カーネル法入門
-正定値カーネルによるデータ解析-

1.カーネル法への招待
1.1 カーネル法の基本的アイデア

08T4082A　西野太樹

古典的なデータ解析

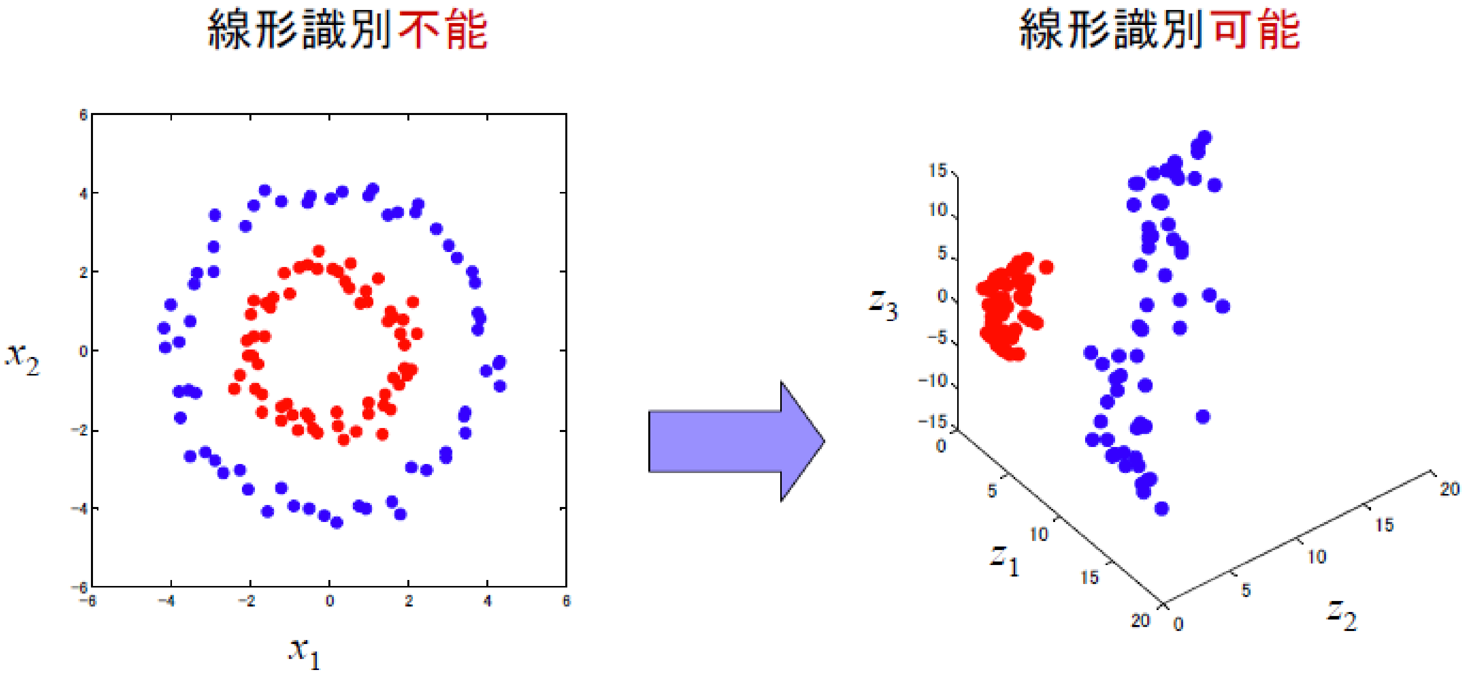
- m次元、N個からなるデータの行列表現

$$X = \begin{pmatrix} X_1^1 & X_1^2 & \cdots & X_1^m \\ \vdots & \vdots & \ddots & \vdots \\ X_N^1 & X_N^2 & \cdots & X_N^m \end{pmatrix}$$

- 多くのデータ解析手法は行列の線形処理を行う
- 現実問題では不十分なことも多い

線形解析の問題例

- 2クラス識別問題



  - 3次元に変換することで線形識別が可能

データの非線形変換

- データに非線形変換を施すと、データの性質がよりよく捉えられる場合がある

- 以下の問題を解決する必要もある
  - どのような変換が適切か
  - 高次元データの扱い

- カーネル法は高次モーメントを有効に抽出し、必要な計算を効率的に実行できる非線形変換法

高次元への非線形写像

$$\Phi : \Omega \to H$$

- 変換Φによって、元データの空間Ωを高次元の実ベクトル空間Hに写像する
  - Φ: 特徴写像
  - H : 特徴空間

- 特徴空間H上のデータ解析ではHが内積を持つことが重要

カーネルトリック

- 高次元なベクトル空間ほど計算コストが莫大に
- カーネルトリック
  - 正定値カーネルによって特徴写像Φが定義（→第2章）
  - 2つのデータ点Φ（X_i）, Φ（X_j）の内積はカーネル関数kの評価によって計算される

$$\left\langle \Phi\left(X_i\right), \Phi\left(X_j\right) \right\rangle = k\left(X_i, X_j\right)$$

  - 計算コストはカーネル値k(x,y)にのみ依存
  - さまざまな線形データ解析手法を変換データに適用可能