

## 4-9 データ活用実践 (教師なし学習)

東京大学 数理・情報教育研究センター  
2020年5月11日

# 概要

- 機械学習，特に教師なし学習の種々のデータ分析手法（データのクラスタリング・確率密度推定）と，それらを用いた一連のデータ分析プロセスを実行するための知識を学びます．
- データ解析例を通じてデータの前処理から分析までどのように実行されるかを体験します．

# 本教材の目次

1. 機械学習	4
2. 教師なし学習	7
2.1. データの収集と前処理	9
3. クラスタリング	11
3.1. K-平均法	12
4. 確率密度推定	13
4.1. 混合正規分布による推定	14
4.2. 異常検知への応用	15
4.3. データ生成への応用	16
5. レコメンデーション	17
6. 過学習とモデル選択	18
7. 階層クラスタリング	20

# 機械学習

Arthur Samuel

「Field of study that gives computers the ability to learn without being explicitly programmed」 (1959)

明示的なプログラムなしにコンピュータに学習能力を与える研究分野.



画像の被写体が車である事をコンピュータに判断させるようなルールを明示的にプログラムすることは非常に困難です.

▶ 代わりに機械学習では蓄積されたデータからルールを発見する学習方法自体をプログラムします.

# 機械学習

- データに潜むルール（パターン）を自動的に見つけます.
- 人がプログラムするのは認識の仕方ではなく **学習方法** です.

画像データ（訓練データ）



学習



汎化

未知画像も正しく認識出来るようなルールの発見.  
(**汎化性**のあるルール)

機械学習でのルール発見に用いられるデータを**訓練データ**と呼びます.  
訓練データにない未知データに対しても正しく**予測**する事（**汎化**）が目的です.

機械学習には大まかに**教師あり学習**と**教師なし学習**があります

※ 今回の講義では教師なし学習を取り扱います.

# 機械学習の流れ

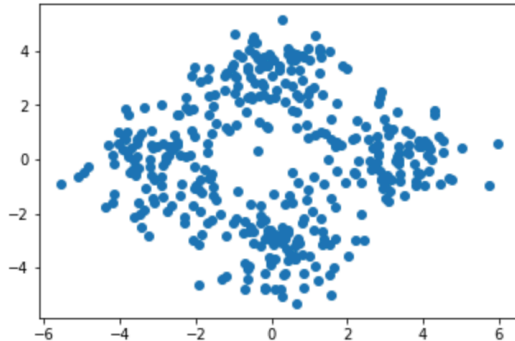
1. 行いたいタスクの特定.  
やりたい事は分類, 回帰, クラスタリング等のうちどれかを特定します.
2. 分析に必要なデータの確認, 対象となるデータの収集.  
収集すべき特徴を検討します.
3. データの分析. 機械学習による学習と評価.  
データの前処理・加工と機械学習によるデータ分析を実行します.
4. データ分析結果の共有, 課題解決に向けた提案.  
データ分析結果のレポートを作成し, それを材料に次の施策を検討します.

# 教師なし学習

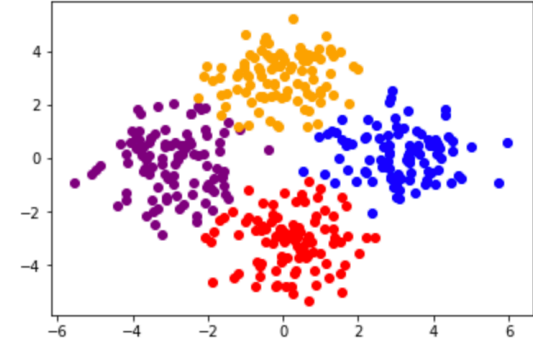
訓練データは特徴ベクトルのあつまり  $x_1, \dots, x_n$  です.

教師値が与えられない点が教師あり学習と異なります.

**教師なし学習によるグルーピング**：データ同士の類似性に基づくグルーピング（クラスタリング）は教師なし学習の代表例です.



教師なし学習による  
グルーピング



クラスタリングの例

**顧客セグメンテーション** → 顧客情報や利用状況によるクラスタリング

**店舗クラスタリング** → 店舗や周辺情報からクラスタリング

# 教師なし学習

クラスタリングの他にも様々な教師なし学習手法があります.

- 確率密度推定  
データの生成分布を推定します. 汎用的なタスクでクラスタリング・異常検知・データ生成への応用が可能な手法もあります.
- 異常検知  
センサーデータ等から異常値の検出をします.
- レコメンデーション  
ユーザーの購入履歴や嗜好情報からアイテムの推薦をします.
- データ生成  
データの生成過程を推定しデータの生成・サンプリングをします.

(補足) 教師あり学習では特徴ベクトル $x$ のラベル $y$ の予測を目的としました. 一方, 教師なし学習では $x$ の背後にある構造や生成過程の推定を目的とします.



# データの収集

クラスタリングに有効な簡単な説明変数（特徴）の作成をし、データを収集します。

## - アヤメの品種予測

特徴：sepal.length（がく片の長さ）  
sepal.width（がく片の幅）  
petal.length（花びらの長さ）  
petal.width（花びらの幅）

※ 教師値の設計は必要ありません。

特徴ベクトル $x$

	sepal.length	sepal.width	petal.length	petal.width
0	5.8	4.0	1.2	0.2
1	5.0	3.3	1.4	0.2
2	7.1	3.0	5.9	2.1
3	4.3	3.0	1.1	0.1

1 行が 1 データ

特徴ベクトルは連続値・カテゴリ値（離散値）として収集します。

特徴ベクトルの設計には対象データの専門知識を有効活用しましょう。

データの設計・収集がうまくいけば、あとは汎用の機械学習で分析可能です。

# データの加工・前処理

機械学習を精度良く正しく実行するにはデータの前処理・加工が必要です。

タイタニックの生存結果データ

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp
1	0	3	Braund, Mr. Owen Harris	male	22.0	1
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1
3	1	3	Heikkinen, Miss. Laina	female	26.0	0
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
5	0	3	Allen, Mr. William Henry	male	35.0	0
6	0	3	Moran, Mr. James	male	NaN	0

カテゴリ変数 欠損値

## データクレンジング

- 外れ値検出・除去
- 欠損値除去・補完

## カテゴリ値の数値への変換

- ダミー変数の追加

例：Male, Female列を追加し0-1値に変換。

## 有効な前処理

- 標準化：列を平均0, 分散1に変換。
- 正規化：列を $[0,1]$ ,  $[-1,1]$ に収める。列or行の平均を0にノルムを1に変換。
- 変数変換
- サンプリング：訓練データの一部を無作為に選び新たに訓練データとします。  
訓練データサイズの調整やアンサンブル学習の際に用います。

# クラスタリング

データには潜在的なカテゴリがあり、それを推定します。

クラスタリングの例

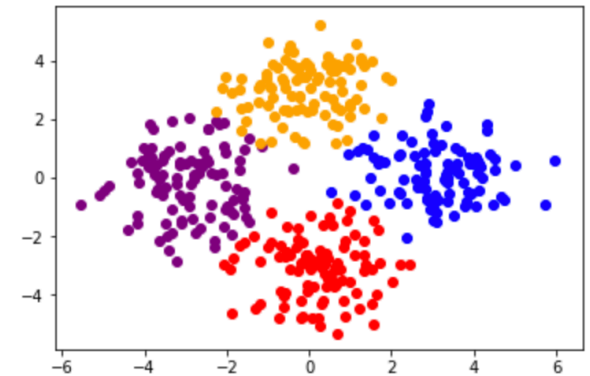
- テキストデータ  
テキストのジャンル（ミステリー・SF・アクション・純文学）
- ECサイトの購入履歴データ  
趣味嗜好やプロフィールに依存した購入パターン

- **データの類似度に基づく方法**

似たデータは同じクラスター，似ていないデータは異なるクラスターに属するという志向のもとでクラスタリングを実行します。  
代表例：K-平均法（K-Means法）

- **確率密度推定に基づく方法**

多峰性を持つ確率密度関数による確率密度推定を行い，各データが属する峰を特定しクラスタリングします。  
代表例：混合正規分布によるクラスタリング



K-平均法によるクラスタリング

# K-平均法 (K-Means法)

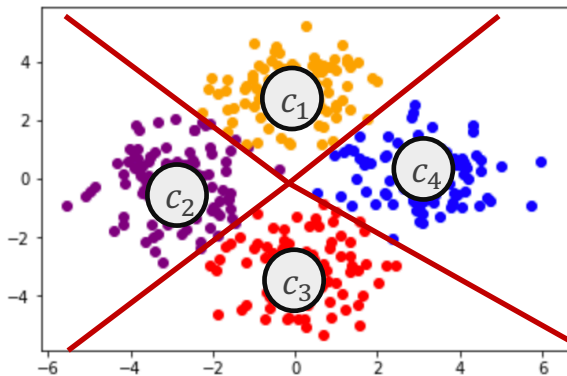
データを以下の方針で指定の $K$ 個にクラスタリングします.

- 各クラスタに $t = 1, \dots, K$ に対応する中心点  $c_t$  を求めます.
- 各データは最も近い  $c_t$  が代表するクラスタに割り当てます.

K-平均法の実行にはデータの類似度を示す距離関数を指定する必要があります.  
距離関数には様々なものがありますが, 代表例はユークリッド距離になります.

(参考) ユークリッド距離は2点  $x_1 = (x_1^{(1)}, \dots, x_1^{(p)})$ ,  $x_2 = (x_2^{(1)}, \dots, x_2^{(p)})$  に対し次で定義される距離関数です.

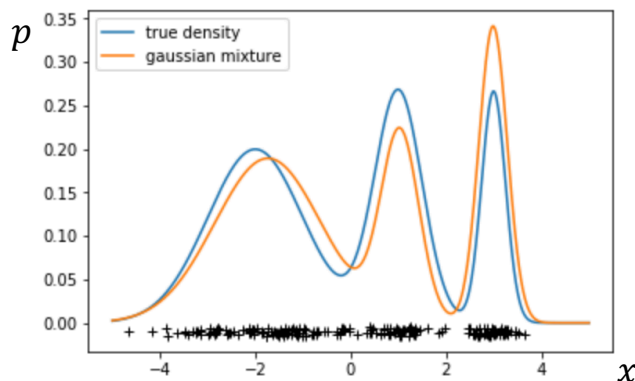
$$d(x_1, x_2) = \left(x_1^{(1)} - x_2^{(1)}\right)^2 + \dots + \left(x_1^{(p)} - x_2^{(p)}\right)^2.$$



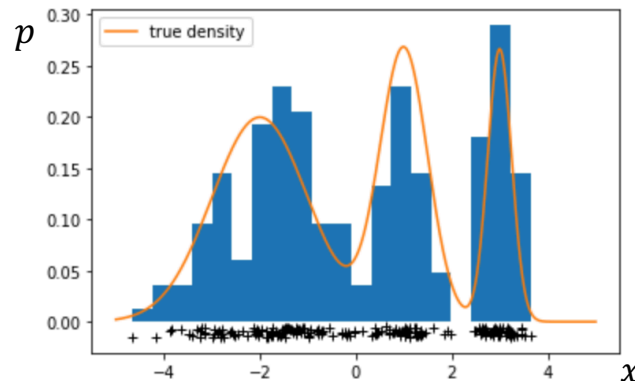
K-平均法によるクラスタリングの様子.  
クラスタの中心点とクラスタリング結果が得られます.

# 確率密度推定

訓練データ  $x_1, \dots, x_n$  からデータを生成する確率密度関数  $p(x|w)$  を推定します。確率密度推定法として最尤推定法やヒストグラムによる推定があります。



混合正規分布による最尤推定



ヒストグラムによる確率密度推定

横軸はデータ  $x$  の座標，縦軸は確率密度  $p(x|w)$  です。橙のグラフは正解の確率密度関数，左図の青のグラフと右図の青のヒストグラムはそれぞれの推定結果です。

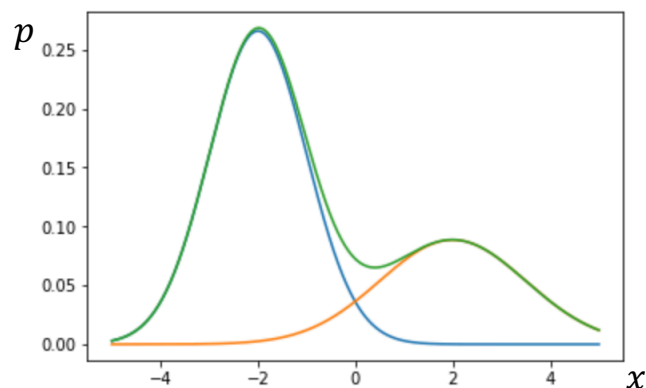
(参考) 確率モデルの最尤推定とは？

確率モデル  $p(x|w)$  はパラメータ  $w$  を持つデータ  $x$  についての確率密度関数で， $w$  に依存して確率密度は変化します。最尤推定は訓練データ  $x_1, \dots, x_n$  を最も高い確率で発生させるようなパラメータ  $w$  を求める手法です。最尤推定によりデータの背後に潜む確率密度関数が推定されることになります。

# 混合正規分布による確率密度推定

混合正規分布とは複数の峰を持つ確率分布（確率密度関数）で、パラメータにより峰の位置と裾の広さが調整されます。

データが複数のまとまったクラスタを持つ場合の確率密度推定に適した確率分布といえます。



緑：2つの峰を持つ混合正規分布  
橙，青：各々の峰を描いたグラフ。

横軸はデータ $x$ の座標，縦軸は確率密度を示しています。

確率密度推定を行う際に混合数はユーザーが指定する必要があります。混合正規分布による確率密度推定ではデータが属す峰も推定する事ができ，結果としてクラスタリングも実行されます。

（参考）峰が一つの混合正規分布は単に正規分布と呼びます。

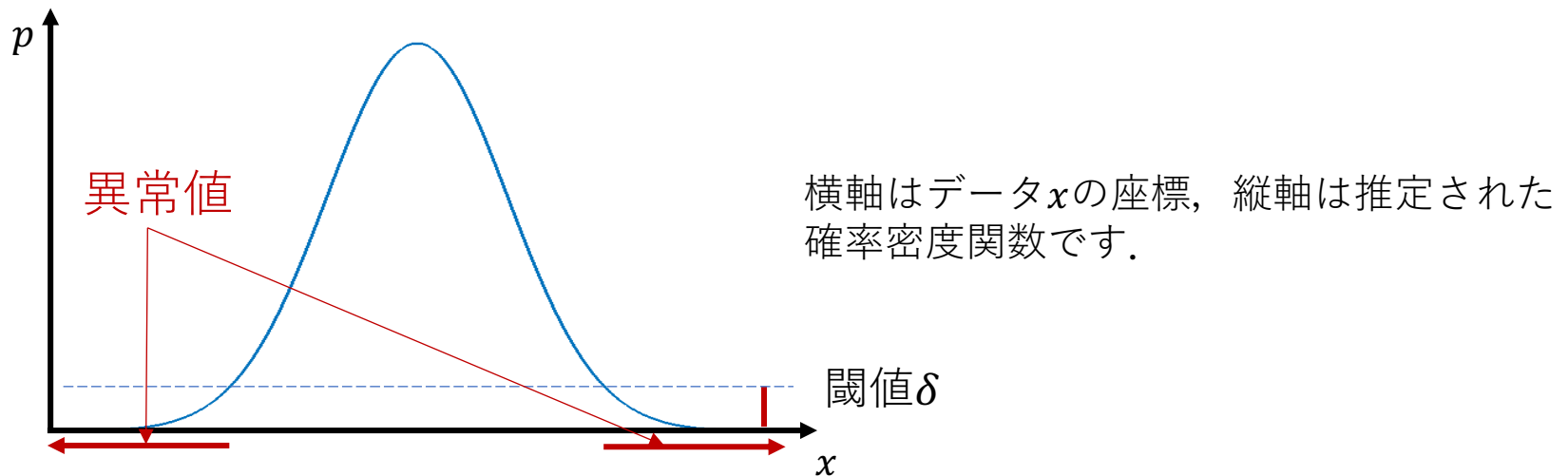
# 確率密度推定の異常検知への応用

異常検知は非常に多くの応用例があります：

工場のセンサーデータ，カードの使用履歴，監視カメラ，為替取引

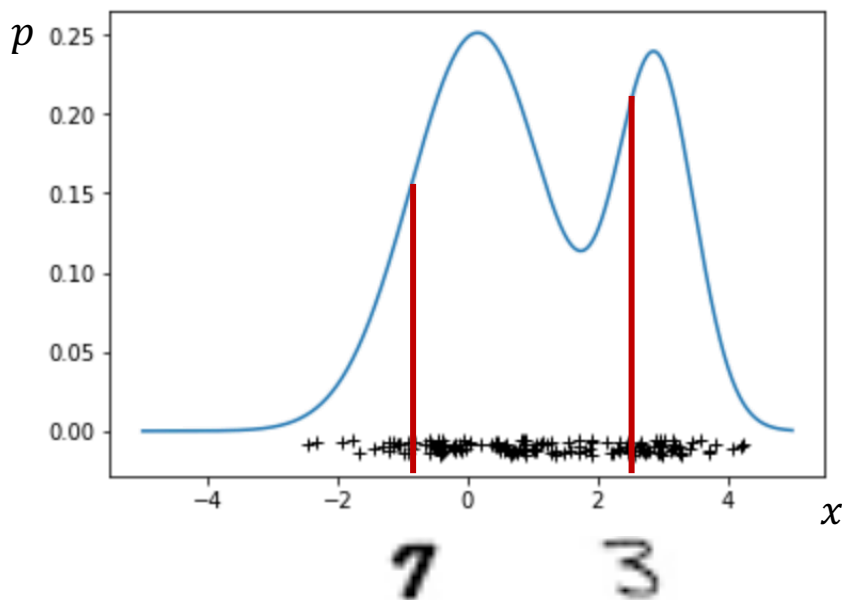
正常データ・異常データが共に十分あれば教師あり学習が適用可能ですが一般に異常データは非常に少ないため教師なし学習を適用します．

**典型的なアプローチ**：確率密度推定の結果  $p(x|\hat{w})$  と閾値  $\delta$  を設定し  $p(x|\hat{w}) \leq \delta$  を満たすデータ  $x$  を異常とみなします．



# 確率密度推定のデータ生成への応用

確率密度推定によりデータの生起確率が分かるため，推定された確率密度関数からサンプリングする事で尤もらしいデータの生成も可能です．



MNIST: 手書き文字データセット

MNISTで学習した確率モデルから確率の高いデータをサンプリングするとそれらしい手書き文字が得られます．



# レコメンデーション

ユーザーによるアイテムの評価値情報が部分的にある時、欠損値を推定することで評価値を推定しアイテムの推薦をします。

	アイテムA	アイテムB	アイテムC	...	アイテムZ
ユーザー1	4	8	*	...	2
ユーザー2	2	*	2	...	*
ユーザー3	4	7	*	...	*
⋮	⋮	⋮	⋮	...	⋮

嗜好の近いユーザー

嗜好の近いユーザー同士は欠損値でも近い評価をするだろうというモデリングに従い欠損補完が可能です。

(参考) この様な欠損値補完手法として行列分解等があります。

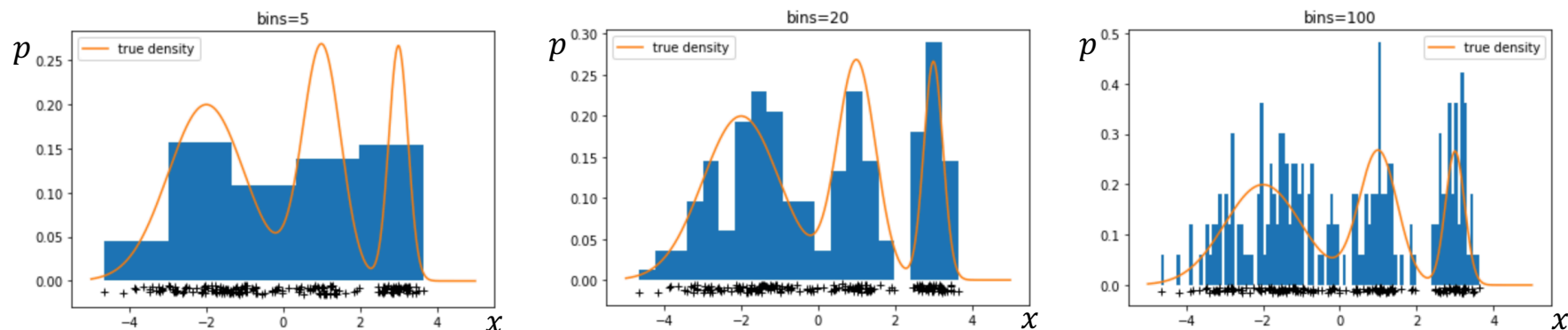
応用例：商品販売サイトでの商品推薦，動画サイトでの動画推薦。

# 過学習とモデル選択

教師なし学習でも有限個の訓練データにモデルを適合するため、訓練データに対しモデルが複雑過ぎる場合に過学習が生じます。過学習を防ぐためにモデルの表現力をコントロールする必要があります。

モデルの表現力をコントロールするハイパーパラメータの例

- K-平均法を中心点の数
- 混合正規分布の混合数
- ヒストグラムのビン幅（ヒストグラムの各帯をビン、その幅をビン幅と呼びます）



ヒストグラムのビン幅（左から5, 20, 100）で未学習・過学習が生じる様子。横軸はデータの座標、縦軸は確率密度、橙のグラフは正解の確率密度関数です。

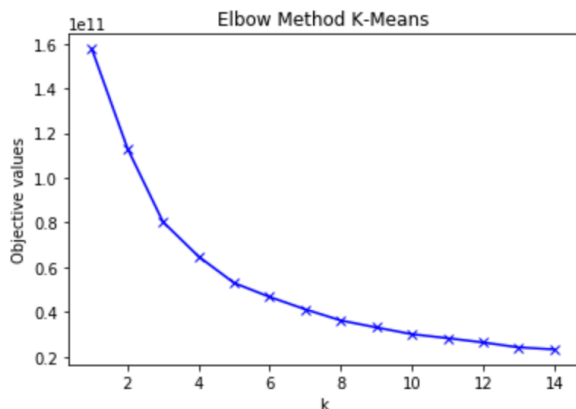
# モデル選択の方法

教師なし学習でも適当な評価指標を用いた交差検証で適用すべきモデルの選択やハイパーパラメータの決定は可能です。

しかし教師あり学習におけるラベルの推定精度のようなモデル非依存の評価指標がない問題（クラスタリングや確率密度推定）では、手法毎にモデル選択法が提案されています。

例：K-平均法のエルボー法

Kの値毎にK-平均法を実行し目的関数値（データへの適合度）をプロットします。基本的にKについて単調に減少しますが減少幅が十分に小さくなってきたKを適切なものとし選択します。



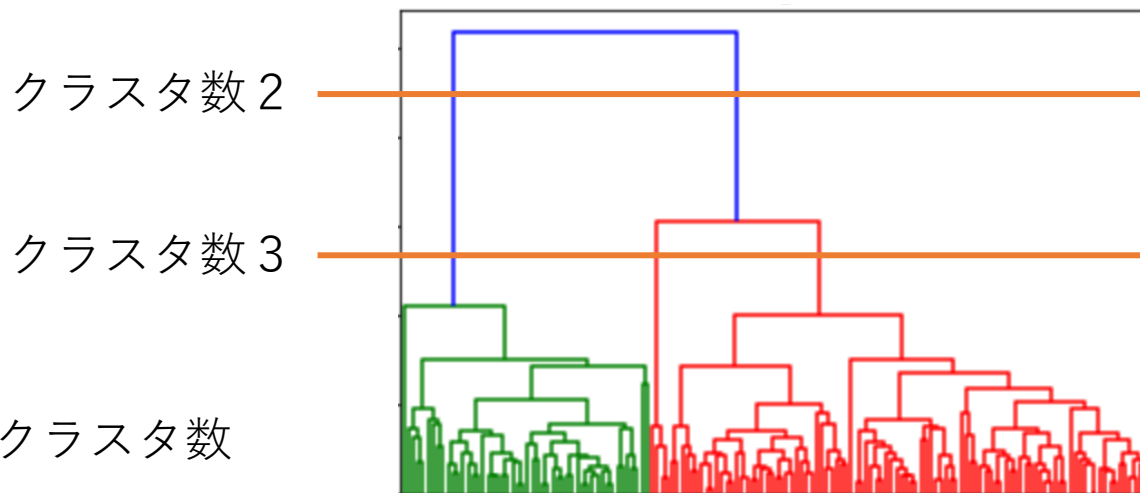
エルボー法の実行例。  
縦軸はK-平均法の目的関数値です。  
K=5あたりで減少が落ち着いてきます。

# 階層クラスタリング

これまで紹介したクラスタリング手法は**非階層クラスタリング**です。  
一方、階層的にクラスタを形成する**階層クラスタリング**という手法もあり、  
データ間の関係性に階層構造がある場合のクラスタリングに有用です。  
例：生物や遺伝子配列の分類。

階層クラスタリング手法では各データが一つのクラスタを形成している状態から近いクラスタ同士を順次併合していき、一つのクラスタになるまで繰り返します。

結果、右図のような**系統樹**が得られます。



階層クラスタリングの利点：  
系統樹の高さを指定する事でクラスタ数  
を学習後に調整する事が出来ます。