

カーネル多変量解析 第一章(2)

発表者

斉藤 章久

概要

- ・カーネル法を使うとどういう利点があるか？
- ・また、それを生かした応用分野にはどのようなものがあるか？

カーネル多変量解析の特徴(1)

(1) サンプルが増えればどんどん複雑にできる

- ・カーネル関数を使ったモデル化では、パラメータ α の次元がサンプル数と同じになっており、それだけ自由度の高いモデルとなっている。

(2) 線形性と非線形性の二つの側面

- ・関数を使ったモデル化では、決めるべきパラメータについては線形性が保たれているが、入力データ x については非線形な関数を実現している。
- ・パラメータは線形性を生かした高速計算によって求めることによって、全体として非線形をもつ複雑なモデルにあてはまることができる。

カーネル多変量解析の特徴(2)

(3)高次元・非数値データをくるむ殻

- ・カーネル関数では1次元に限らず、高次元のデータでも良い。
- ・実数ベクトルでなくても良いので、文字列やグラフ構造などの複雑なデータも扱える。

(4)カーネル計算のモジュール

- ・カーネル関数がどのような関数になっているかはその後の処理には影響しないことが多い。
- ・手持ちのデータが複雑だろうが、カーネル関数を計算すれば簡単に処理ができる。

カーネル法の応用分野

- 手軽に分類や識別、予測などのデータ解析を扱うことができる



【検索エンジンや商品推薦システムなど】

- DNAやアミノ酸の配列から情報を読み取る



【有用な遺伝子の発見。医療や創薬を目的とする】

カーネル法の種類

・多変量解析の手法は

- ①問題設定
- ②計算法

によって分類できる。

・カーネル法では計算アルゴリズムの側面を重視しており、計算法でまとめたほうが共通点が多い

問題設定による分類(1)

(1)関数近似

入力 x から出力 y への関数を推定する。

関数近似には実数値の出力値を学習するものと、離散地を出力するものとある。

- 実数値の出力を学習 → 回帰
- 離散値の出力を学習 → クラス識別

問題設定による分類(2)

(2)情報圧縮

- ・情報圧縮とは、入力から必要な情報だけを抽出する処理

《目的》

情報を図に表現する可視化のため

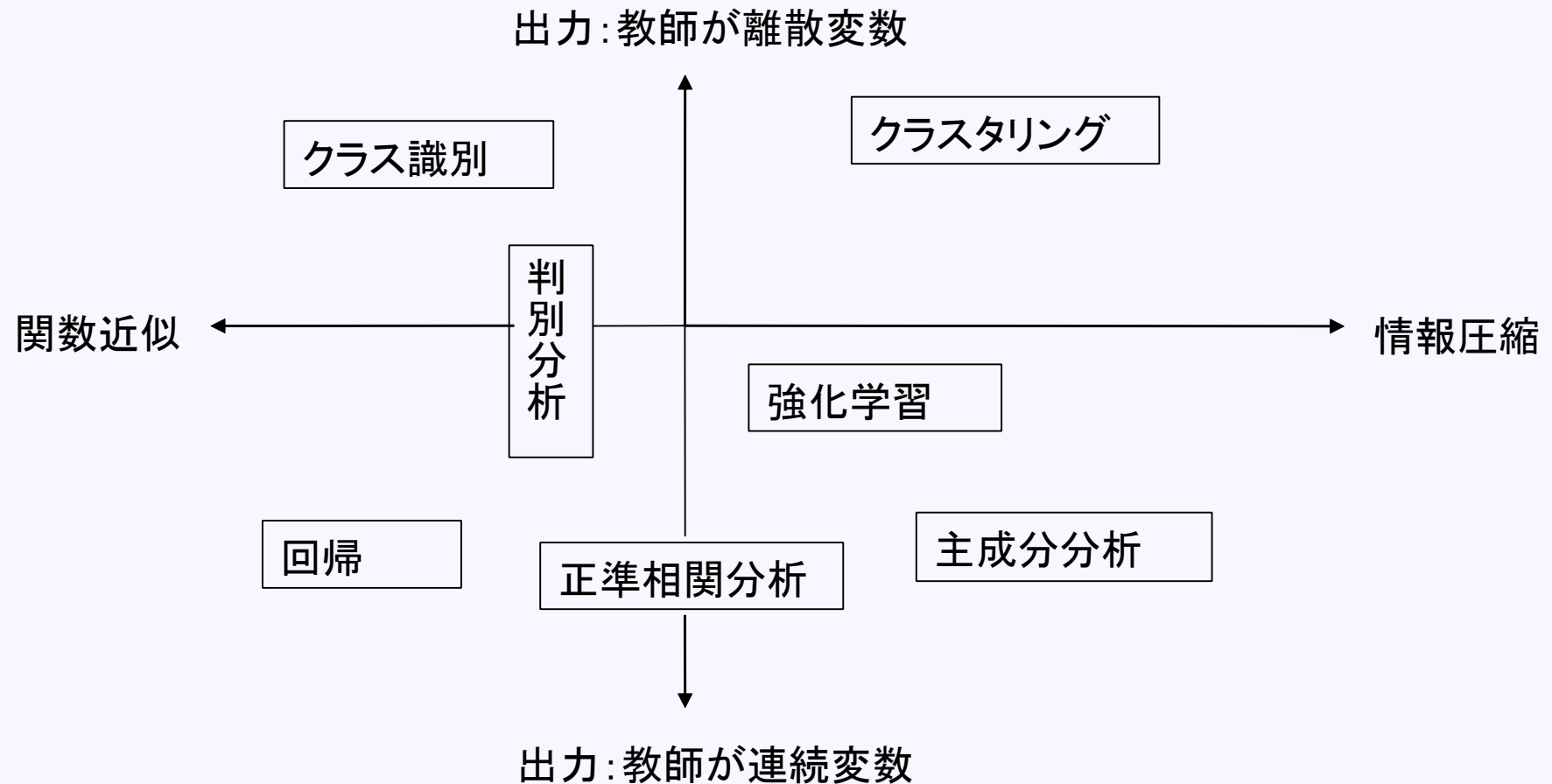
不要な情報を取り除き、その後の処理を単純化する

問題設定による分類(3)

その他の分析法

- クラスタリング
- 主成分分析
- 判別分析
- 正準相関分析

問題設定による分類(4)



計算法による分類(1)

- ・さまざまな問題を解く際に共通していること



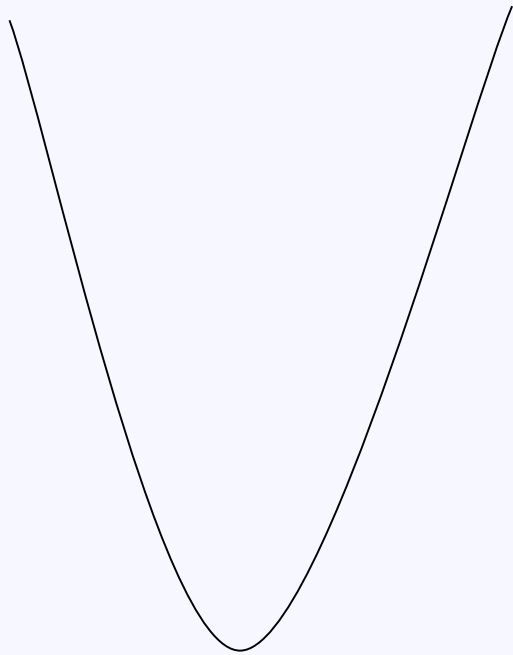
〈なんらかの損失関数を定義し、それをパラメータについて最小にするという最適化問題を解くことに帰着できる〉

- ・陽に解が求まらない場合



〈容易に解を求めるために、損失関数を鉢状の凸関数を利用する〉

計算による分類(2)



凸な損失関数



凸でない損失関数

計算量による分類(2)

- 計算量という観点での難易度

(やさしい)関数近似

(難しい)次元圧縮になったり、離散変数ができてたりする問題



そういう問題は固有値問題として定式化される

(ここまできるといろいろな問題を解くことができる)

また、サポートベクトルマシンなどは、固有値問題よりも一段階難しい問題に帰着される