

# 教師なし学習

正田 備也

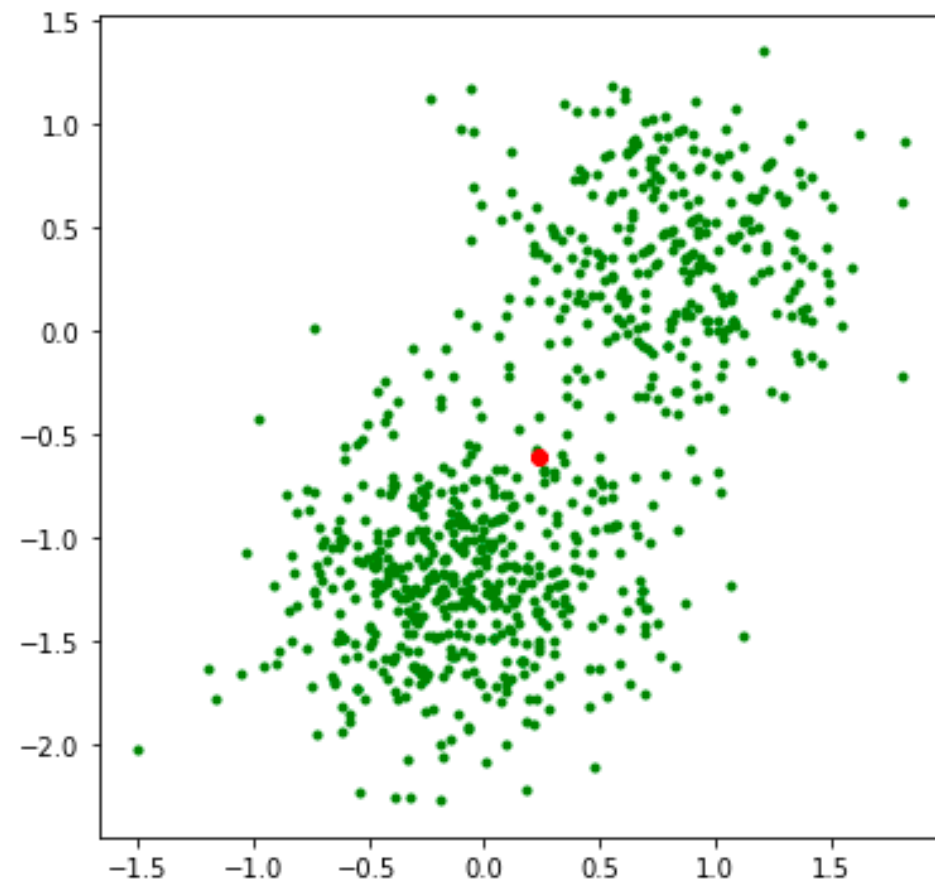
[masada@rikkyo.ac.jp](mailto:masada@rikkyo.ac.jp)

## 追加：多値分類について

- 分類先が 3 つ以上あっても . . .
- `sklearn.linear_model.LogisticRegression` を使えばいいだけ
- SVC など他の分類手法も同様

# データ集合を要約する

- 平均ベクトルで要約する
  - すべてのベクトルの和を求めて個数で割る



例) 右図のような2次元ベクトルの集まり

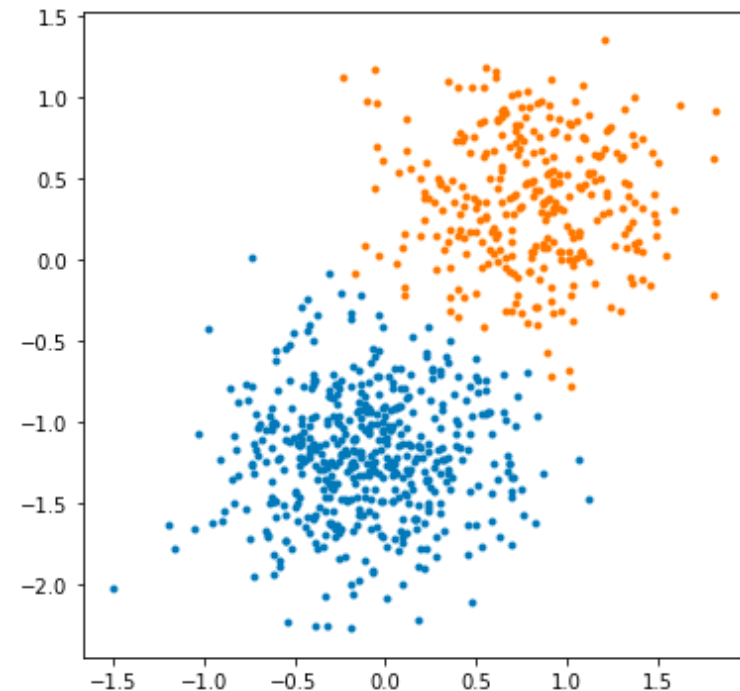
- 平均は、全体をうまく要約していると言えるか？
- データがひとつにまとまって分布していないように見える
- どうする？

# 教師なし学習 (unsupervised learning)

- 与えられているのは、入力ベクトルの集まりだけ

$$\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$$

- 正解となるクラスラベル $y_i$ は対応づけられていない
- そうであっても・・・
- データの自然なまとまりを見つけたい
  - 右図はK-meansクラスタリングの実際の結果。
- そこで、教師なし学習の出番！
  - 次元圧縮も、もうひとつの重要な用途。



# 階層的クラスタリング

- 入力ベクトルがバラバラの状態から始める
  - つまり、どのクラスタも1つのデータ点だけを含む状態から始める。
- 一番距離が近いクラスタのペアをマージする
  - どのような基準でクラスタのペアを選ぶかで、クラスタリング結果が違ってくる。
- 「一番距離が近い」というところの近さの基準をどうするか？
  - これをlinkage criterionと呼ぶ

# linkage criterion

どの2つのクラスタをマージするか？

## 1. Ward criterion (分散を利用した方法)

- クラスタの重心からの距離の2乗の和を、すべてのクラスタにわたって合算したものが、最小になるように、クラスタをマージする。

## 2. Complete linkage criterion (最大値を最小化する方法)

- 異なるクラスタに属するデータ点の距離の最大値が最小のクラスタをマージする

## 3. Average linkage criterion (平均値を最小化する方法)

- 異なるクラスタに属するデータ点の距離の平均値が最小のクラスタをマージする

## 4. Single linkage criterion (最小値を最小化する方法)

- 異なるクラスタに属するデータ点の距離の最小値が最小のクラスタをマージする

# 階層的クラスタリングの応用例

- 単語ベクトルをクラスタリング
  - Dan Jurafsky and James H. Martin.  
Speech and Language Processing  
(3rd ed. draft), Section 6.9.
  - 元々階層構造や派生構造をもつ対象に  
使うと、うまくいくことがある。
  - 意味や範疇が近い単語は階層構造の上  
でも近くに位置している（???)

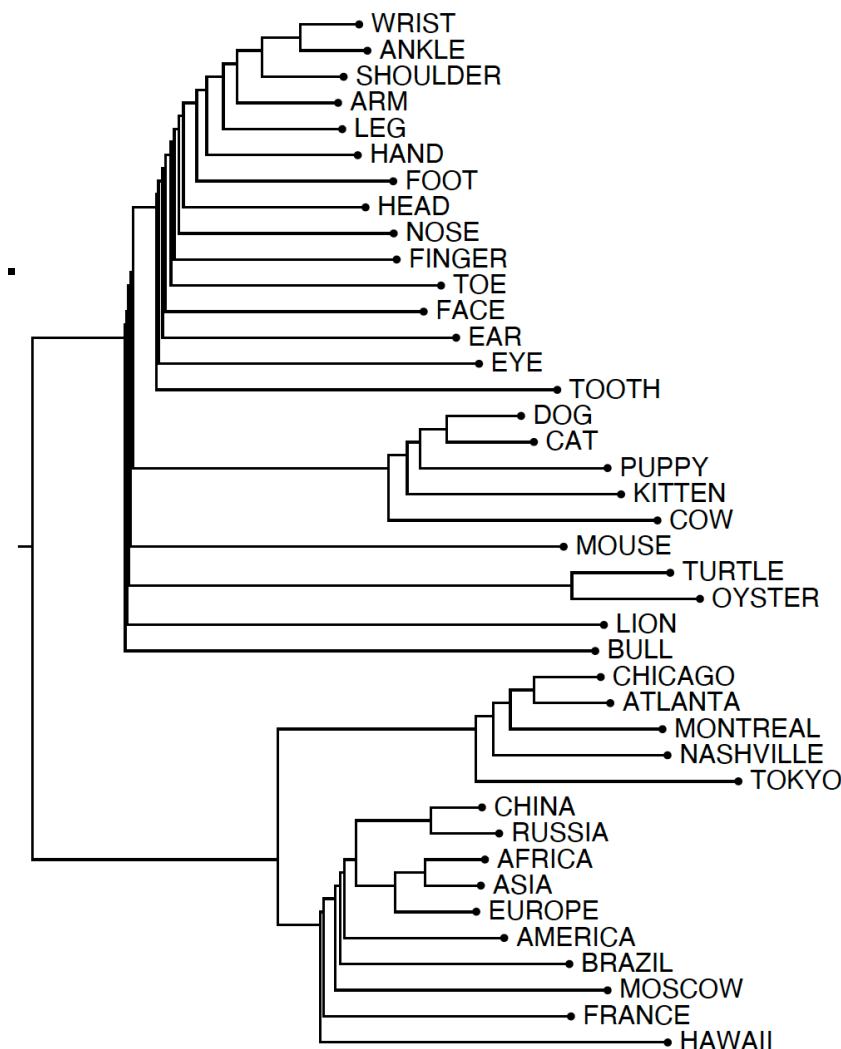
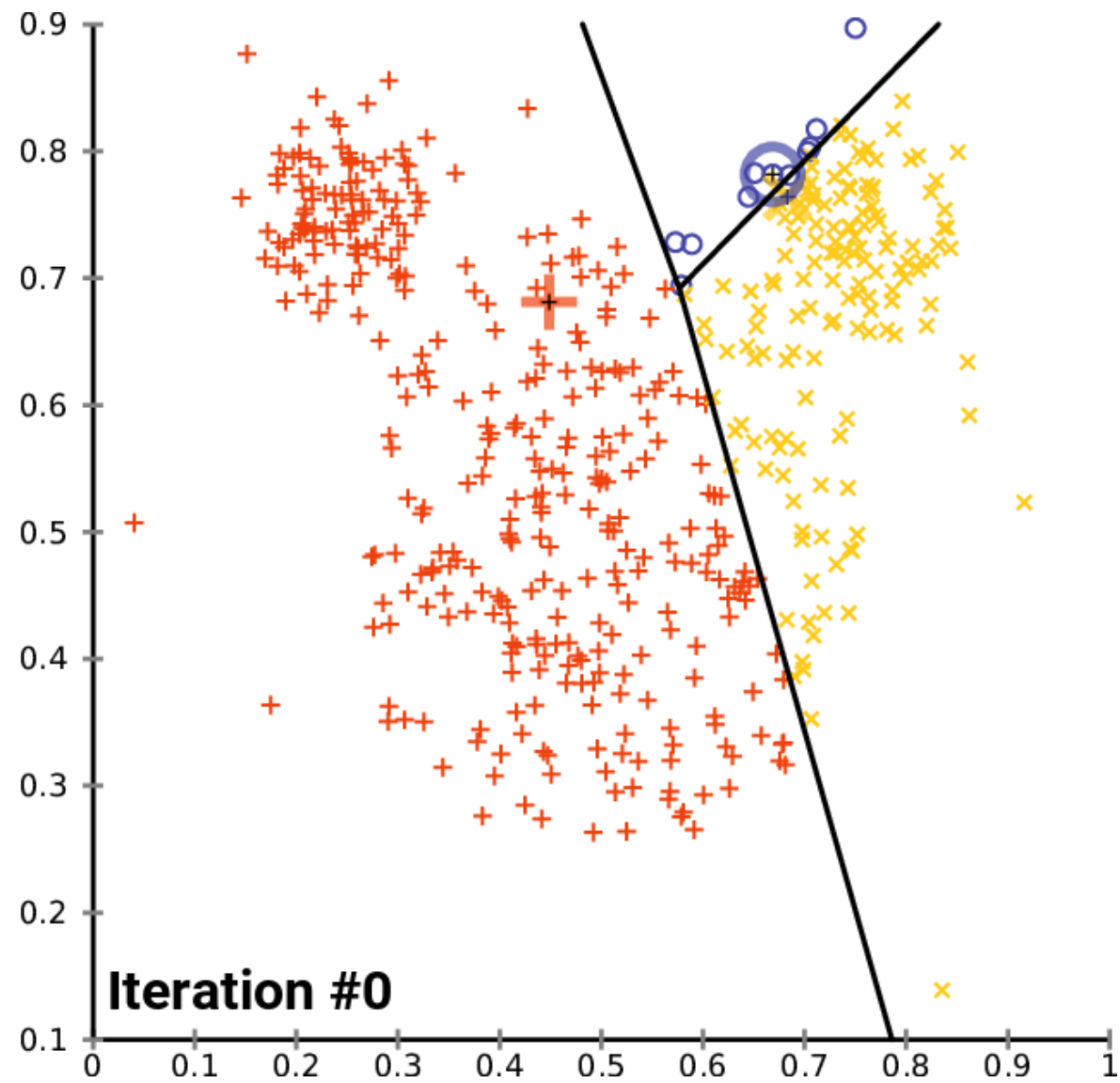


Figure 9: Hierarchical clustering for three noun classes using distances based on vector correlations.

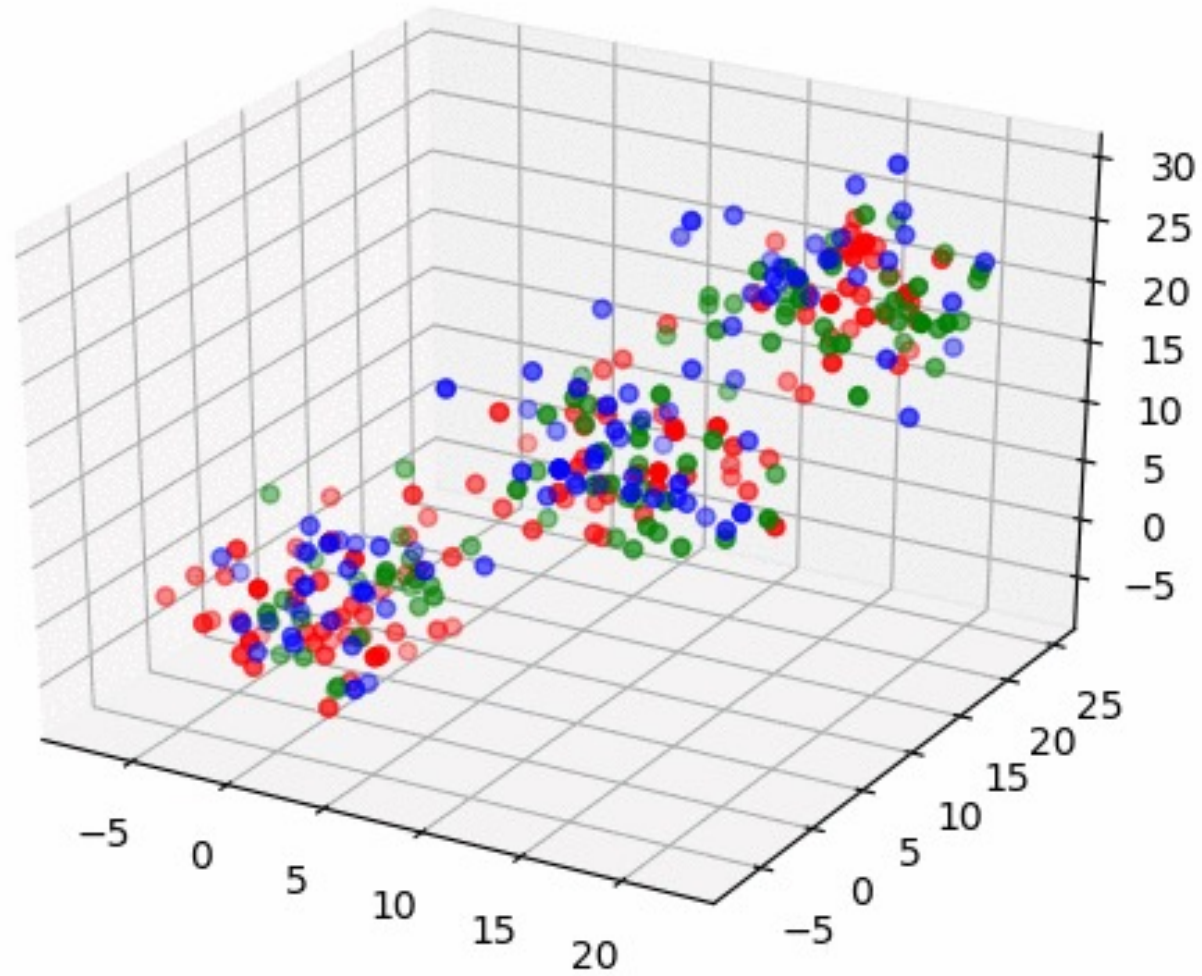
# K-平均法 (K-Means)

- とても良く知られているクラスタリング手法
- 与えられたベクトル集合を $K$ 個の排他的なクラスタに分ける
  1.  $K$ 個のクラスタの重心(centroid)を適切に初期化
  2.  $N$ 個のベクトルの各々を、最も近い重心に対応するクラスタへ属させる ( $K$ 個の重心までの距離を求め、最小値を探す)
  3. 構成されたクラスタの重心を求め直す
  4. クラスタの重心がまだ変動しているなら、2. へ戻る





[https://commons.wikimedia.org/wiki/File:K-means\\_convergence.gif](https://commons.wikimedia.org/wiki/File:K-means_convergence.gif)



# クラスタリング手法の出力

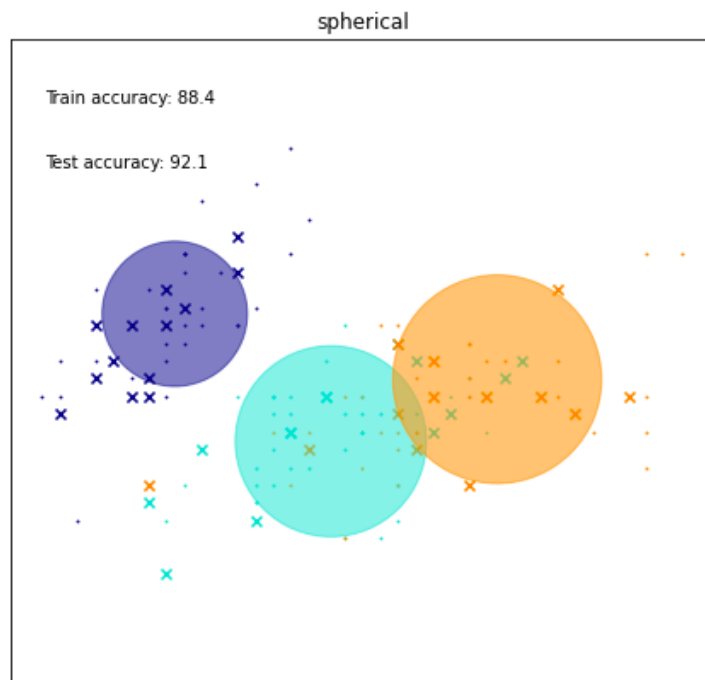
- model-basedなクラスタリングは2種類の情報を出力
  1. 各入力ベクトルのクラスタへの所属の情報
  2. クラスタの数理的な表現（例えば重心）
- K-平均法の場合
  1. 各入力ベクトルは、ちょうどひとつのクラスタに属する
  2. 各クラスタは、その重心で表現される

# 混合ガウス分布によるクラスタリング

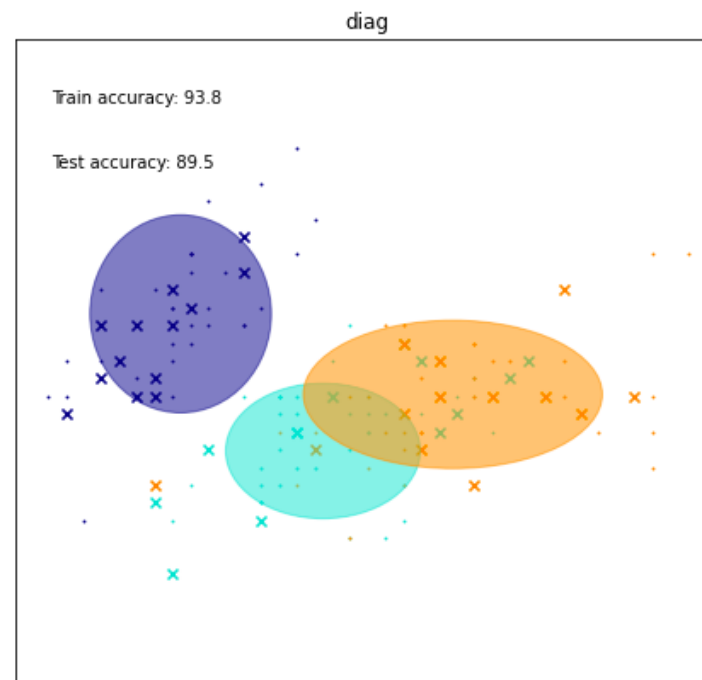
- K-平均法の仮定を緩めた手法
  - 詳細は秋学期の「統計モデリング1」で。
- 1. 入力ベクトルはちょうど一つのクラスタに属さなくてよい
  - 各クラスタへの所属確率（所属の度合い）が得られる
- 2. 各クラスタの分布は同心円状の広がりを持っていないくてよい
  - 横に広がった円でもいいし、傾いていてもいい
    - クラスタ形状の情報は多変量正規分布の共分散行列によって与えられる

cf. [https://scikit-learn.org/stable/auto\\_examples/mixture/plot\\_gmm\\_covariances.html](https://scikit-learn.org/stable/auto_examples/mixture/plot_gmm_covariances.html)

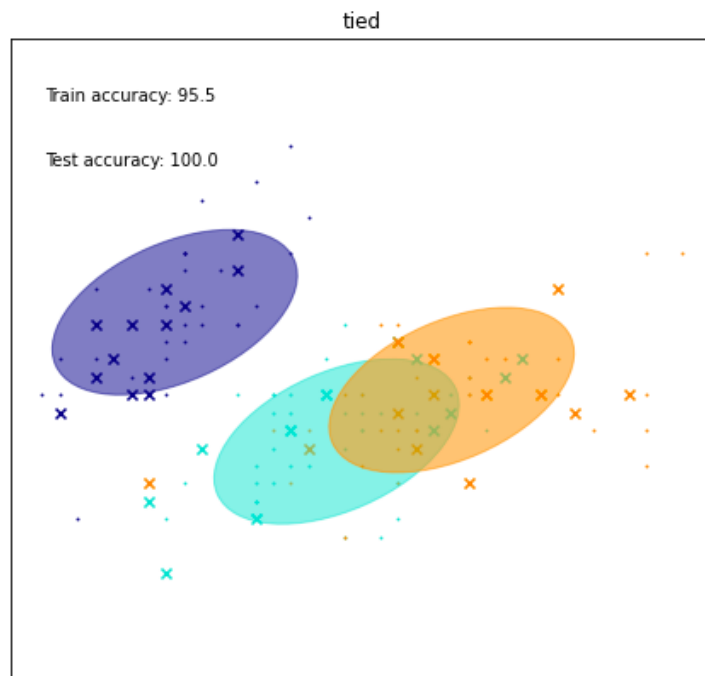
共分散行列が  
すべての対角成分  
が等しい対角行列



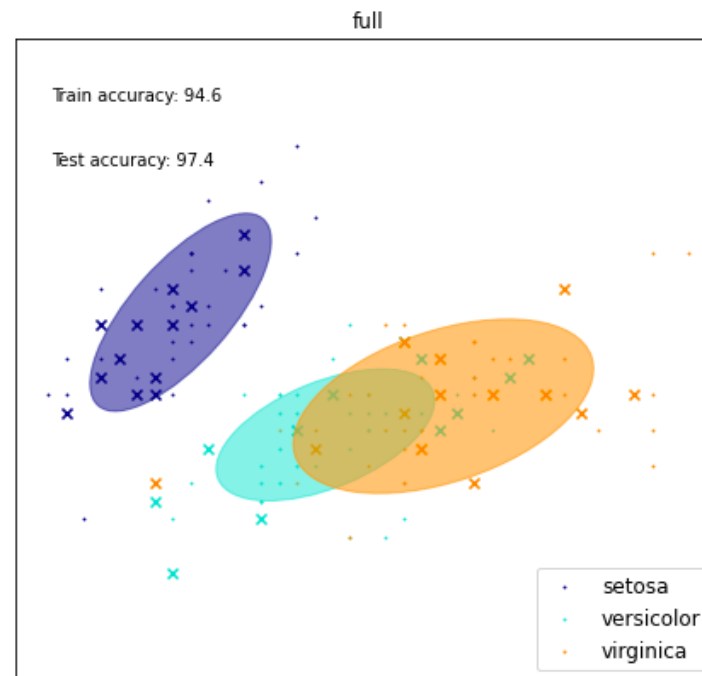
共分散行列が  
対角成分が様々な  
対角行列



共分散行列が  
対角行列でないが  
成分は共有



共分散行列が  
対角行列でない  
成分も共有しない



# 今日の課題

- word2vecという手法を使うと、単語をベクトル化できる
  - word2vecの詳細は、秋学期の「自然言語処理特論」で。
- 日本語の単語について、word2vecでベクトルを求めたデータセットがあるので、このデータセット上でクラスタリングをおこなってみる
- 各クラスタに属する単語ベクトルを、クラスタの重心に近い順に並べたとき、上位に来る単語がどんな単語かを見してみる
  - クラスタ重心に近い単語を、簡単に「重要語」と呼ぶことにする。
- クラスタリングの手法を変えたり、クラスタ数を変えたりしたとき、各クラスタの重要語がどう変化するか、調べてみる