

# 統計・機械学習の理解を促すテクニック

MathWorks Japan

アプリケーションエンジニアリング部 テクニカルコンピューティング

大開 孝文

# アジェンダ

- すぐ使えるデータセットのご紹介
- 統計/機械学習を理解を促す可視化機能
- 例題

# 統計で使えるデータセット (34種類)

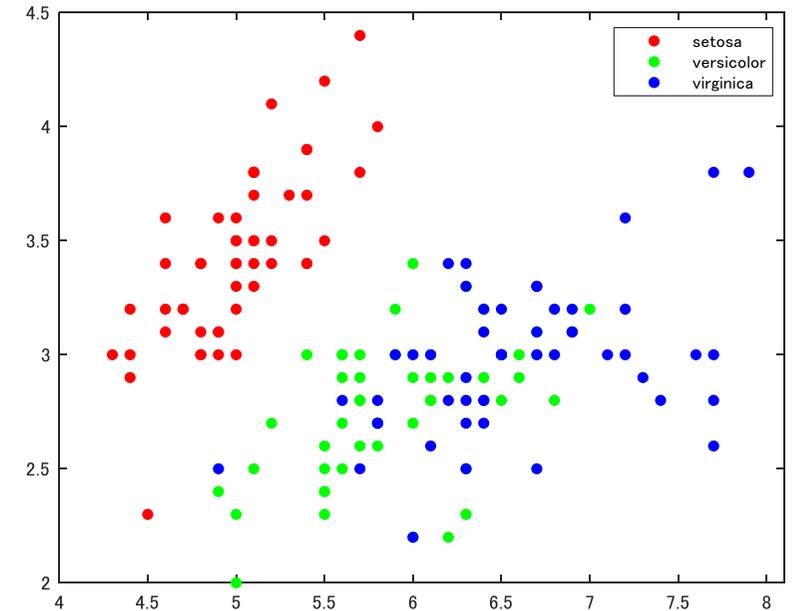
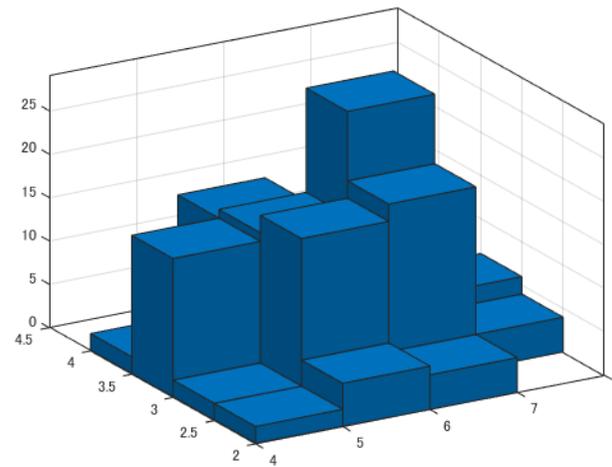
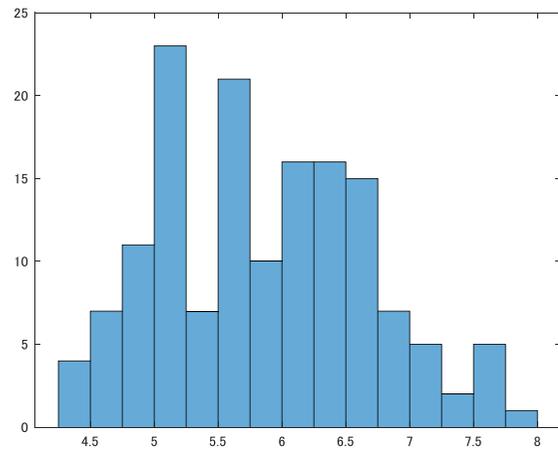
<https://jp.mathworks.com/help/stats/sample-data-sets.html>

ファイル	データセットの説明
acetylene.mat	相関予測子をもつ化学反応データ
arrhythmia.mat	UCI Machine Learning Repository からの心臓不整脈データ
carbig.mat	自動車の寸法 (1970~1982 年)
carsmall.mat	carbig.mat のサブセット。自動車の寸法 (1970、1976、1982 年)
census1994.mat	UCI Machine Learning Repository の Adult データ
cereal.mat	朝食シリアル成分
cities.mat	米国大都市圏の生活の質のランキング
discrim.mat	判別分析のための cities.mat のバージョン
examgrades.mat	0 ~ 100 のテストの採点
fisheriris.mat	フィッシャーのアヤメの分析データ (1936 年)
flu.mat	Google の「インフルトレンド」は米国のさまざまな地域の ILI (インフルエンザ様疾患) の比率を推定、CDC は監視組織のレポートに基づき ILI の比率を重視
gas.mat	1993 年のマサチューセッツ州周辺のガソリン価格
hald.mat	セメントの発熱と混合成分
hogg.mat	出荷別の牛乳内のバクテリア数
hospital.mat	病院のデータのシミュレーション
humanactivity.mat	座る、立つ、歩く、走る、踊る、という 5 種類の動作に対する行動認識データ
imports-85.mat	UCI リポジトリからの 1985 年の自動車輸入データベース
ionosphere.mat	UCI Machine Learning Repository からの電離層データセット
kmeansdata.mat	4 次元にクラスター化したデータ
lawdata.mat	法律学校 15 校の GPA と LSAT 得点
mileage.mat	2 社の 3 モデルの車両のマイレージ データ
moore.mat	5 つの予測子に基づく生物化学的酸素要求量
morse.mat	非符号器によるモールス符号の区別の認識
nlpdata.mat	MathWorks® ドキュメンテーションから抽出した自然言語処理データ
ovariancancer.mat	4000 個の予測子に基づきグループ分けされた意見
parts.mat	36 個の円形パーツの寸法のずれ
polydata.mat	多項式近似用の標本データ
popcorn.mat	タイプ別とブランド別によるポップコーン生産量
reaction.mat	Hougen-Watson モデルのための反応力学
sat.dat	性別とテスト別の大学進学適性試験平均値 (表)
sat2.dat	性別とテスト別の大学進学適性試験平均値 (csv)
spectra.mat	60 件のガソリン標本の NIR スペクトル観測とオクタン価
stockreturns.mat	株式収益のシミュレーション

# 統計/機械学習を理解を促す可視化機能

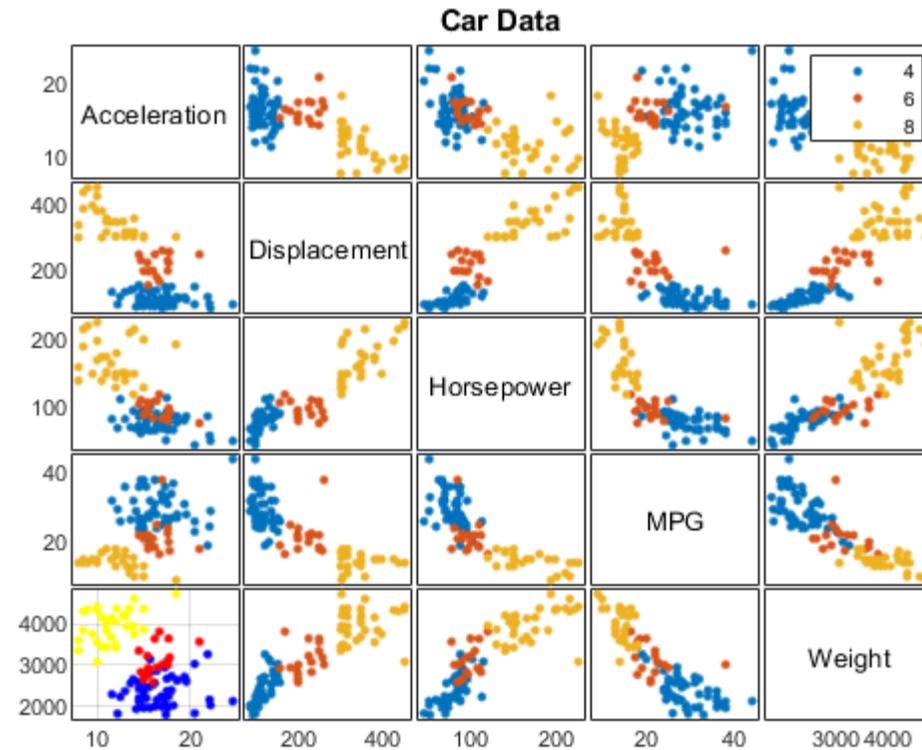
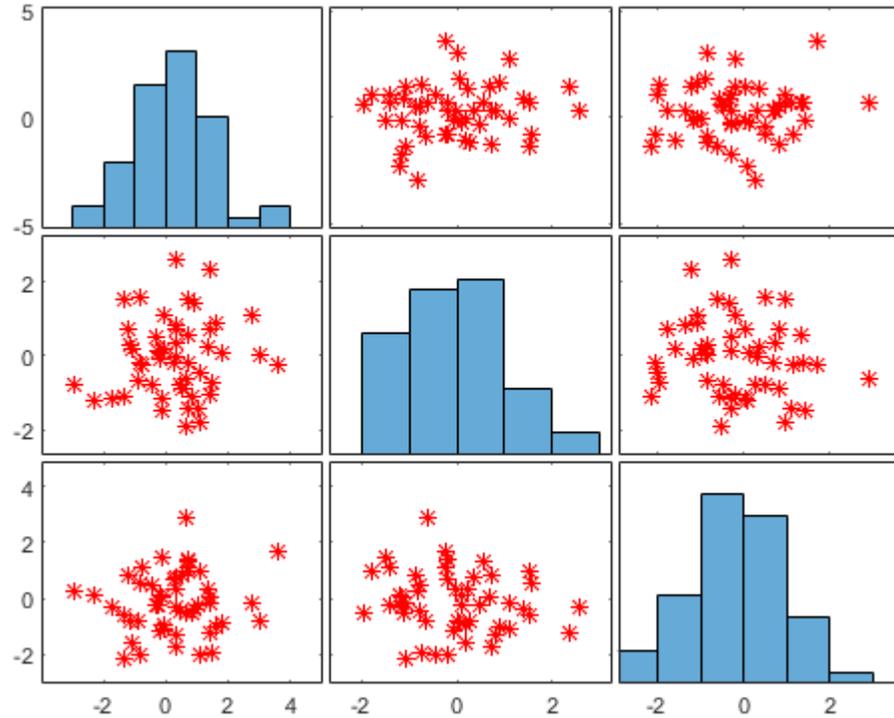
# histogram / histogram2 / gscatter

- ヒストグラム



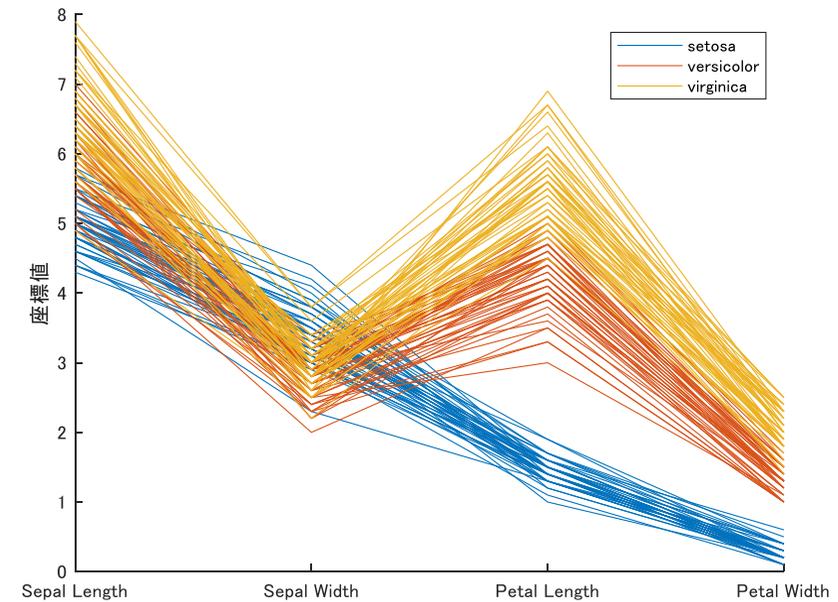
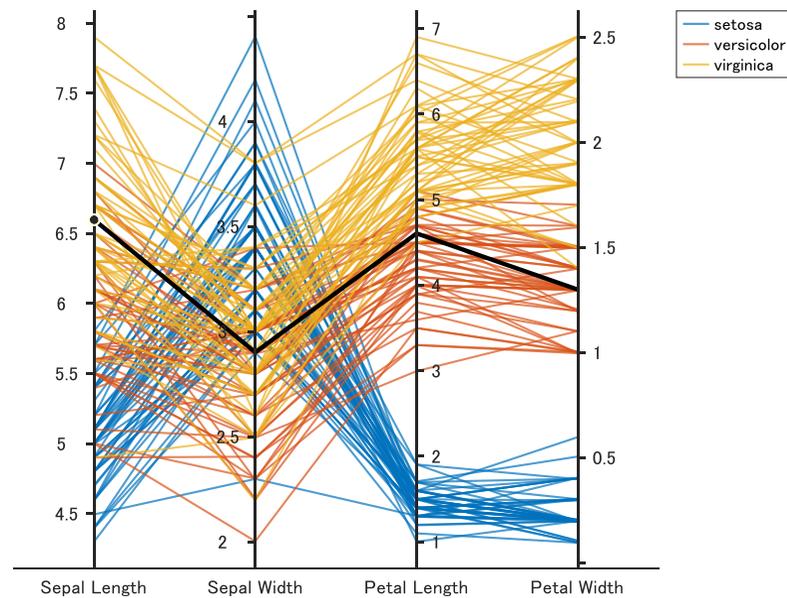
# plotmatrix / gplotmatrix

- 散布図プロット行列



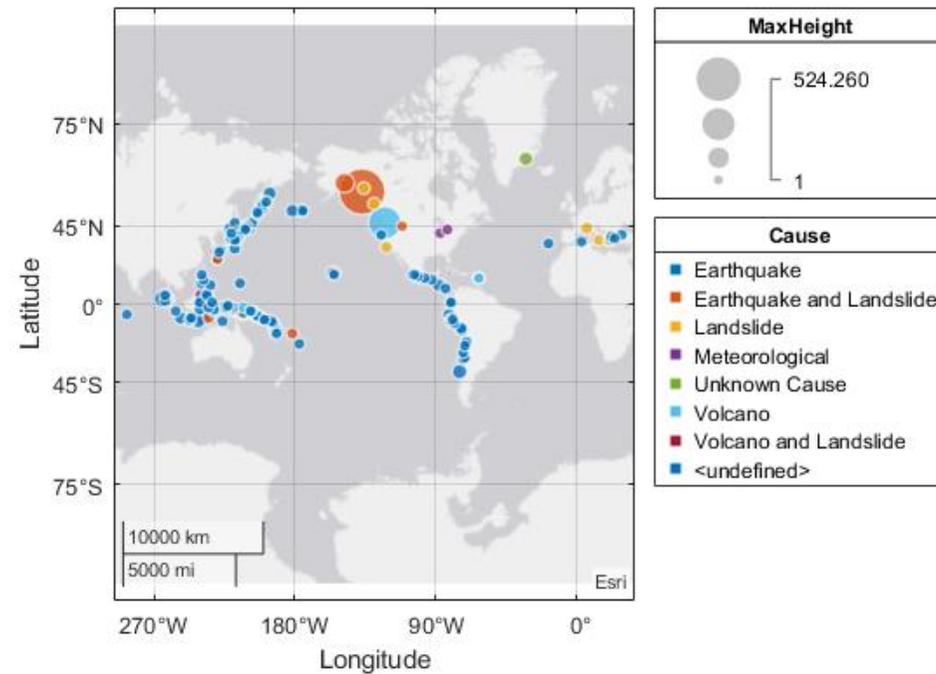
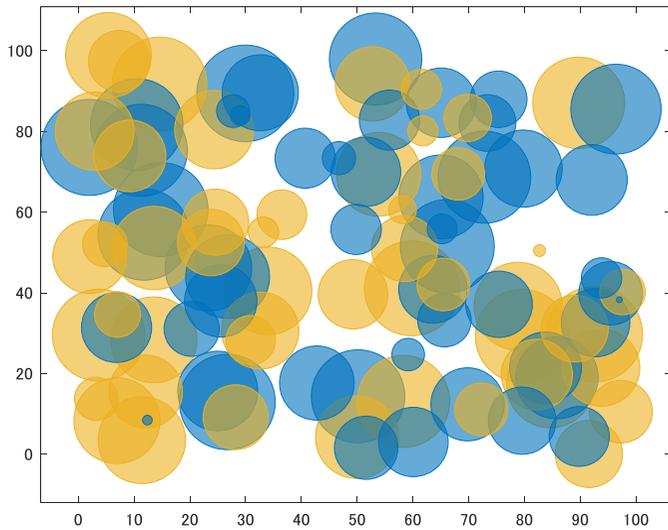
# parallelplot / parallelcoords

- 平行座標プロットを作成



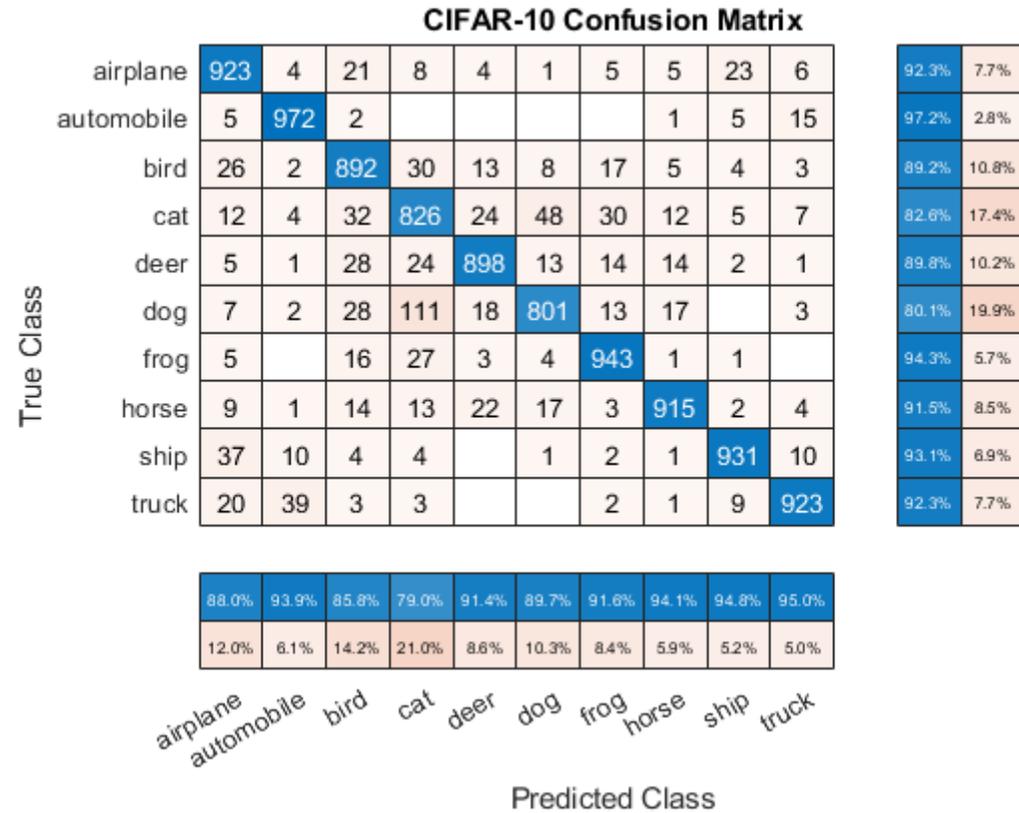
# bubblechart/ geobubble

- バブルチャート



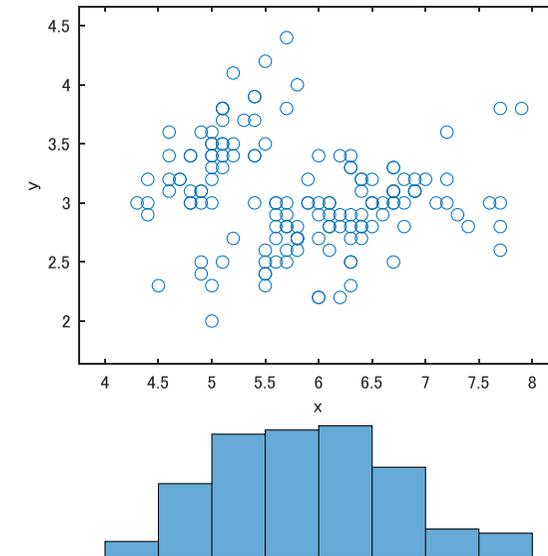
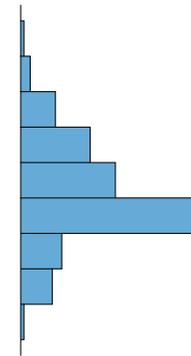
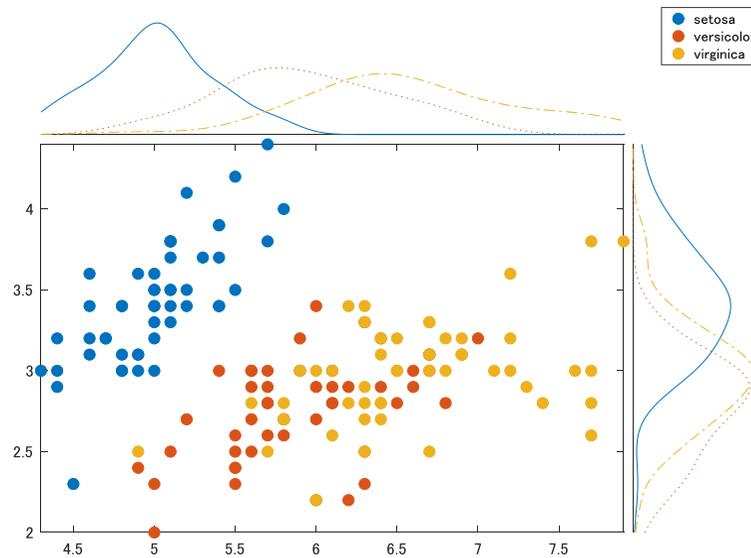
# confusionchart

- 混同行列チャートの作成
  - 分類モデルの精度の可視化



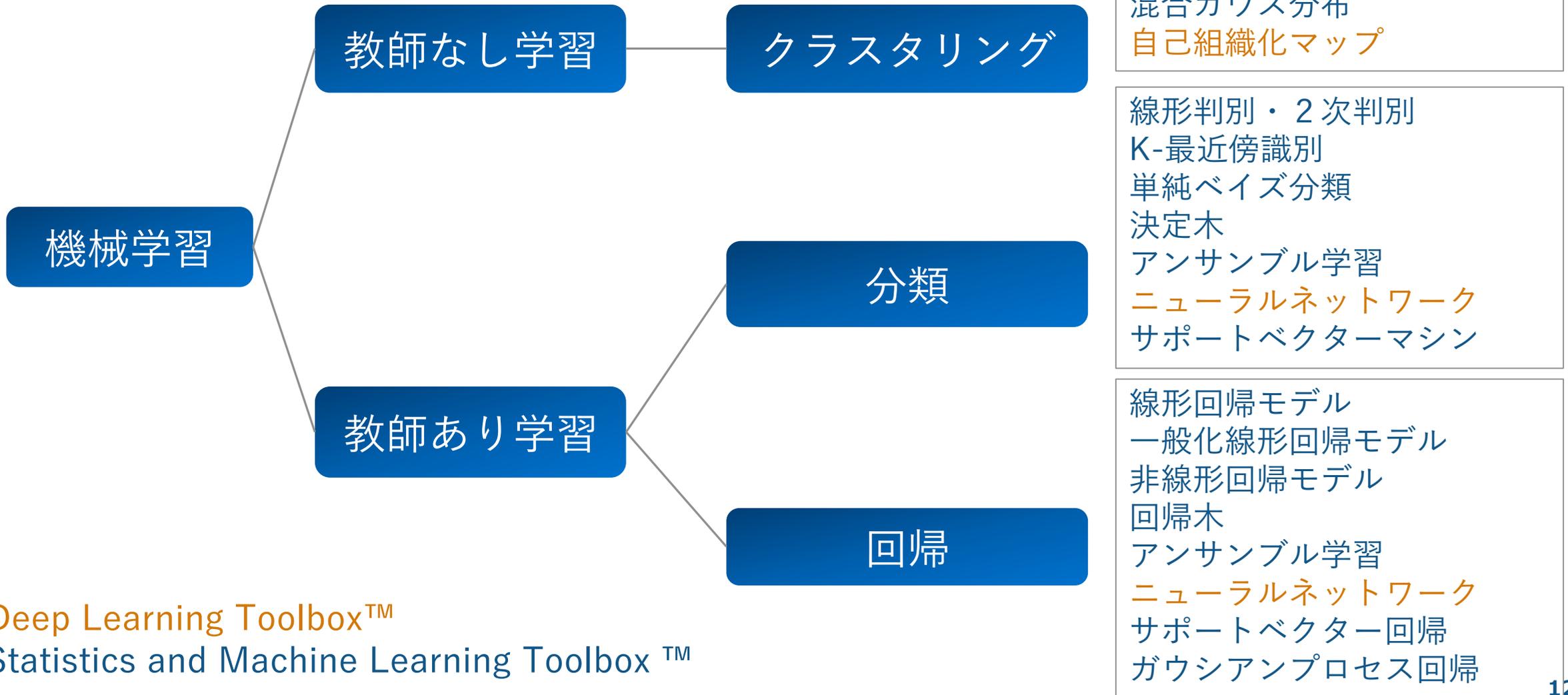
# scatterhistogram / scatterhist

- 周辺ヒストグラムをもつ散布図



# 例題

# 機械学習アルゴリズム



Deep Learning Toolbox™

Statistics and Machine Learning Toolbox™

# AI/機械学習モデルの作成

データのアクセスと探索

ファイル



データベース

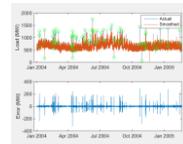


センサー



データの前処理

データクレンジング



データ削減 / 変換



特徴抽出



予測モデル開発

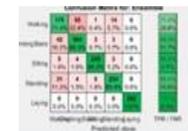
AIモデル学習



パラメータ最適化



モデル検証



# SVMのガウシアンカーネル

## MATLAB Answers

MATLAB Central ▾ | ホーム | My MATLAB Answers ▾ | 質問する | 回答する | ブラウズ | その他 ▾

## Gaussian kernel scale for RBF SVM

⊕ フォロー

207 ビュー (過去 30 日間)



vince65201 2017 年 3 月 8 日

👍 投票 | 0 | 🚩 フラグ | 🗨️ 翻訳

コメント済み: Walter Roberson **MVP** 2019 年 6 月 1 日

採用された回答: Akira Agata **STAFF**

✅ 採用された回答

Hi All, I'm using RBF SVM from the classific anyone knows how Matlab came up with the predictors. Additionally, it says "The software mean the kernel scale is simply  $1/\sigma^2$ ? <https://www.mathworks.com/help/stats/temp>



Akira Agata **STAFF** 🌟 2017 年 3 月 9 日

👍 投票の取り消し | 3 | 🔗 リンク | 🚩 フラグ | 🗨️ 翻訳

Assuming the RBF kernel function with scaling parameter (gamma) as follows:

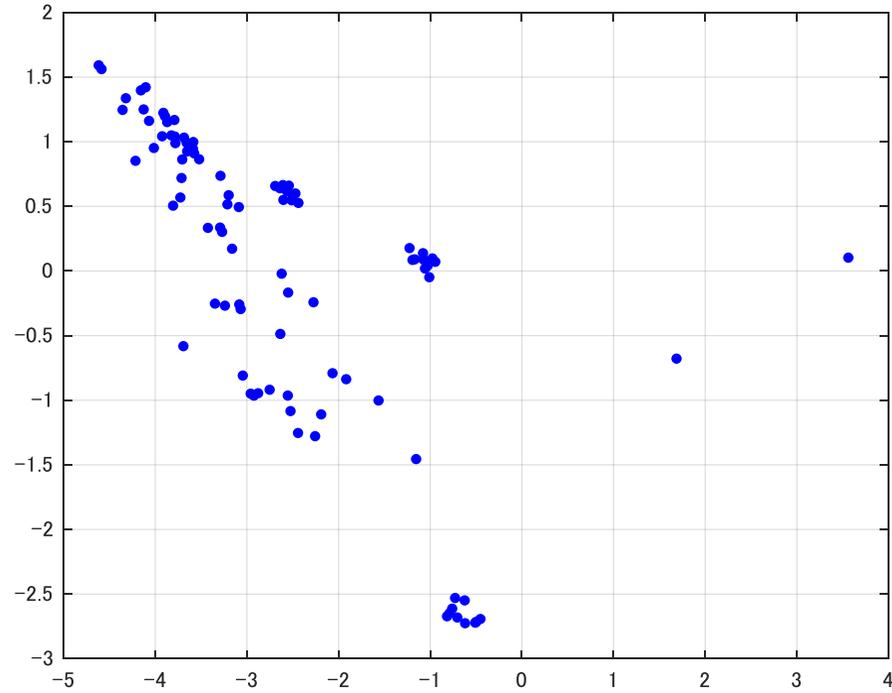
$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

Then, the SVM model should be set using "KernelScale" like this.

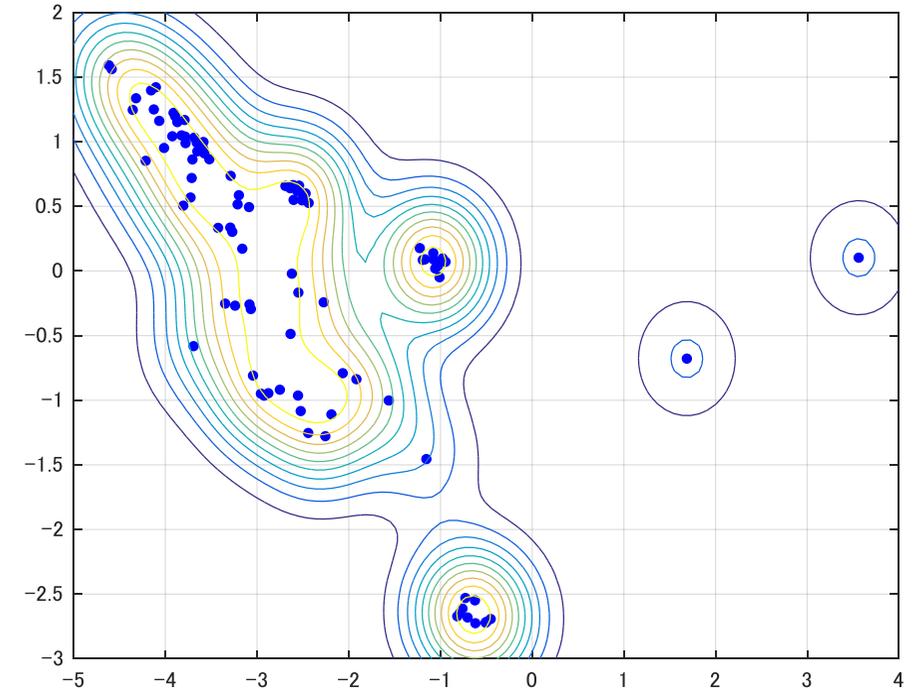
```
mdlsvm = fitcsvm(..., 'KernelScale', 1/sqrt(gamma));
```

# 1-Class SVM とは？

カーネル法を使った正常度スコアの推定アルゴリズム



異常データを含むデータの例



スコアの値が高い方が正常度が高い

1-Class SVM

## 授業開発についてご質問・ご要望ございましたら

お問い合わせ先

MathWorks Japan

Education Customer Success Engineering 部

Email: [cse-jp@groups.mathworks.com](mailto:cse-jp@groups.mathworks.com)