmrmr(allFeatures, 'class');  
trainFeatures = allFeatures(:, [mrmrFeatures(1:numPredictorsToUse);true]);
```

% Step3: 自動で最適なモデルを100回の反復から選択

```
modelAuto = fitcauto(trainFeatures,'class', 'Learners','all',  
'MaxObjectiveEvaluations',100);
```

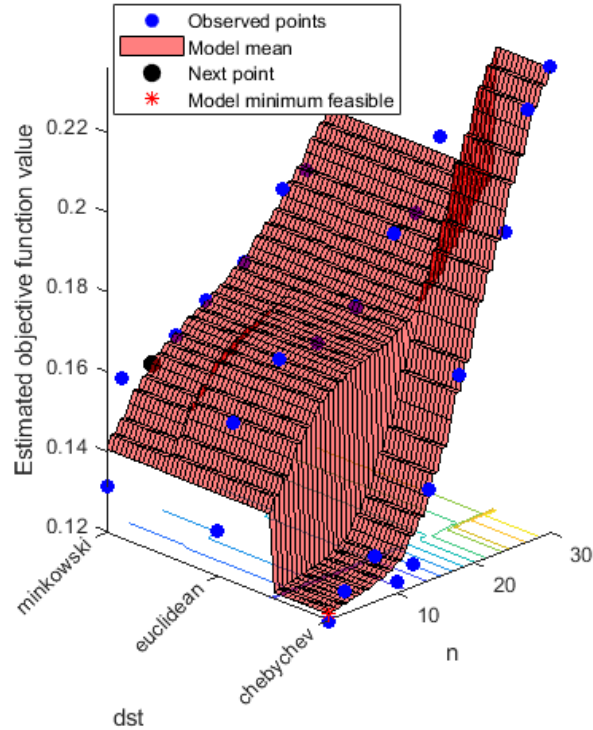
1行のコマンドで探索

ハイパーパラメータの空間
×
モデルの空間



“MATLAB®の”ベイズ最適化機能

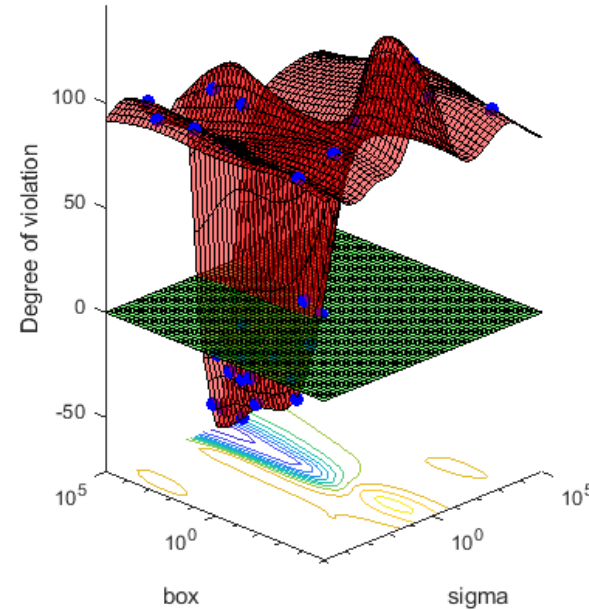
目的関数モデル



目的関数のガウス過程回帰

- 未知の点の事後確率分布関数を用いた高効率な解の探索
- 解の実現可能性も考慮した探索

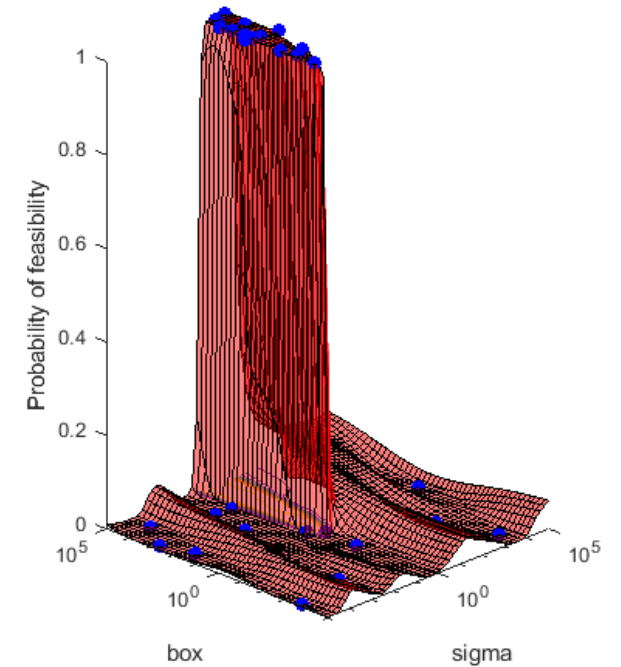
制約条件モデル



制約条件

- 確定的制約
- 条件付き制約
- 連結制約

解の実現可能性確率



解の実現可能性

- 制約やエラー値などから、解の存在可能性の確率分布関数

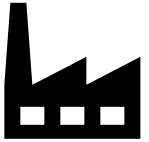
“結果”を出した機械学習のアプリケーション例



フリートデータ解析



電力予測



製造アナリティクス

製造アナリティクスとは？

定義: 作業効率を最大化するために、プロセスやセンサーデータにモデリング (AI) を適用すること

主要な活用事例:

1. 生産プロセスの監視の自動化
2. 製品品質の保証
3. 複雑な製造プロセスの生産量の最適化

スマートマニュファクチャリング
インダストリー4.0

AIを製造業に適用する際の課題

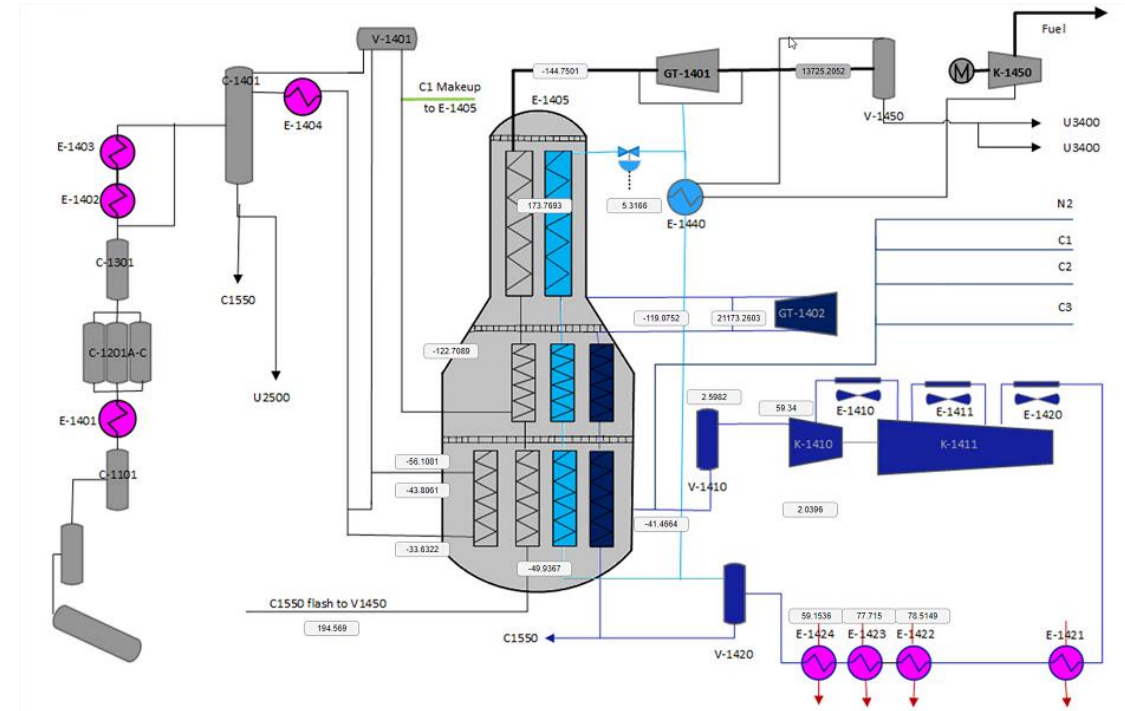
1. “データヒストリアン” (SCADA, LIMS, OSISoft PI) に残る大量のデータ

2. 信頼できる測定やモデリングが困難

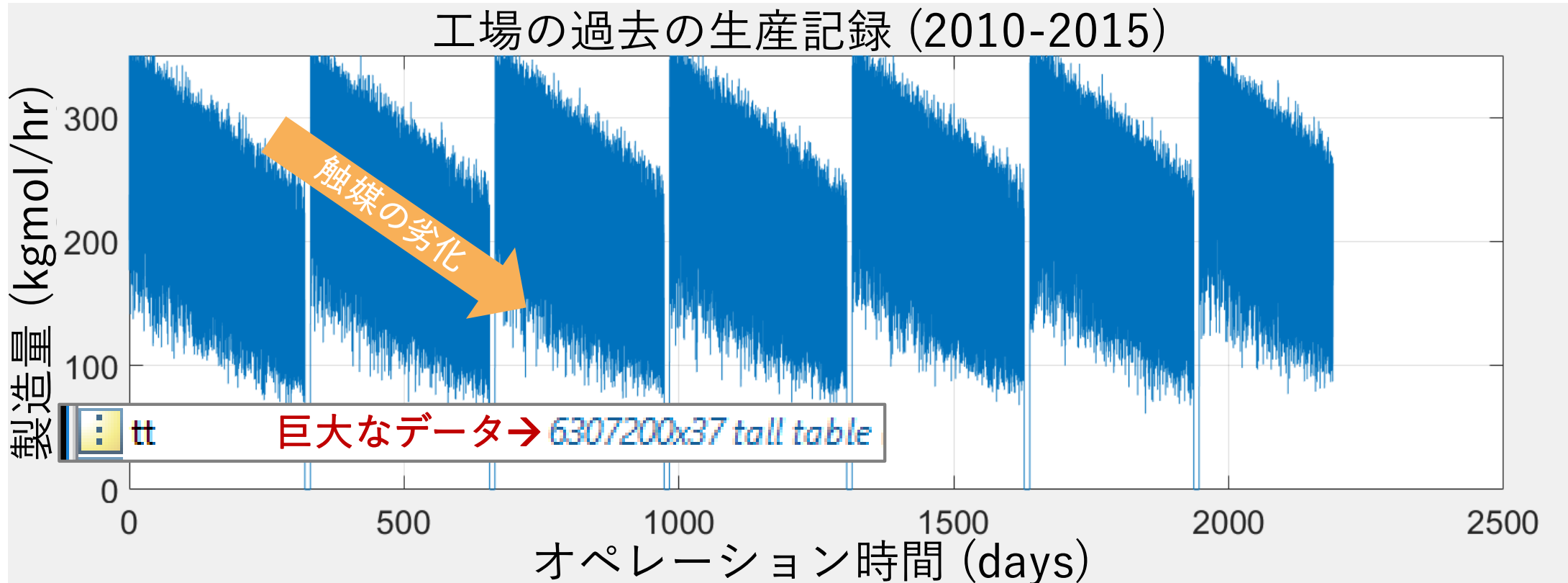
- センサーの故障
- 隠れ変数 (経年数など)

3. 多数の異種のツールの使用

- 限定的な予測モデリング (自動化などが難)
- ストリーミングデータの処理
- カスタマイズ性 (コーディング無しでの利用)

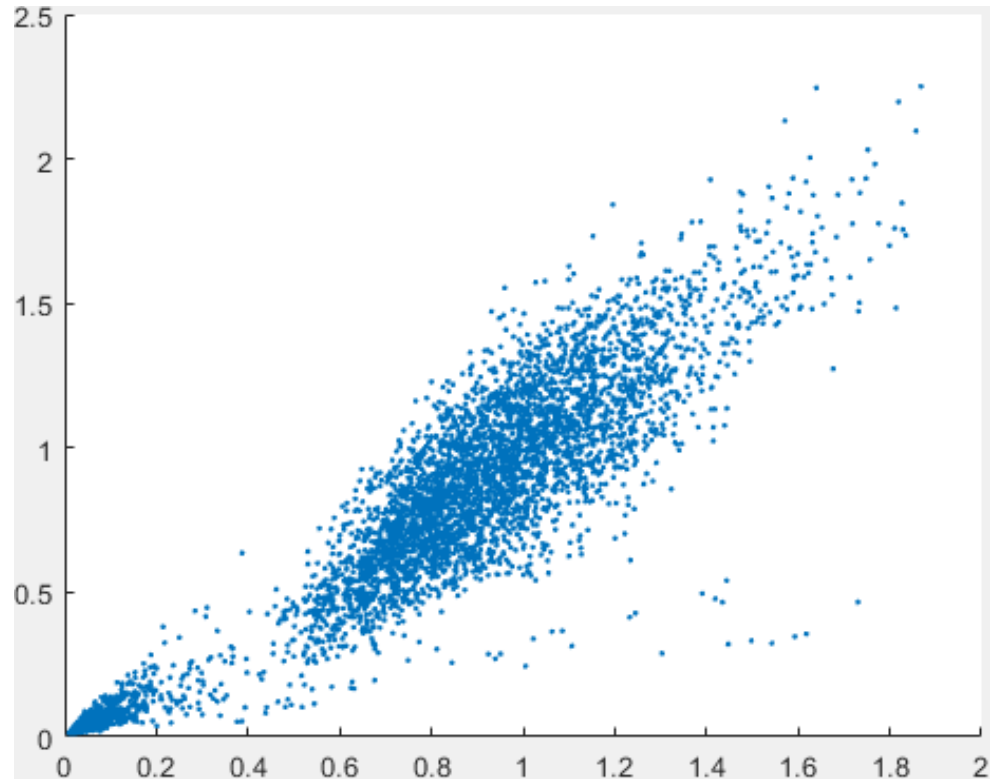


プロセスモデリングで隠れ変数を考慮する



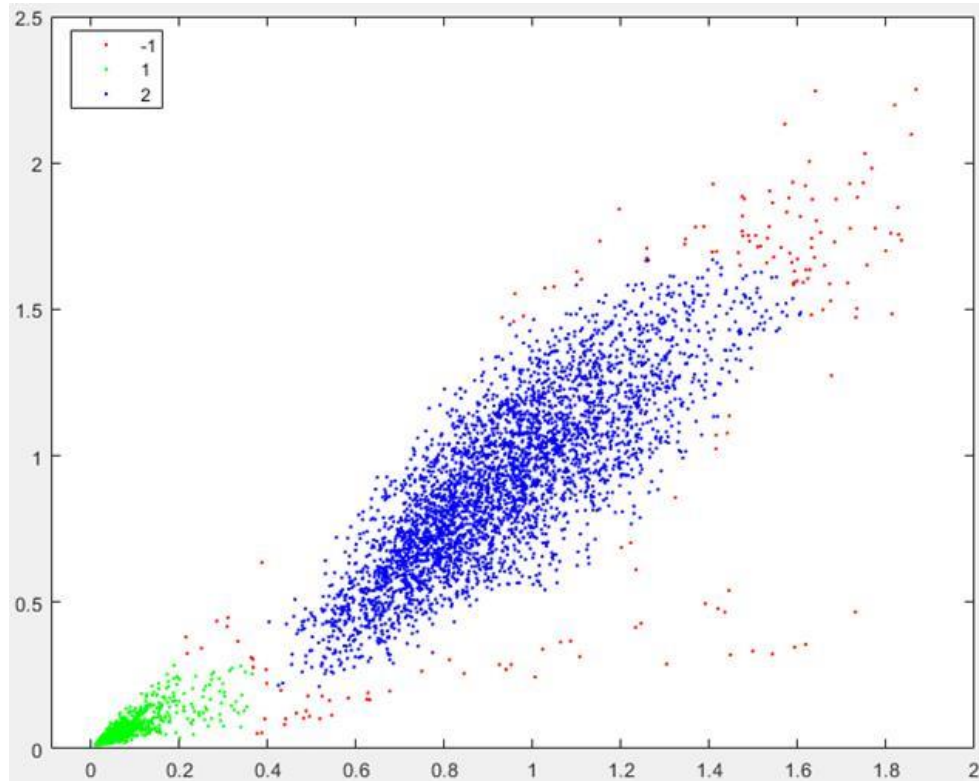
← MATLAB®のビッグデータ用ツールを使ってモデリング →

ケーススタディ: 異常検知

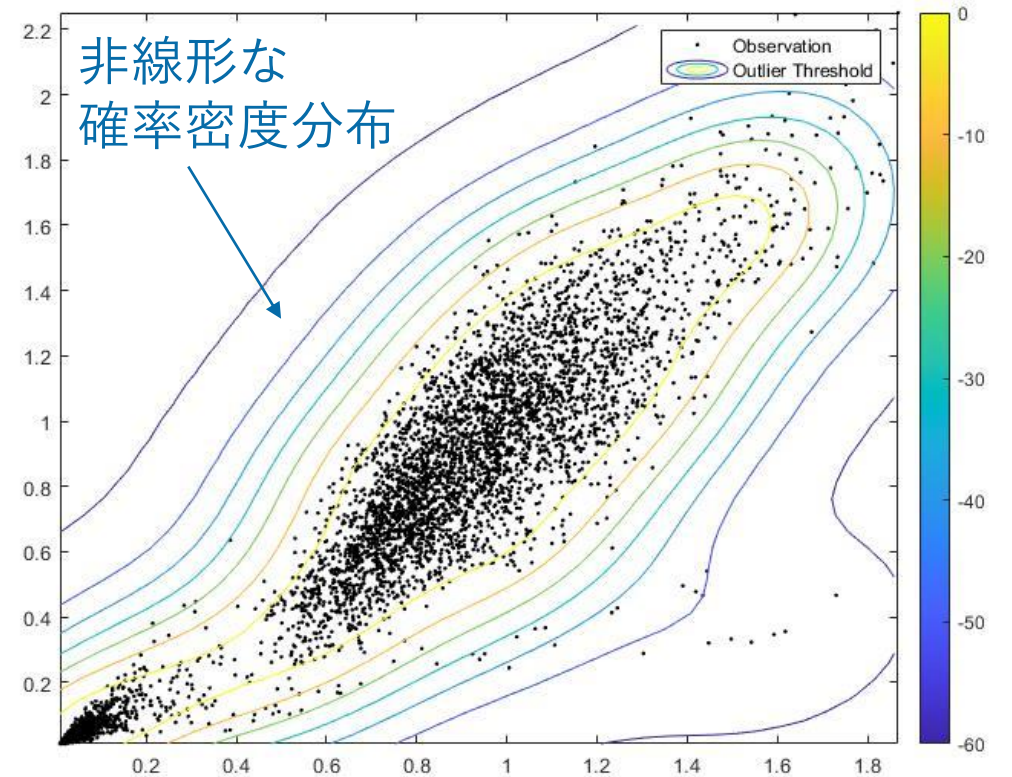


ケーススタディ: 異常検知

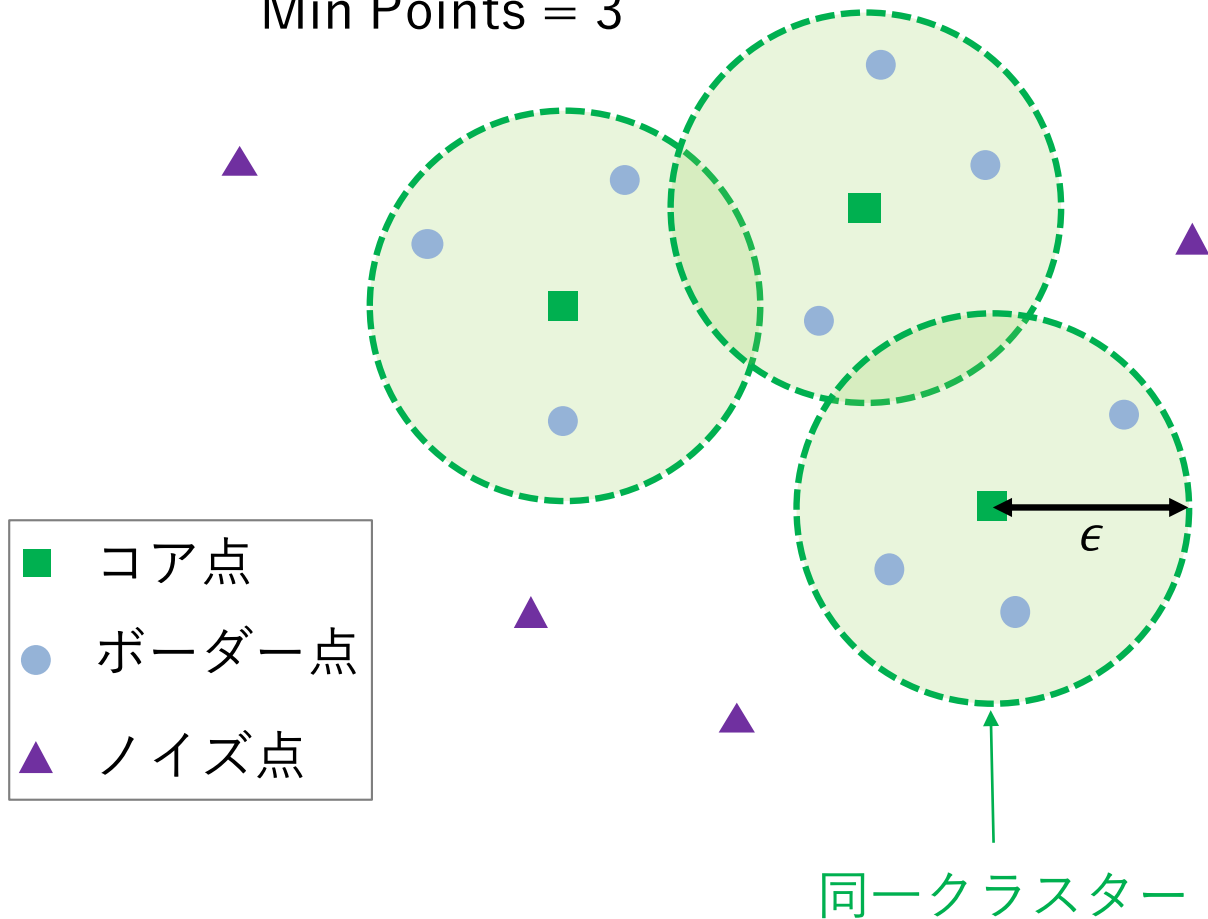
1. DBSCANを使ったクラスター



2. One-class SVM



Min Points = 3

**STEP 1: 点を定義**

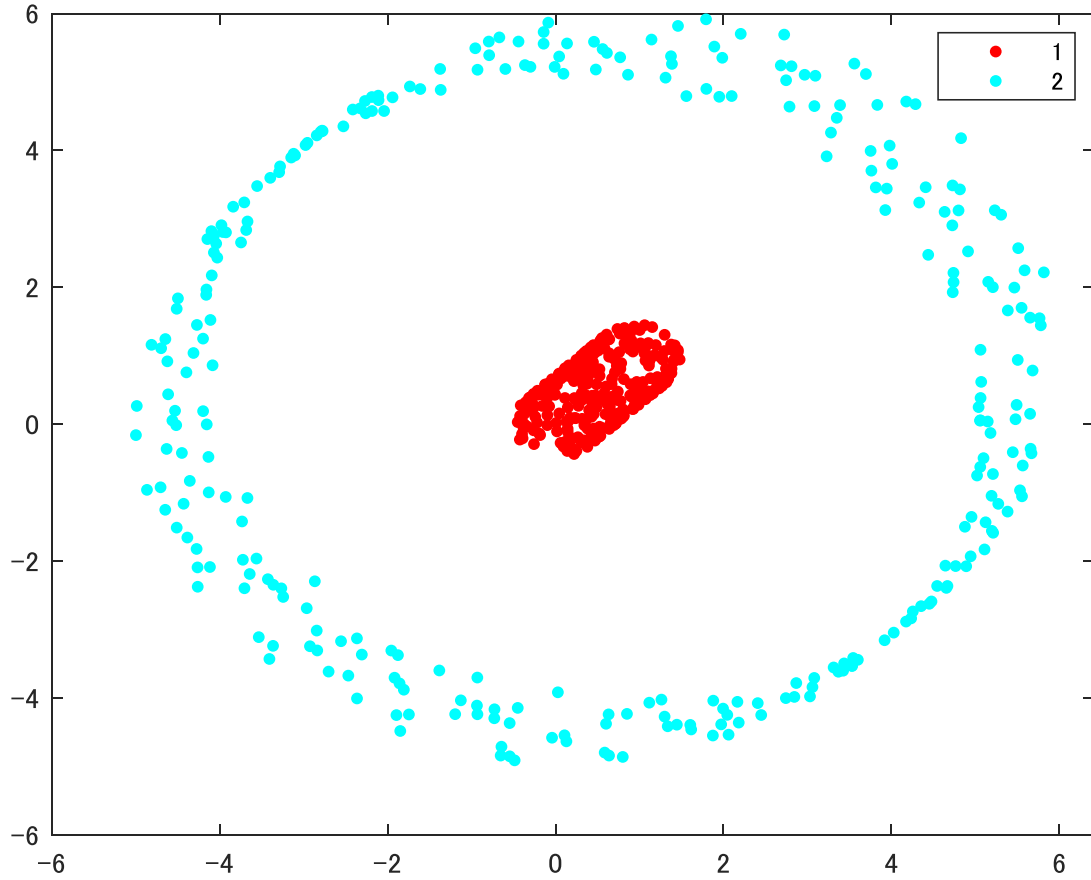
1. 半径 ϵ 以内に, Min Points以上の隣接点を持つものをコア点とする
2. コア点に満たない点で、コア点の作る半径以内に存在する点をボーダー点とする
3. それ以外をノイズ点

STEP 2: クラスタ分類

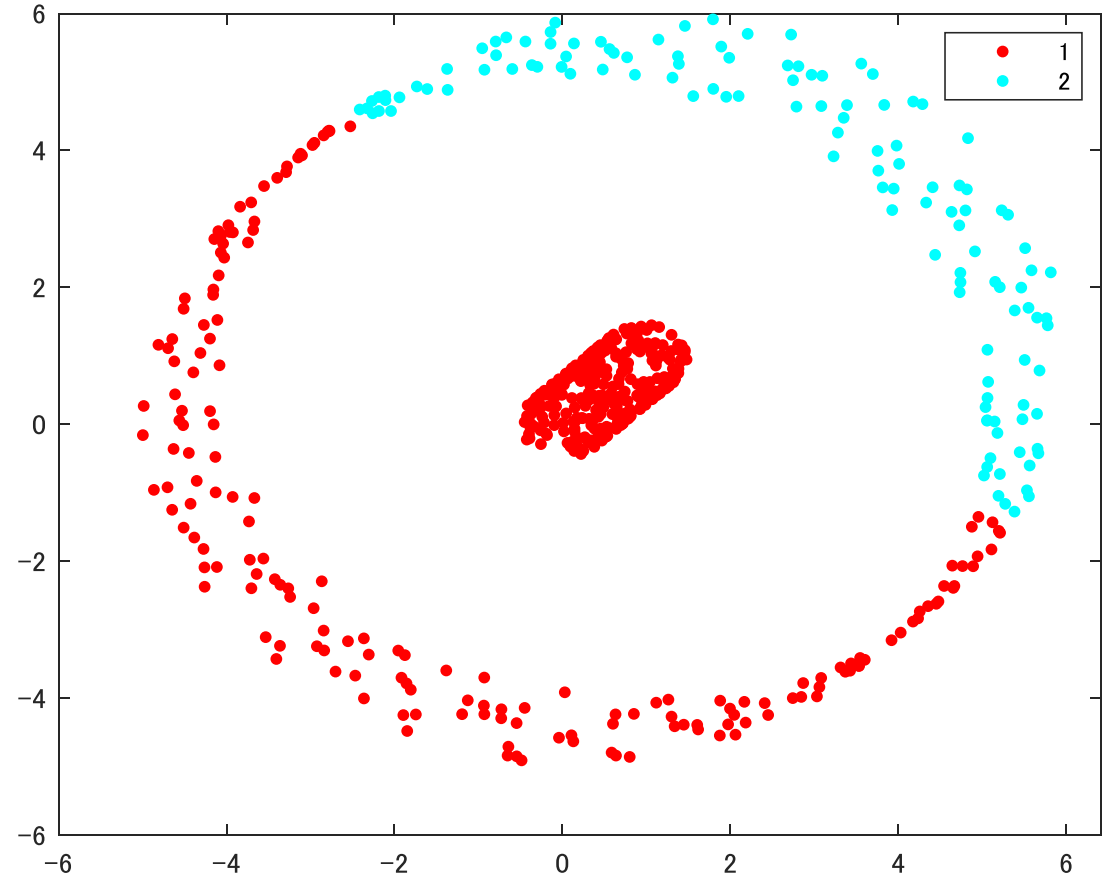
1. 二つのコア点が ϵ 以内にお互い存在している場合、接続関係があるとして同じクラスターに分類
2. 各ボーダー点をそれとついでになるコア点のクラスターに割り当てる

DBSCAN vs K最近傍法

DBSCAN R2019a



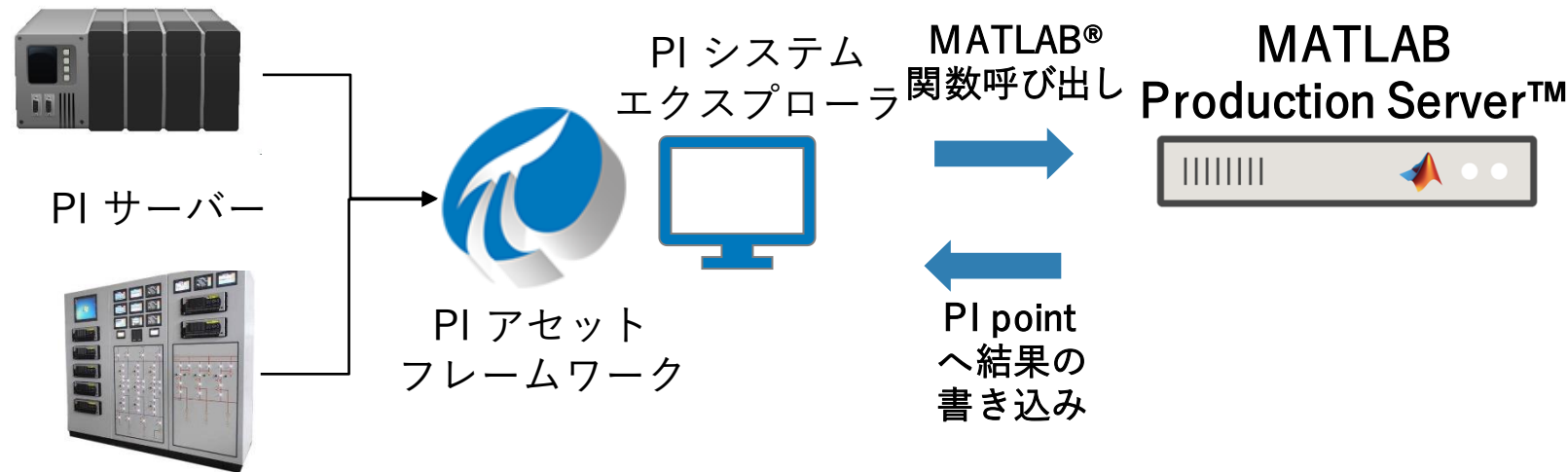
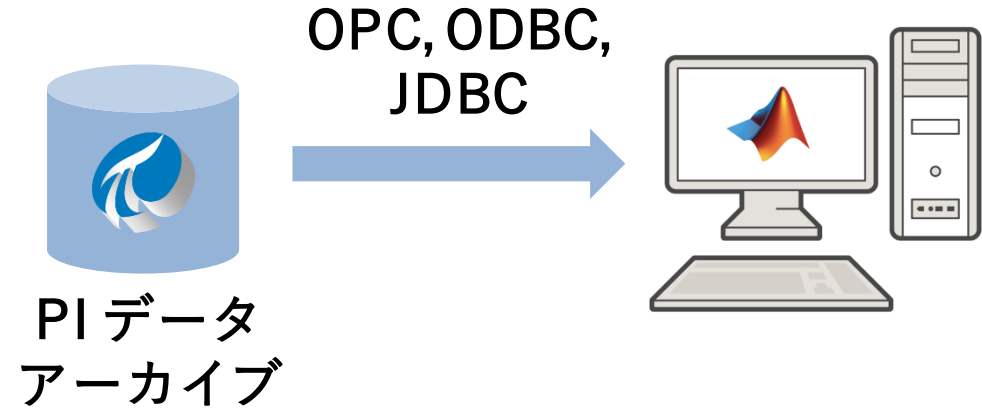
K最近傍法



展開

データヒストリアンとの統合

- OPC Toolbox™ (Database Toolbox™ via ODBC/JDBC)でPI サーバーに接続



解析手法をカスタムして共有

- 工場スタッフとプロセスエンジニアがGUIで解析
- App Designer を使ってカスタムダッシュボードを作成

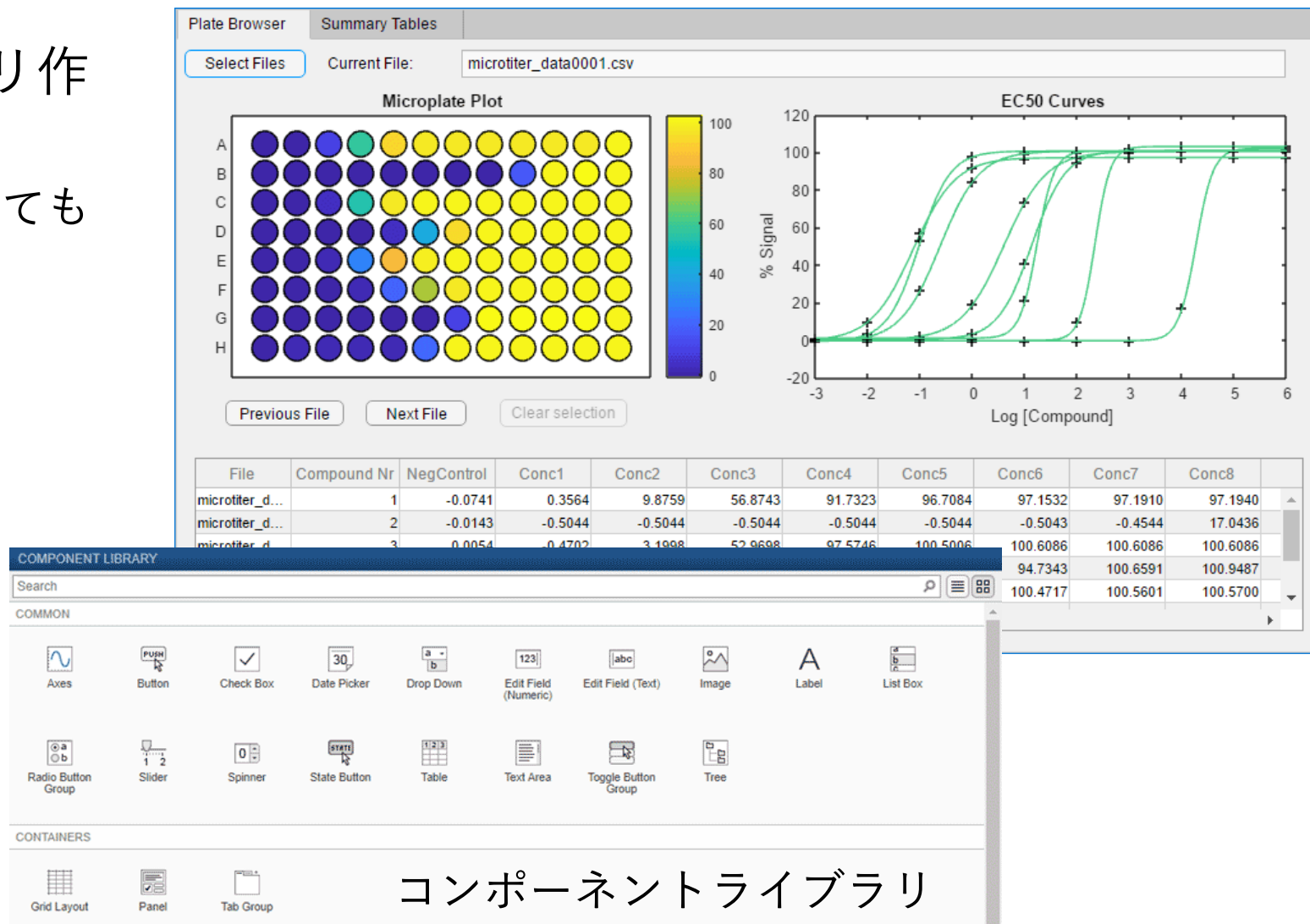
MATLAB App Designerでカスタムダッシュボードを作成

- MATLAB®を使用したアプリ作成

- ソフトウェアの専門家でも完成度の高いアプリを作成可
- スタンドアロンアプリ化
- Webアプリ化

- コンポーネント操作

- コンポーネントライブラリ
- コールバック関数
- マウス・キーボード操作



フリート解析

装置のノウハウ

設計仕様
オペレーションモード
オペレーション条件

機械学習

統計解析
教師無し学習

電力予測

電力網のノウハウ

季節性
天候の影響
発電機の特徴

機械学習

時系列モデリング
回帰

製造アナリティクス

製造業のノウハウ

プロセス装置
変数 & 設定値
パラメータの影響

機械学習

異常検知 回帰
多変量統計解析

デスクトップからプロダクションへ



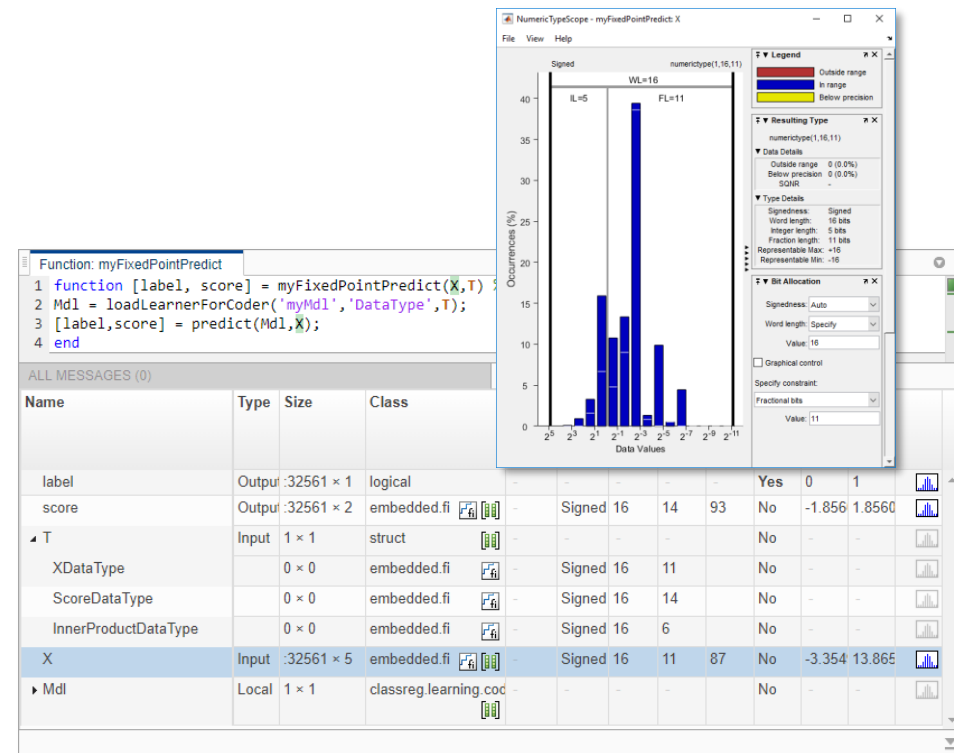
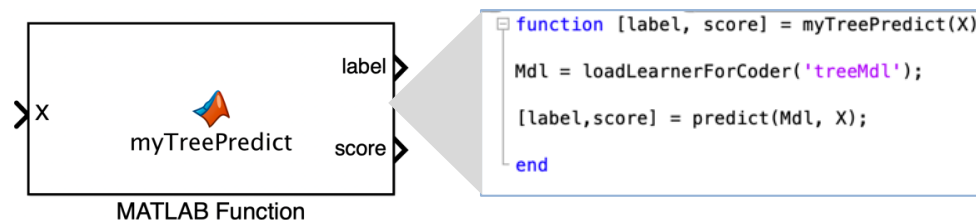
更新を行う理由:

- よりよいモデルが見つかる
- 新しいデータが利用可能
- ビジネスが変化
- ...

更新の度にモデルをアップデートするのは高コスト!

自動 C/C++ コード生成

1. ほぼ全ての分類と回帰モデルの推論
2. 展開されたモデルを再コーディングなしで更新
 - SVM, 決定木, 線形モデル
3. 固定小数点サポート
 - SVM, 決定木, アンサンブル木
 - 浅いニューラルネットワーク (Simulinkを通して)
4. MATLAB Function BlockとしてSimulinkモデルと統合



[MATLAB®を他の言語と統合する](#)

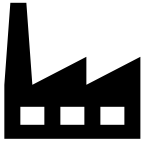
“結果”を出した機械学習のアプリケーション例



フリートデータ解析



電力予測



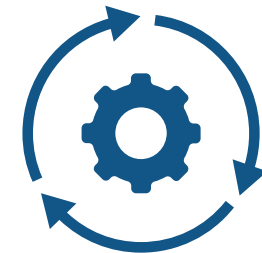
製造アナリティクス

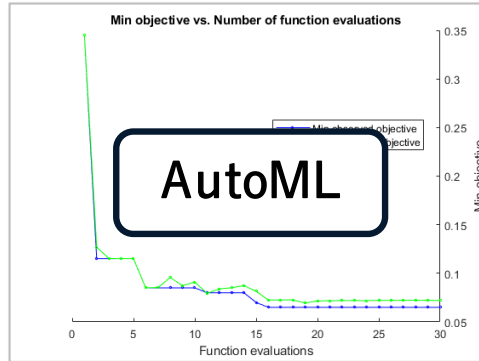
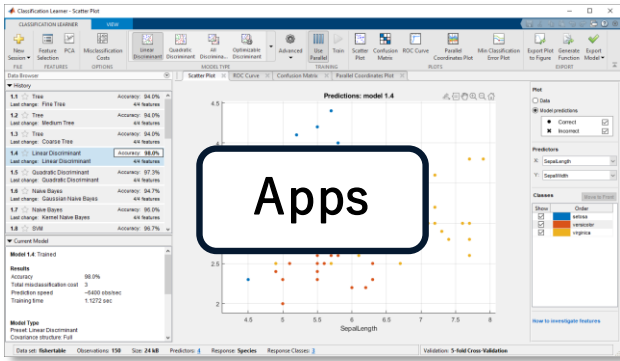
新機能

- MATLAB® アプリ
- AutoML
- 信号処理を機械学習に用いる
- C/C++ コード生成



Innovation





```

Function: myFixedPointPredict
1 function [label, score] = myFixedPointPredict(X,T) %codegen
2 Md1 = loadLearnerForCoder('myMd1', 'DataType', T);
3 [label, score] = predict(Md1, X);
4 end
  
```

Name	Type	Size	Class	DT	Signed	Min	Max	Always	Sim	Sim
						Whole	Min	Num	Min	Max
label	Out							Yes	0	1
score	Out							No	-1.856	1.8560
T	Input							No		
X	Input	32561 × 5	embedded fi		Signed	16	11	87	No	-3.354 13.865
Md1	Local	1 × 1	classreg learning coc					No		

機械学習

フリートデータ解析

信号処理

業界のノウハウ

電力予測

製造アナリティクス

アプリケーションのノウハウ

医療機器

素材製造

更に学びたい方へ



MATLAB 入門



機械学習入門



ディープラーニング入門

- 無料の入門コース
- 受講にライセンス不要
- Webブラウザで操作



Stateflow 入門



Simulink 入門

<https://jp.mathworks.com/services/training.html>

本格的に学びたい方・“成功”への最短ルートをお求めの方へ

各種トレーニングコース

MATLAB® 基礎 (3日間)

MATLAB®によるデータ処理と可視化 (1日間)

MATLAB®によるビッグデータ処理 (1日間)

MATLAB®による統計解析 (2日間)

MATLAB®による機械学習 (2日間)

MATLAB®によるデータの前処理および特徴抽出 (1 day)

MATLAB®によるディープラーニング (2日間)

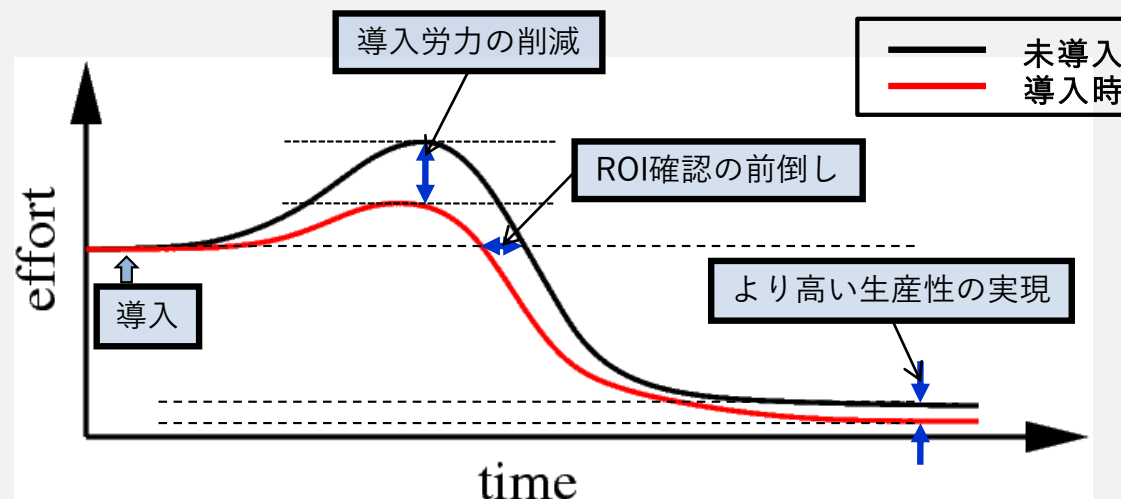
MATLAB®コードの高速化と並列処理 (2日間)

<https://jp.mathworks.com/services/training.html>

技術コンサルティングサービス

MathWorks社内の専門知識を活用

- 透明性の高いアプローチ
- カスタマイズされたサポート
- ROI 向上



<https://jp.mathworks.com/services/consulting.html>

MATLAB EXPO

