3.7 さまざまなカーネル法
3.8 カーネル法の性質
3.9 カーネル法の選択

08T4083T 真下飛瑠

3.7 さまざまなカーネル法

- カーネル **K-means** クラスタリング
  - K-meansクラスタリングとは、N個のデータをK個のクラスタ$C_1,\ldots,C_K$に分割する手法

    $\mathbf{R}^m$のデータ$\{X_i\}_{i=1}^N$に対して、シード点$\{z_a\}_{a=1}^N$を用意し、

    データX_iに最も近いシード点がZ_aであるとき$X_i \in C_a$とする。

    シード点は

    $$\sum_{a=1}^{K} \sum_{X_i \in C_a} \left\| X_i - z_a \right\|^2$$

    が最小になるように、繰り返し法によって求める。

  - これを再生核ヒルベルト空間に変換したデータに拡張したもの

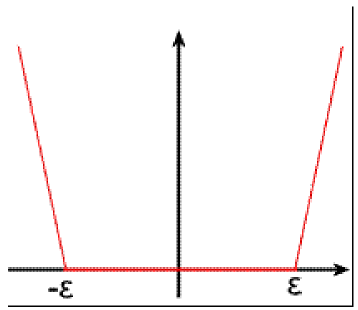
- カーネル**PLS**
  - 回帰分析の1つ、PLS回帰をカーネル化したもの
  - PLS回帰とは、m次元説明変数XからYへの線形回帰を行うさい、Yとの相関が高くなるようなXの部分空間へデータを射影し、射影されたデータによってYへの線形回帰を行う方法
  - 説明変数Xが共線性を持っている場合に有効
  - 計量化学分野でよく用いられる

- サポートベクター回帰
  - SVMを、連続値をとる応答変数の場合に拡張したもの

線形関数 $f(x; w, b) = w^T x + b$ をモデルとし、損失関数として

$\varepsilon -$不反応損失$(\varepsilon - insensitive\ loss)$

$$l_\varepsilon(y, f) = \max\{|y - f| - \varepsilon, 0\}$$



を用いて

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} l_\varepsilon(Y_i, f(X_i; w, b))$$

を最小化する$f(x; w, b)$を解として用いる

- カーネルロジスティック回帰
  - 線形判別の手法の1つ、ロジスティック回帰をカーネル化したもの
  - ロジスティック回帰は2値判別の場合、損失関数として

$$l(y, f) = \log(1 + \exp(-yf))$$

    を用いる
  - 確率を直接扱うことができ、また多クラスへの拡張も容易である
  - データ数が多いと最適化の計算量が増大する

3.8 カーネル法の性質

- カーネル法はデータの非線形性や高次モーメントを扱うためのデータ変換の方法論
- 計算効率を重視した変換
- 正定値カーネルの定める特徴写像によって変換されたデータの内積は、正定値カーネルの値の評価によって計算可能（カーネルトリック）
- グラム行列を用いた計算によって、多くの線形手法が再生核ヒルベルト空間へ拡張可能

- 多くの場合、再生核ヒルベルト空間はデータの張る部分空間の元で解が与えられる（**Representer**定理）

- 計算量は基本的にデータ数に依存。データ数が非常に大きい場合は低ランク近似などの工夫が必要

- 正定値カーネルは任意の集合上に定義可能。いったん正定値カーネルが与えられれば、どのようなタイプのデータに対しても同じ方法論が適用可能。

3.9 カーネルの選択

カーネル法の性能は、選択した正定値カーネルに大きく依存する。そのため正定値カーネルの選択は重要なモデル選択の問題である

- **3.9.1** 構造化データに対するカーネル
  - 正定値カーネルは任意の集合上に定義できるため、さまざまな集合上のデータに応用可能
  - 特にシンボル列や代数構造などの、特定の構造をもったデータ（構造化データ）に対しては、その構造を反映した正定値カーネルを用いたほうが、より特徴を捉えた解析が期待できる

- **3.9.2** 教師あり学習に対するクロスバリデーション
  - SVMなどの識別や回帰の問題は、応答変数 Yi を Xi に対して正解を与える教師とみなし、教師あり学習と呼ぶことがある
  - これらの問題におけるカーネルの選択には、クロスバリデーションを用いることが多い
  - クロスバリデーションは、教師あり学習アルゴリズムが持つハイパーパラメータ（アルゴリズムの外的変数）を適切に決めるための方法
  - SVMでは、制約の破れに対する係数Cやカーネルに含まれるパラメータ（ガウスRBFカーネルのσなど）がハイパーパラメータにあたる

- **Leave-one-out** クロスバリデーション（LOOCV）とは、ハイパーパラメータ λ のもとでの期待損失を

$$\frac{1}{N}\sum_{k=1}^{N} l(Y_k, \hat{f}_\lambda^{[-k]}(X_k))$$

によって推定する方法。すべての λ に対して上式を計算し、最小値となるハイパーパラメータを求める

- データD ＝ {(X1,Y1),…,(Xn,Yn)}
- $\hat{f}_\lambda$:Dに対して得られる関数
- $D^{[-k]}$:Dからk番目のデータを除いたデータ
- $\hat{f}_\lambda^{[-k]}$:D[-k]によってアルゴリズムが与える結果

- LOOCVでは各 λ に対してデータ数 N だけ $\hat{f}_{\lambda}^{[-k]}$ を求める必要があり、計算上の困難が生じることがある
- そのため **K-fold** クロスバリデーションという手法もよく行われる

- これはデータをK分割し、

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\left|D_k\right|}\sum_{i\in D_k} l(Y_i, f_{\lambda}^{(-k)}(X_i))$$

によって期待損失を推定する

- $f_{\lambda}^{(-k)}$:k 番目のグループ D_k 以外のデータを用いた結果

- **3.9.3** 教師なし学習に対するカーネル選択
  - 主成分分析や正準相関分析などがこれ
  - カーネル選択に関する汎用的な方法で、理論的に性能が保証されたものはない
  - カーネル法の応用に依存して、カーネルを適切に選択することは可能
- **3.9.4** その他の方法
  - ハイパーカーネル：正定値カーネルを最適化する方法。正定値カーネルの族の上にさらに正定値カーネルを定義
  - Multiple Kernel Learning（MKL）：いくつかの異なる正定値カーネルに対し線形結合や凸結合を行い、新しい正定値カーネルを定める。そしてその結合係数を最適化