

MATLAB EXPO 2018

センサー/テキストデータ解析分野への
機械学習の適用

アプリケーションエンジニアリング部
井原 瑞希



本日のトピック

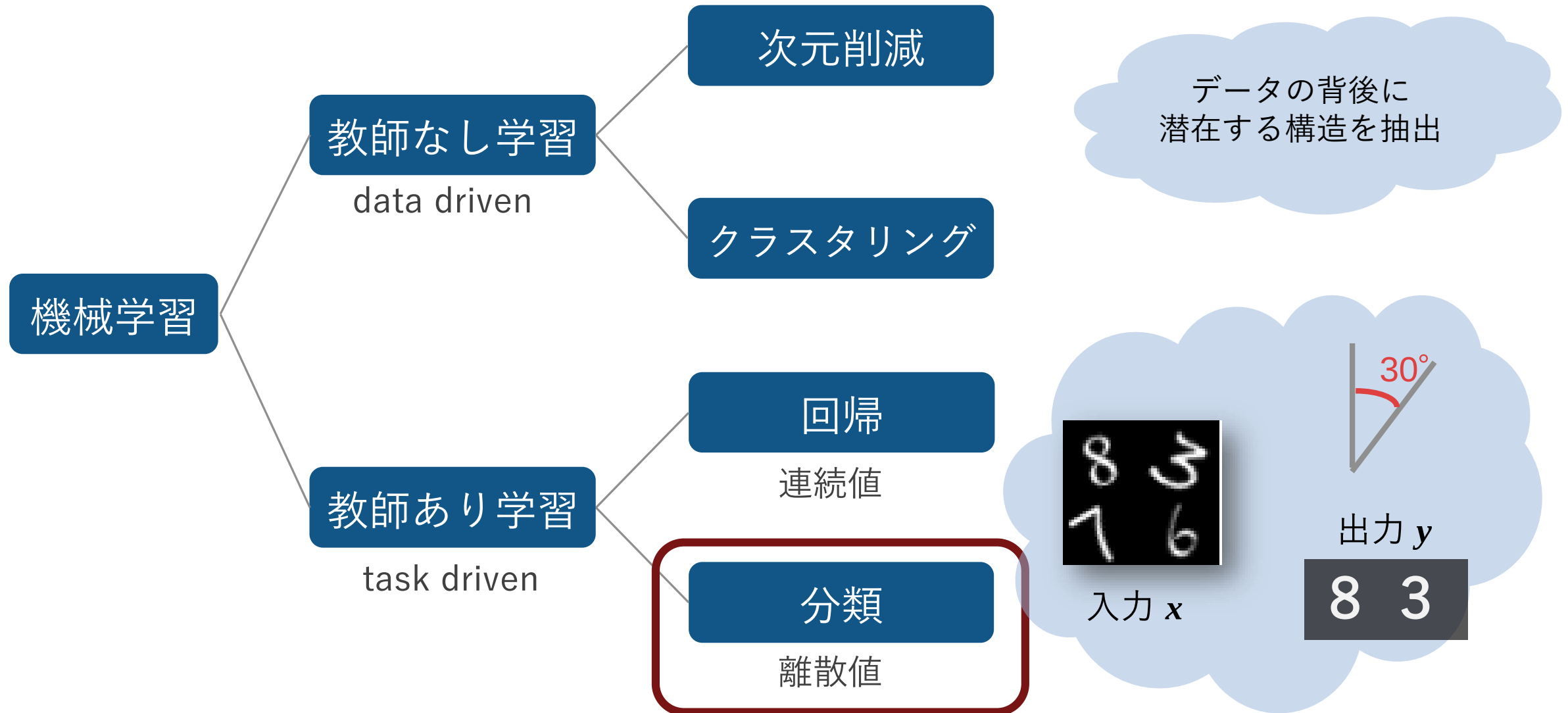
■ センサーデータの分類

- 手法の選択 – 従来の機械学習手法とディープラーニング
- センサーデータ解析特有の課題
- 分類モデルの選択

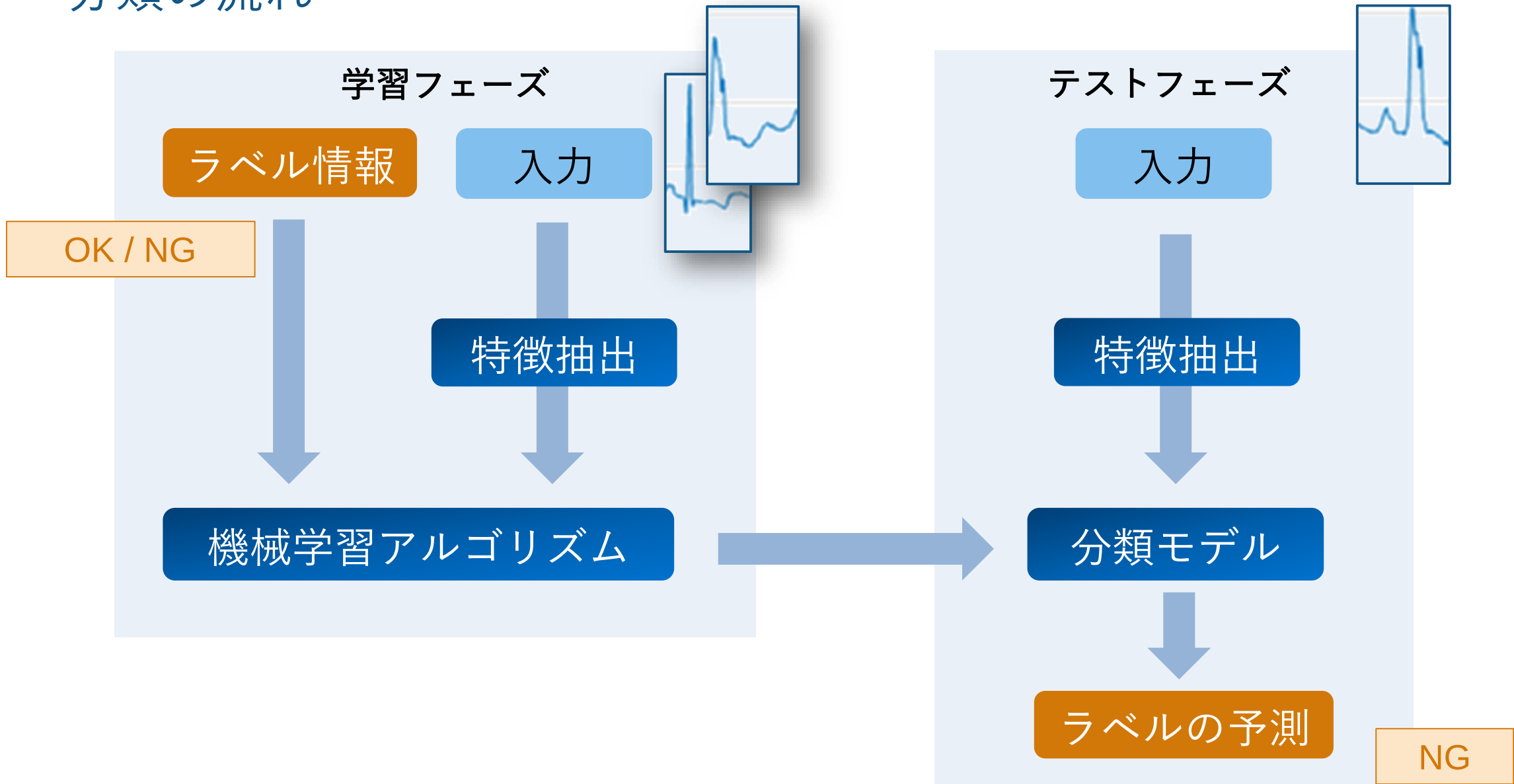
■ テキストの類似性分析

- テキストからの特徴抽出
- テキストデータ特有の課題
- テキスト解析ワークフロー
- 機械学習手法とディープラーニングの使い分け

課題の見極め



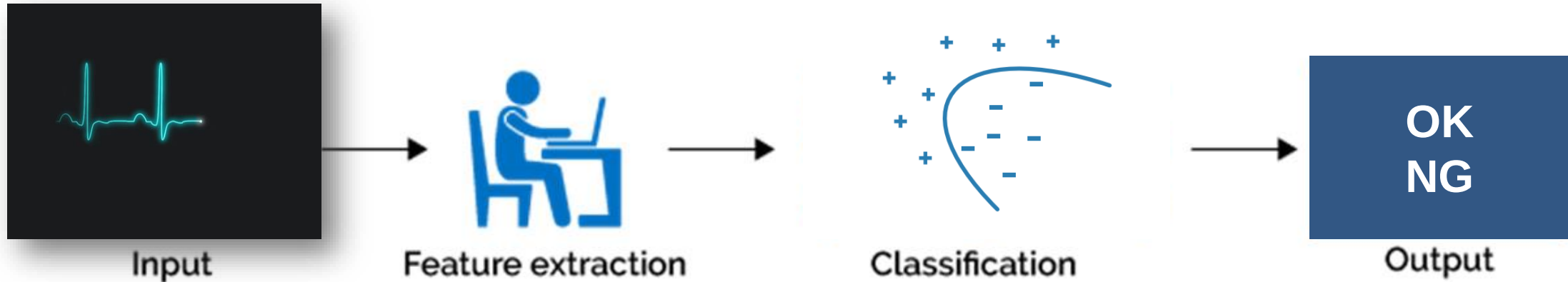
分類の流れ



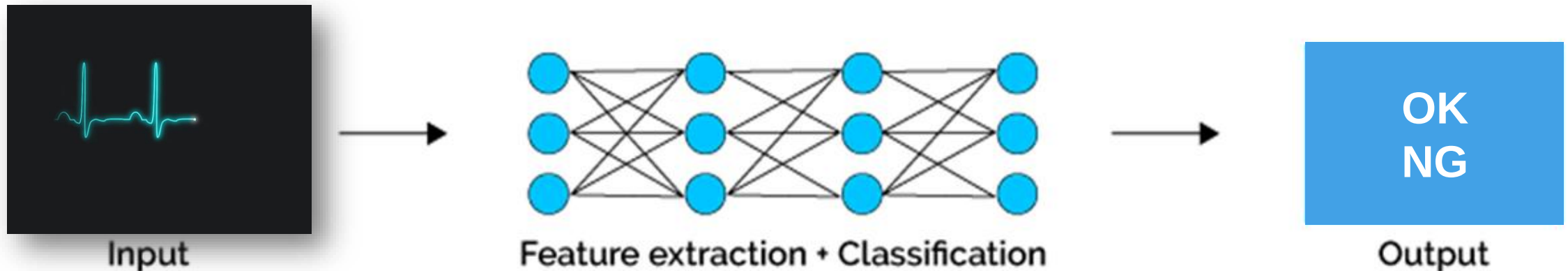
センサーデータ解析のアプローチ

従来の機械学習手法とディープラーニング

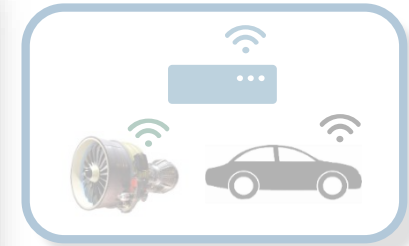
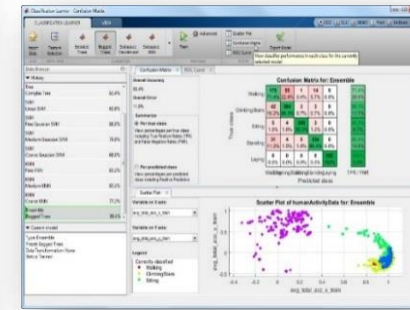
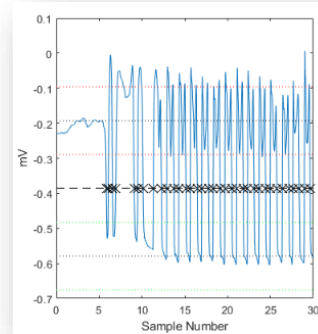
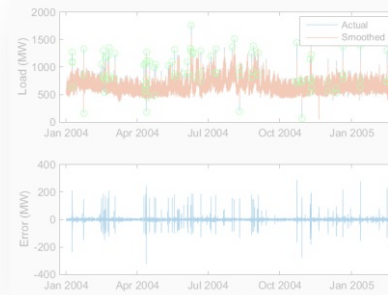
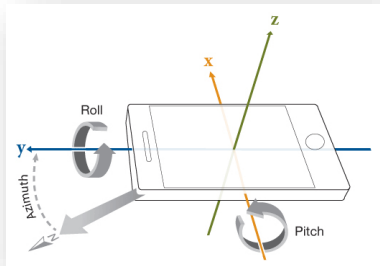
従来の機械学習



Shallow & Deep Neural Networks



センサーデータ解析特有の課題



データ
アクセス

前処理

特徴抽出

モデルの構築

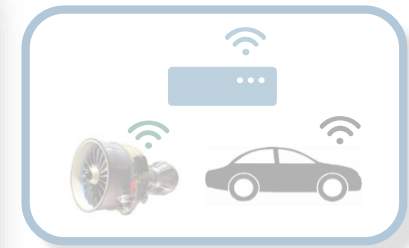
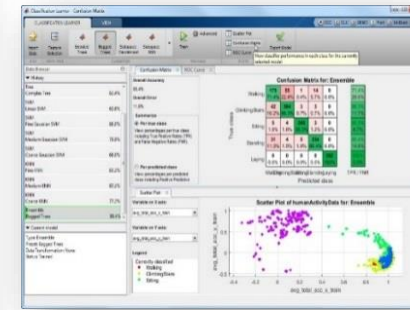
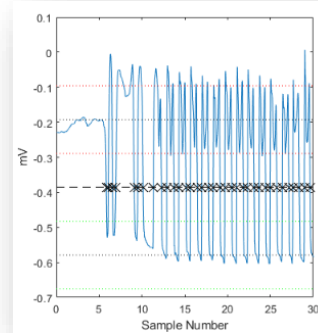
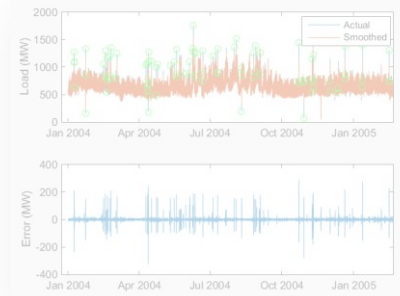
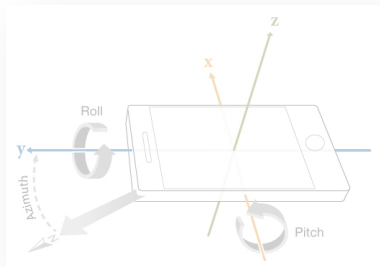
実環境展開

データが膨大になりがち

ドメイン知識を活用したい
良い特徴量がわかりにくい

どの手法を選べばよいか不明

センサーデータ解析特有の課題



データ
アクセス

前処理

特徴抽出

モデルの構築

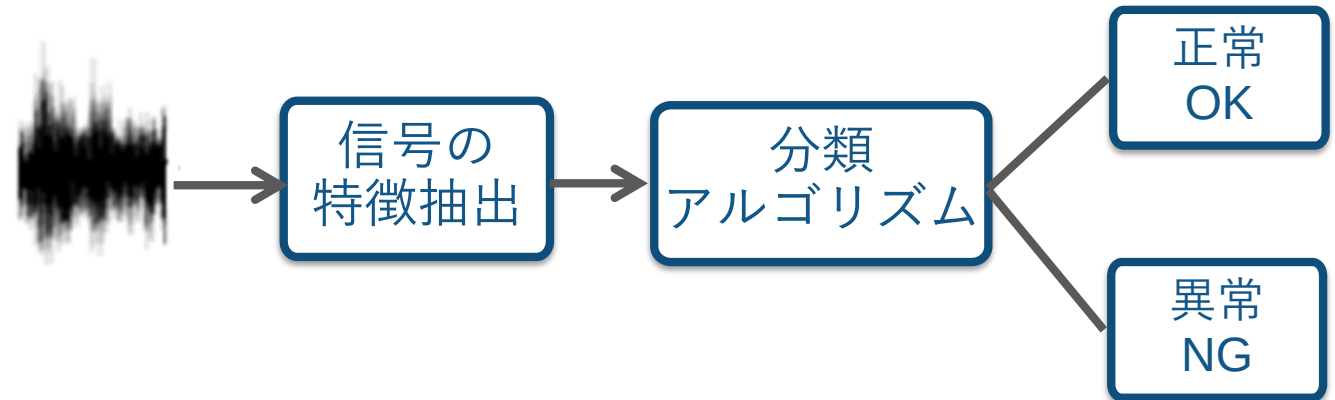
実環境展開

例1: 心電図の判別

従来の機械学習

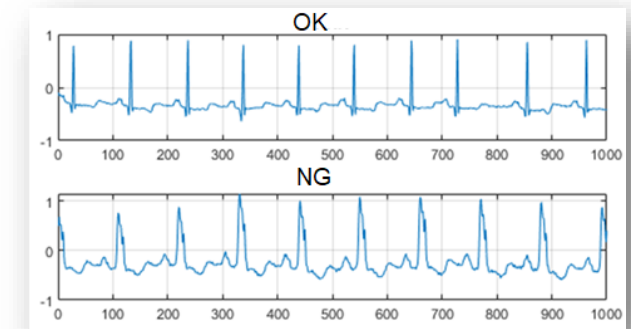
目的

- 心電図の信号から、その心電図の状態を判断する分類モデルを作成



1. 従来の機械学習アプローチ

- センサーデータの特徴抽出
- データの細分化や周波数特性の可視化
- 機械学習を使用した分類モデルの作成



例1: 心電図の判別

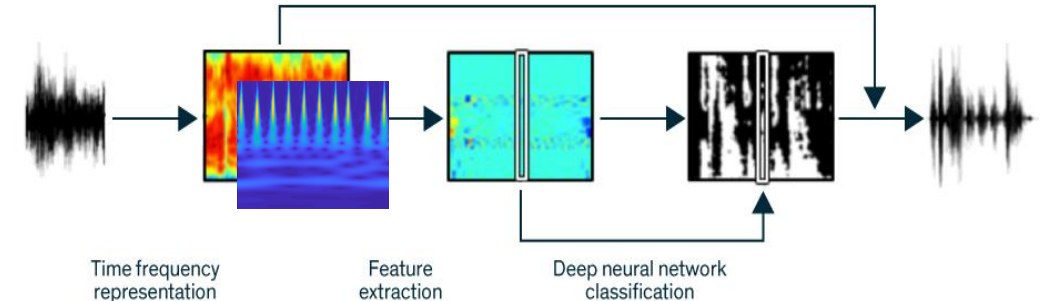
ディープラーニングによる分類

2-1. 畳み込みニューラルネットワーク (CNN) のアプローチ

- 信号から画像への変換（時間-周波数領域表現）
- ネットワークパラメータの学習

利点: 画像の転移学習ワークフローが利用可能

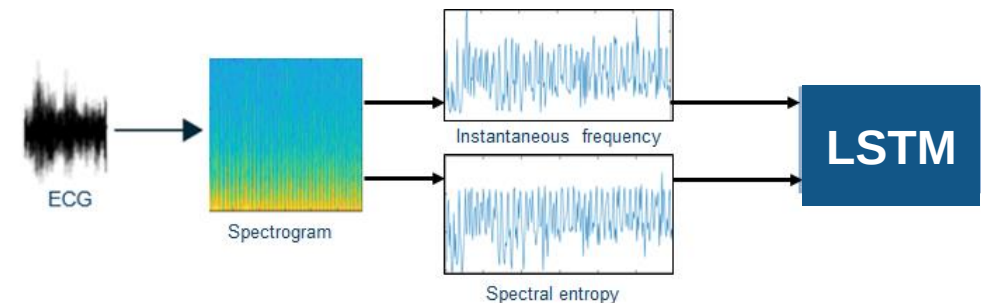
スケール/振幅の異なる信号も
容易に扱える



2-2. Long Short-Term Memory (LSTM) のアプローチ

- Raw データ、あるいは信号特徴量の抽出
- ネットワークパラメータの学習

利点: 適切な特徴量であれば高精度

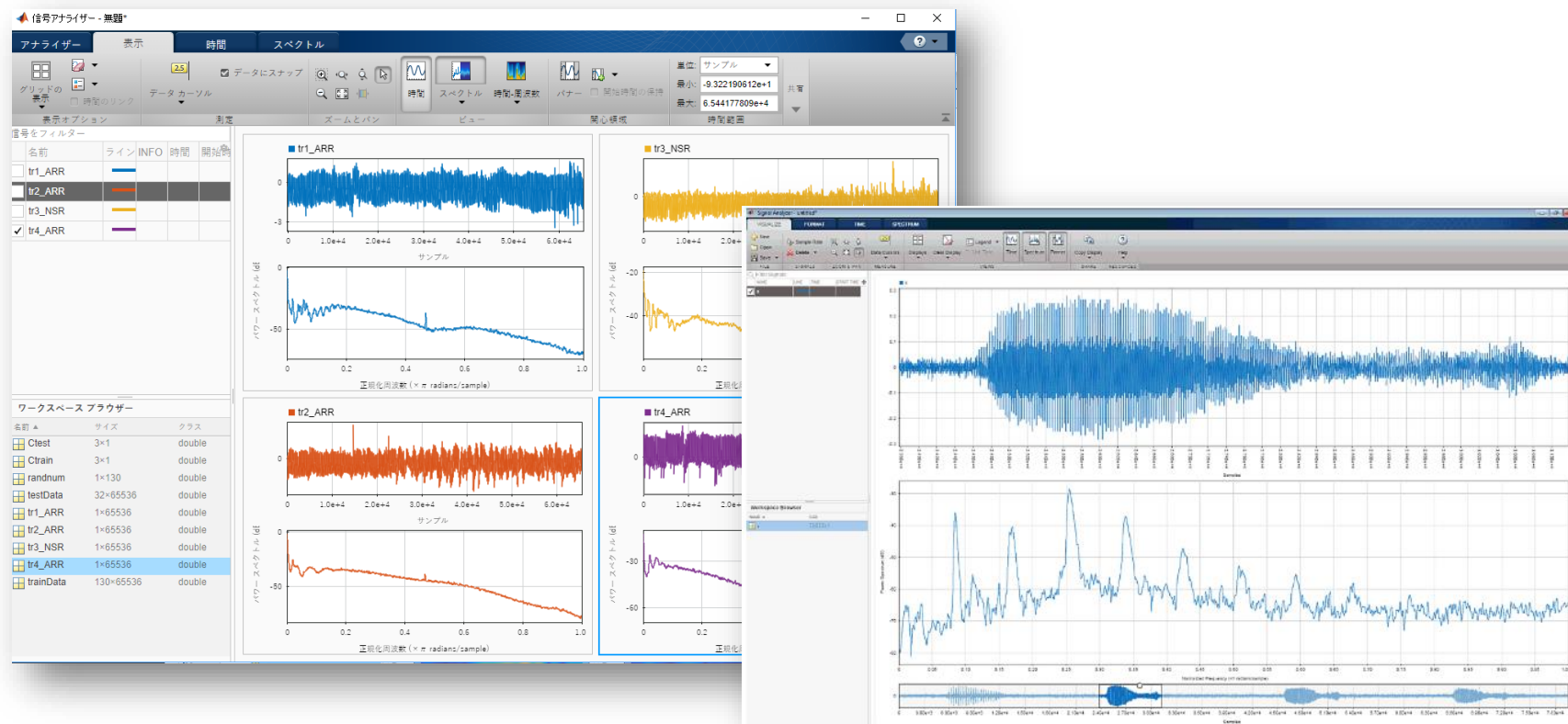


信号アナライザー

センサー信号の特徴の確認と抽出



- 信号とスペクトルの可視化や比較ができるアプリ



R2015a

R2017a



分類学習器

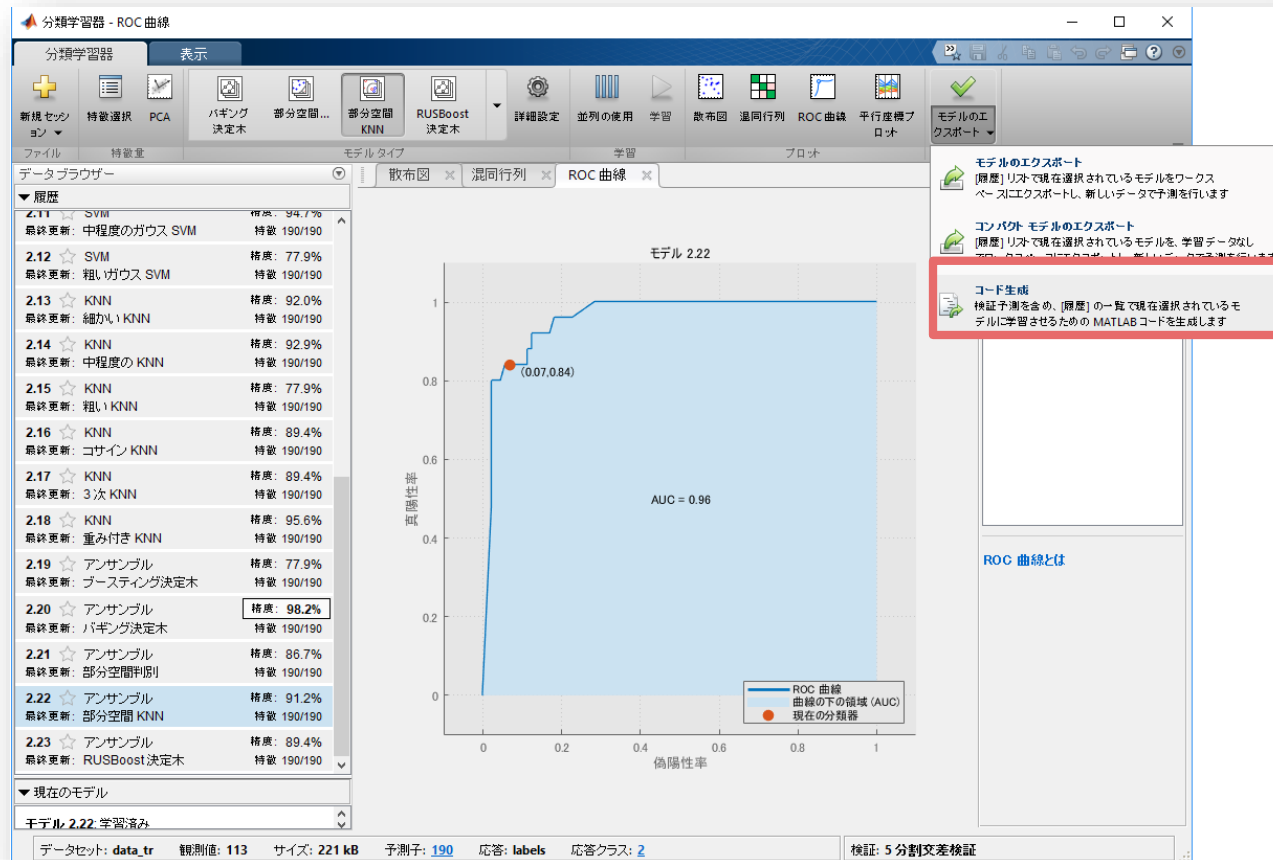


回帰学習器

分類学習器

機械学習アルゴリズムの比較

- データを分類するためのモデル学習と比較



```

52
53 % 分類器の学習
54 % このコードは、すべての分類器オプションを指定し、分類器に
55 classificationKNN = fitcknn(...
56     predictors, ...
57     response, ...
58     'Distance', 'Euclidean', ...
59     'Exponent', [], ...
60     'NumNeighbors', 1, ...
61     'DistanceWeight', 'Equal', ...
62     'Standardize', true, ...
63     'ClassNames', categorical({'urgent'; 'short'; 'medium'; 'lo
64
65 % 関数 predict_ア結果の構造体を作成
  
```

MATLAB プログラムの自動生成



ディープ
ネットワーク ...

ディープネットワークデザイナー

ディープラーニングの転移学習の層変更

- 学習済みネットワークの層のパラメータを変更

The screenshot displays the Deep Learning Network Designer interface. On the left, a network diagram shows a sequence of layers including Inception modules, pooling, and fully connected layers. On the right, the 'ANALYSIS RESULT' table provides details for each layer.

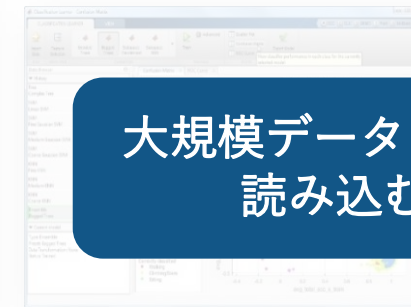
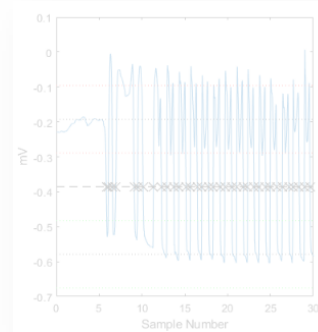
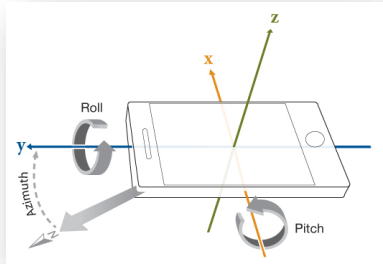
NAME	TYPE	ACTIVATIONS	LEARNABLE
inception_4a-relu_5x5_reduce	ReLU	Unknown	-
inception_4a-5x5	Convolution	Unknown	Weights : Bias
inception_4a-relu_5x5	ReLU	Unknown	-
inception_4a-pool	Max Pooling	Unknown	-
inception_4a-pool_proj	Convolution	Unknown	Weights : Bias
inception_4a-relu_pool_proj	ReLU	Unknown	-
inception_4a-output	Depth concatenation	Unknown	-
inception_4b-1x1	Convolution	Unknown	Weights : Bias
inception_4b-relu_1x1	ReLU	Unknown	-
inception_4b-3x3_reduce	Convolution	Unknown	Weights : Bias
inception_4b-relu_3x3_reduce	ReLU	Unknown	-
inception_4b-3x3	Convolution	Unknown	Weights : Bias

The 'Issues' panel on the right lists several problems with the network:

- Network: Missing input layer. The network must have one input layer.
- conv1-7x7_s2: Missing input. Each layer input must be connected to the output of another layer.
- inception_4b-3x3: Unused output. Each layer output must be connected to the input of another layer.
- inception_4b-relu_3x3: Missing input. Each layer input must be connected to the output of another layer.

The 'Properties' panel on the right shows the configuration for the 'FullyConnectedLayer' (fc7), including parameters like Name, InputSize, OutputSize, Weights, Bias, WeightLearnRateFactor, WeightL2Factor, BiasLearnRateFactor, and BiasL2Factor.

センサーデータ解析特有の課題への対策



大規模データに対する処理を
読み込む前に指定

データ
アクセス

前処理

特徴抽出

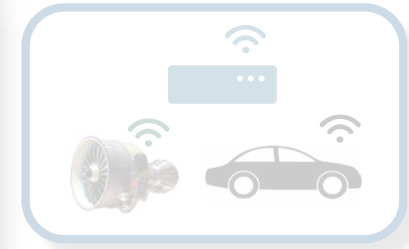
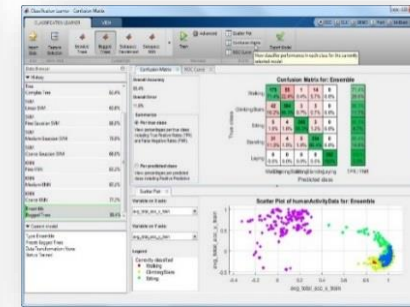
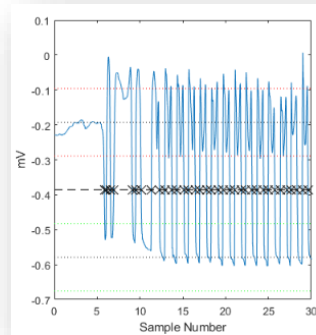
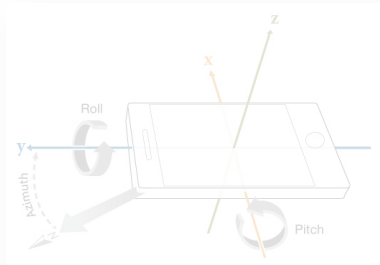
モデルの構築

実環境展開



```
>> imds = imageDatastore( fullfile(foldername, ...  
                                {'classA', 'classB'}), 'LabelSource', 'foldernames');  
  
>> [train_imds, test_imds] = splitEachLabel(imds, 0.7);
```

センサーデータ解析特有の課題への対策



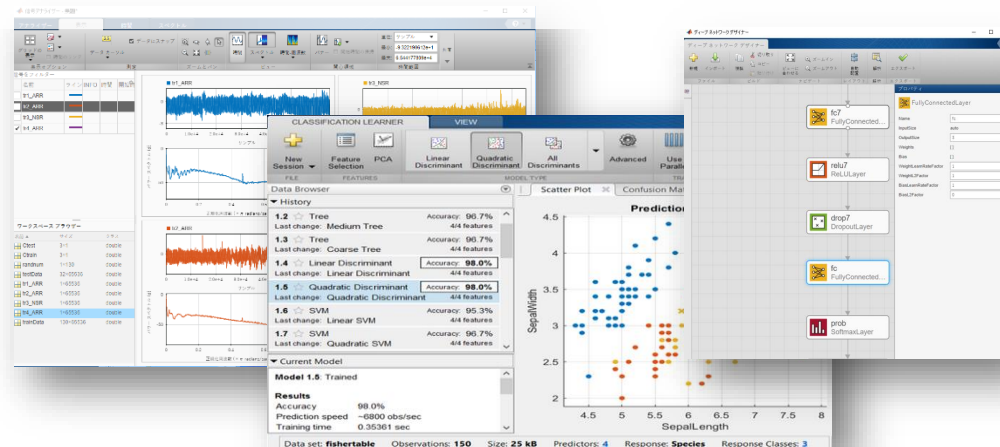
データ
アクセス

前処理

特徴抽出

モデルの構築

実環境展開



アプリとコーディングの
組み合わせで解析の効率化

本日のトピック

■ センサーデータの分類

- 手法の選択 – 従来の機械学習手法とディープラーニング
- センサーデータ解析特有の課題
- 分類モデルの選択

■ テキストの類似性分析

- テキストからの特徴抽出
- テキストデータ特有の課題
- テキスト解析ワークフロー
- 機械学習手法とディープラーニングの使い分け

テキスト解析のワークフロー

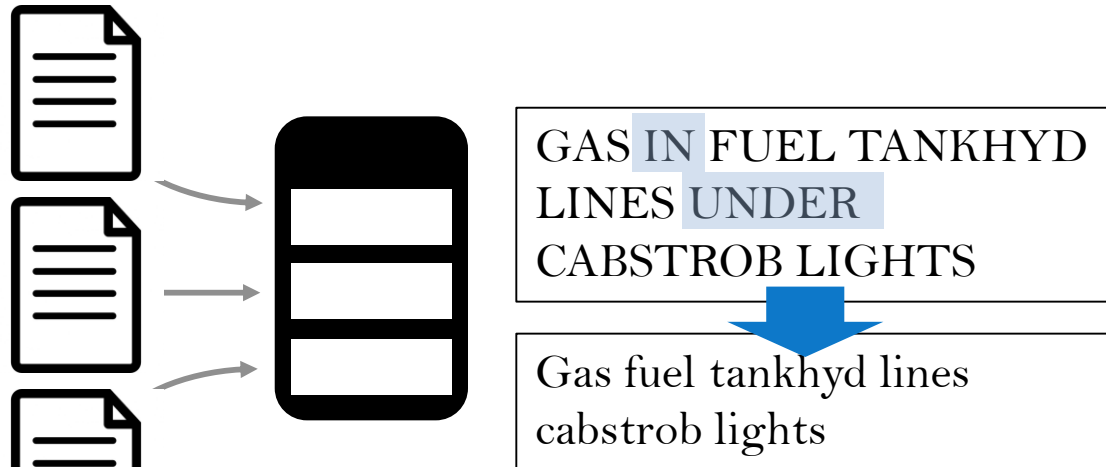
データアクセス

データの前処理

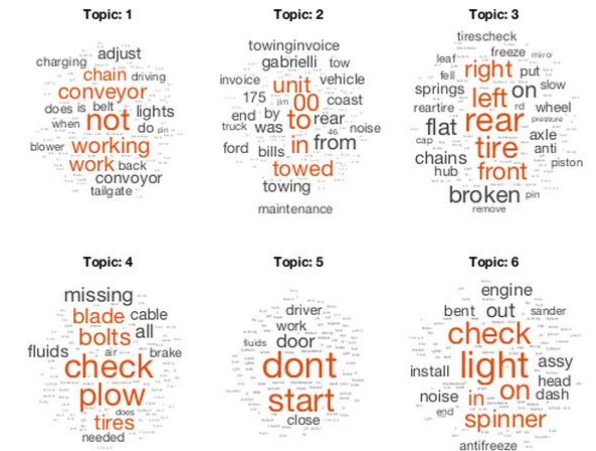
予測モデルの構築

テキストの整形

テキストの数値化



	gas	fuel	tank	hyd
doc1	1	0	1	0
doc2	1	1	0	1



テキストを解析する際の課題

1. 単語によって文字数が異なる

```
num = [1 3 5];
```

```
str = ["チーズ" "電車" "たこやき"];
```

2. 出現する単語の数が膨大

3. 目的によっては不要な単語が存在（句読点、記号、...）

- 彼が、私のたこやきを食べる ☹

4. 複数言語が混合

- この **Simulation** の結果を教えてください。

例2: コミュニティサイトで類似質問の表示

- MATLAB Answers – 技術的な Q&A サイト

The screenshot displays the MATLAB Answers website interface. The top navigation bar includes the MATLAB Answers logo, a search bar, and a language dropdown menu. The left sidebar contains filters for language (日本語, 英語), status (回答済み, 回答採用済み, 未回答), source (MathWorks Support, コミュニティ), and product (MATLAB). The main content area shows a list of questions under the heading '最近の追加'. One question is highlighted, showing its title 'HOF特徴量を出す関数がありますか。', author 'Maya Nakamura', and details such as the date and time of the question, the number of views, and the status of the answer.

MATLAB Answers™

Clear Filt... x Search Answers Answers Q

絞り込み 閉じる

言語 クリア

✓ 日本語
英語

ステータス

回答済み	2367
回答採用済み	2156
未回答	113

ソース

MathWorks Support	1657
コミュニティ	823

製品

MATLAB	785
--------	-----

最近の追加

並べ替え: 更新日 (新しい順) ▼

1 回答

MATLAB関数のdll化について

Masamichi Tanaka さんによって質問されました

最新アクティビティ michio STAFF さんによって

タグ dll化、inpolygon、日本語

製品 MATLAB Compiler、MATLAB Compiler S

0 回答

augmentedImageDatastore を通

MATLAB Answers™

Search Answers Answer

MN

+ フォロー 投票 0 フラグ 編集 閉じる 削除

HOF特徴量を出す関数がありますか。

Maya Nakamura さんによって質問されました 2018 年 10 月 3 日 11:53

最新アクティビティ Maya Nakamura さんによって コメントされました 2018 年 10 月 4 日 2:05

Hirokazu Tanaka さんの 回答が採用されました

23 ビュー (過去 30 日間)

HOF(Histograms of Optical Flow)を出す関数はMATLABに入っていますか。入っていない場合、どのようにプログラムを組めばよいでしょうか。

タグ

特徴量

オブティカルフ

日本語

製品

製品が未入力で

リリース

No release entered yet.

- ユーザコミュニティQ&Aサイトで類似度の高い過去の質問を表示することでユーザが回答にたどり着くまでの時間を短縮

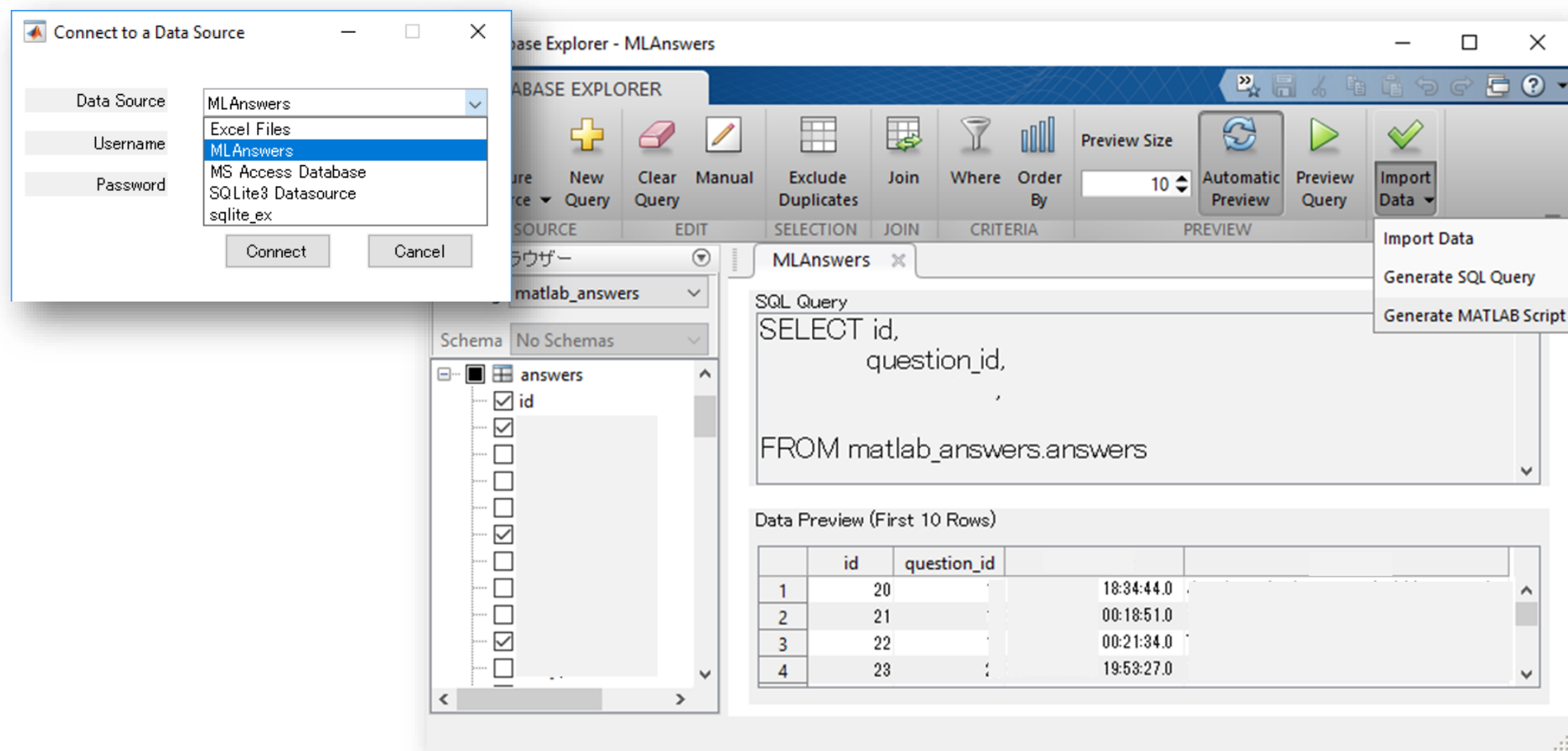
- データベースからデータの取得
- テキストの前処理
- 文書の類似度を計算
- 新規文書に対して類似度の高い質問をおすすめ



Database
Explorer

例2: コミュニティサイトで類似質問の表示 データの読み込み

- Database Explorer App を使用したデータベース接続



- 文書セット全体に対する前処理
 - HTML タグ
 - 日本語質問の抽出
- 不要な単語の削除
 - 意味をなさない単語（ストップワード）
 - 記号
 - 短い単語
- 特定の属性の単語の除去
 - 品詞やタイプ（文字、数値、ハッシュタグなど）

例2: コミュニティサイトで類似質問の表示

前処理と解析

- 文書の数値化 **Bag-of-words**

- 文書に単語が何回含まれているかを行列で表現した文書の特徴量

	すもも	も	もも	の	うち
文章 1	1	2	2	1	1
文章 2			...		

- 単語が出現する順序は考慮しない

- 複数単語のセット抽出 **N-gram**

- N文字が連続した文字列
- 例: 機械 学習

- 文章の類似性 **コサイン類似度**

- ベクトルの向きの類似性
- 文章（ベクトル）1 と文章 2 がどれくらい違うか
- 長さの異なるドキュメントも扱うことが可能



テキスト解析のサポート機能

データアクセス

データの前処理

予測モデルの構築

テキストの整形

テキストの数値化

テキストファイル
スプレッドシート
Web

削除/抽出
正規表現

可視化

Word Cloud Text Scatter

Word ドキュメント
PDF

絵文字の抽出
HTML parsing

bag-of-words

潜在意味解析 (LSA)

bag-of-ngrams

潜在的ディリクレ配分法 (LDA)

TF-IDF

単語の分散表現

品詞分解

単語の正規化

R2017b

R2018b 日本語対応

MeCab (3rd party の形態素解析器)
& IPA辞書 に対応

データベース
音声
画像

機械学習アルゴリズム

ディープラーニング

テキスト解析のまとめ

- 単語を数値化
 - Bag-of-words, bag-of-ngrams
- テキスト前処理の多様性
 - 句読点、記号、ストップワードの削除
 - 特定の属性の単語の削除（品詞、タイプなど）
 - 複数単語から成る熟語の使用
 - 可視化や機械学習アルゴリズムの活用
- テキスト解析の前処理部分に日本語が対応



過去に使用した類似の文書を表示することで同じ作業を減らす

Key Takeaways

- 画像だけでなくセンサー/テキストでも機械学習 & ディープラーニングによる解析が可能
- データの理解や手法の使い分けが結果に影響
- アプリ活用による作業の効率化
- お持ちの知識を活かしたデータ解析を！





© 2018 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

補足資料

Learn More

- センサーデータ解析のセクションで参考にした例
 - [Signal Classification with Wavelet-Based Features and Support Vector Machines](#)
 - [Signal Classification with Wavelet Analysis and Convolutional Neural Networks](#)
 - [Classify ECG Signals Using Long Short-Term Memory Networks](#)
- テキスト解析のセクションで参考にした例
 - [Analyze Japanese Text Data](#)

Read more:

- [ディープラーニングを学ぶためのリソースリスト](#)
- [MATLAB で自然言語処理や解析を行う機能はありますか？](#)



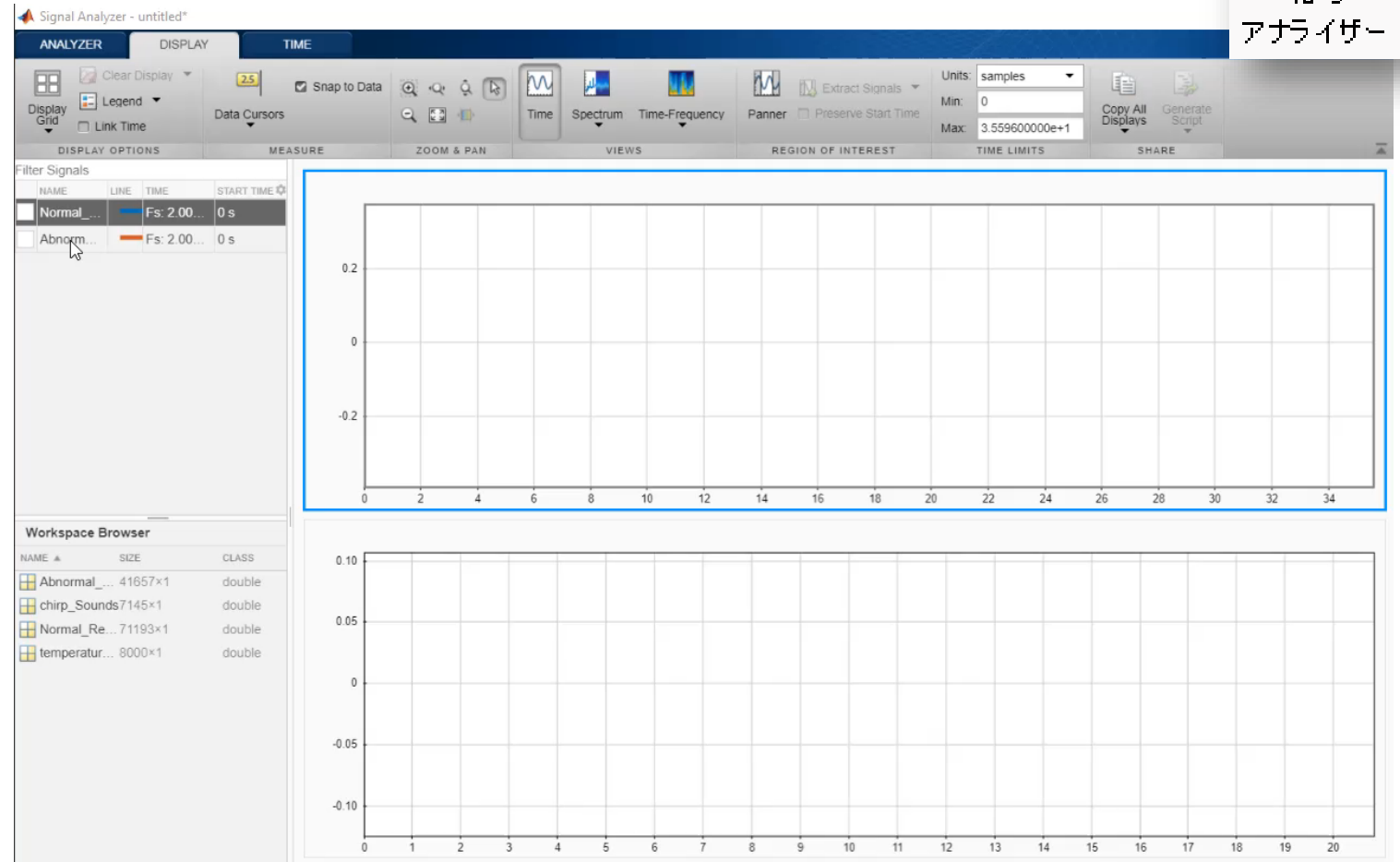
Step 1: データの取得

■ 課題

- 異なるサンプリングレート
- 信号マネジメント
- 大きなデータセット

■ データの可視化

- 時間領域
- 周波数領域
- 時間-周波数領域



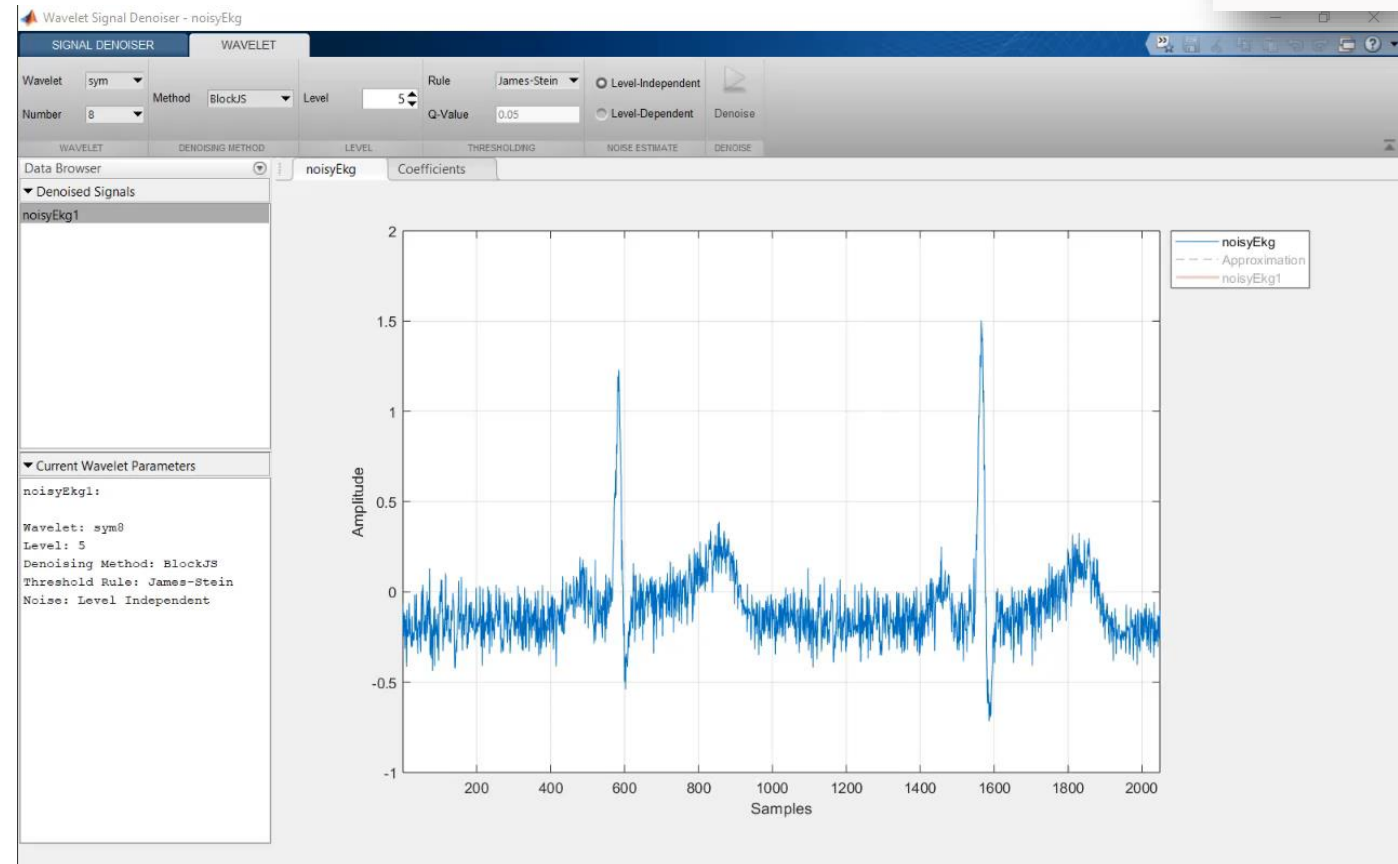
信号アナライザアプリによるセンサー信号の可視化

>> signalAnalyzer

Step 2&3: センサー信号の前処理 & 特徴抽出



- 課題
 - シャープな特徴の維持
 - 信号 & ノイズスペクトルの重なり
- 自動ノイズ除去
- コード上でのカスタマイズ
 - アプリから MATLAB コードの自動生成



センサー信号の前処理

R2015a

R2017a



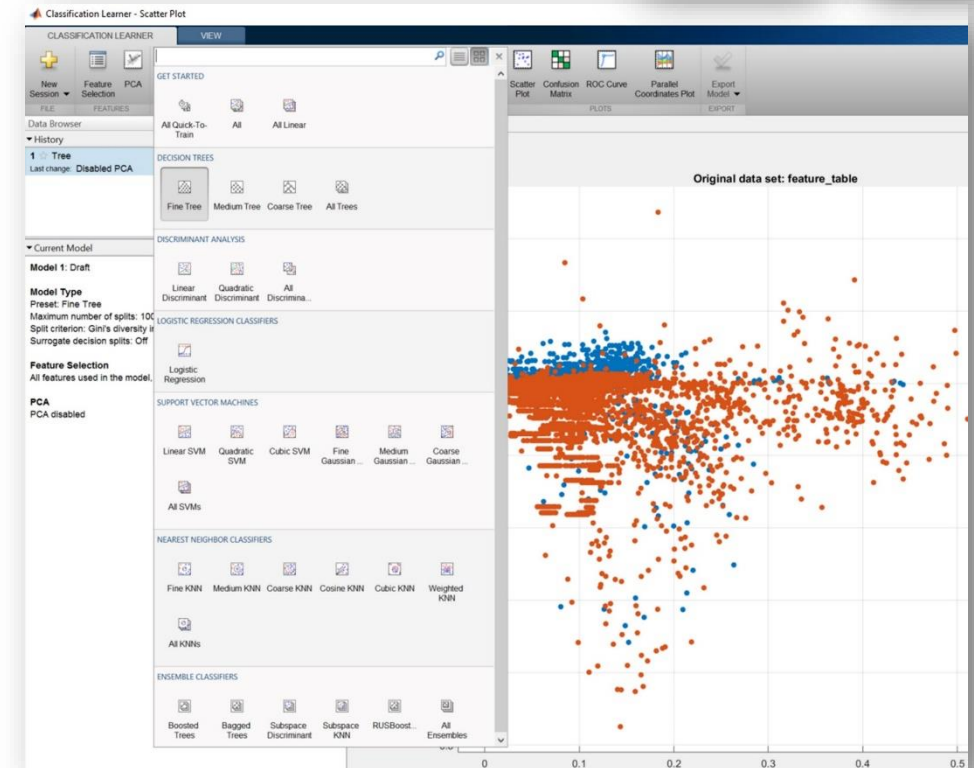
分類学習器



回帰学習器

Step 4-1: モデルの学習

- 課題
 - 機械学習アルゴリズムの知識習得
 - 大規模データセットへのスケーリング
- アプリによるクイック学習
 - クロスバリデーションの定義
 - 一般的なアルゴリズムの比較
 - パフォーマンスの分析
- 大規模データセットへのスケールアップ



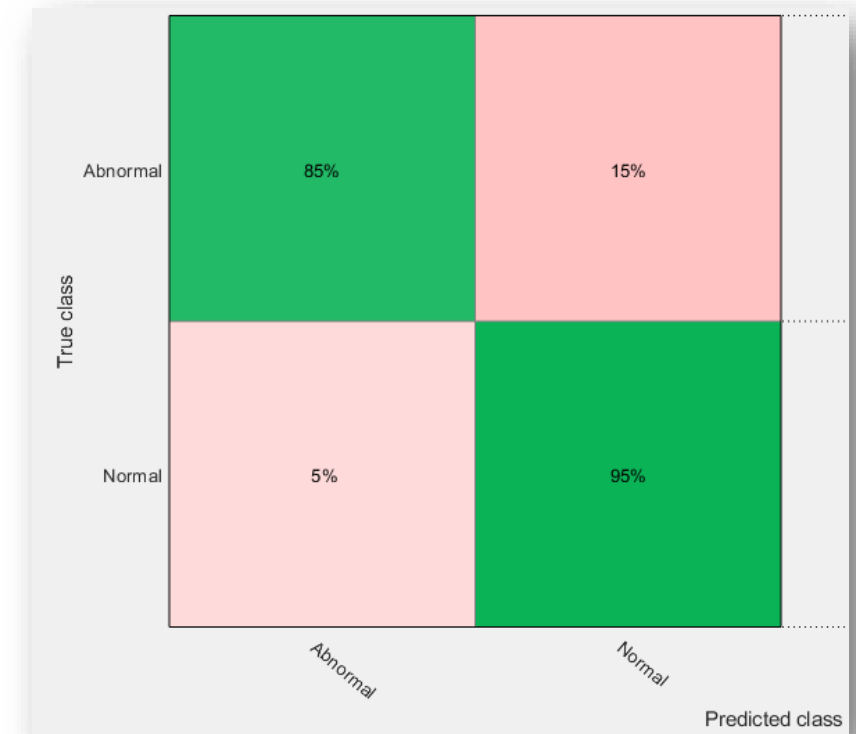
分類学習器によるモデル学習

>> classificationLearner

Step 4-2: パフォーマンス改善

- 課題
 - 手動パラメタチューニングの煩わしさ
 - 追加の改善点の特定
- モデルの反復最適化
 - ベイズ最適化
 - パフォーマンスの分析
 - アンバランスの調整（誤分類のサンプル/重要性）

クラス	割合
正常	75%
異常	25%



Step 5: 実行環境への展開



- 課題
 - 異なるプラットフォームへの展開
 - ハードウェア要件
(サイズ、速度、固定小数点など)

- 展開オプション
 - 組み込みシステムへのコード生成 (C/C++, HDL, GPU)
 - エンタープライズシステムへの展開 (Webブラウザ)