カーネル法の基礎 4.4節

参考文献

- 金森, 統計的学習理論, 講談社, 2015.

Table of contents

- 表現定理

線形モデル$\mathcal{M} = \{ f(x) = \beta^{\mathrm{T}} \phi(x) \mid \beta \in \mathbb{R}^D \}$における推定量$\hat{f}(x)$

$$\hat{f}(x) = \sum_{i=1}^{n} \hat{\alpha}_i k(x_i, x)$$

- $\hat{\alpha} = (K + \lambda I_n)^{-1} Y$
- $k(x_i, \cdot),\ i = 1, \ldots, n$: データ点$x_i$に対応するカーネル関数

$\rightsquigarrow$ カーネル関数の線形結合

$\rightsquigarrow$ この性質は，一般のRKHSにおいて表現定理としてまとめられる．

- $\mathcal{X}$: 入力空間
- $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$: $\mathcal{X}$上のRKHS
  - k: 対応する再生核
- $\mathcal{H}+\mathbb{R}=\{f+b\,f\in\mathcal{H}, b\in\mathbb{R}\}$: $\mathcal{H}$から構成される統計モデル

$\rightsquigarrow$ データ $D=\{(x_1,y_1),\dots,(x_n,y_n)\}$が与えられた下で以下の関数を最小化.

$$\min_{f,b} L(f(x_1)+b,\dots,f(x_n)+b;D)+\lambda\|f\|_{\mathcal{H}}^2 \quad (\lambda\geq 0) \quad (4.7)$$

$\mathcal{H}$の部分空間で,

- $S=\mathrm{Span}\{k(x_1,\cdot),\dots,k(x_n,\cdot)\}\subset\mathcal{H}$
- $S^\perp=\{v\in\mathcal{H}\mid\langle v,w\rangle_{\mathcal{H}}=0\ \forall w\in S\}$: Sの直交補空間

を考える.

$\rightsquigarrow f \in \mathcal{H}$は，

$$f = f_S + f_{S^\perp} \ (f_S \in S,\ f_{S^\perp} \in S^\perp)$$

と一意に分解できる (∵ 射影定理).

部分空間Sの定義から，

$$\langle f, k(x_i, \cdot)\rangle = \langle f_S, k(x_i, \cdot)\rangle, \quad \|f\|_{\mathcal{H}}^2 = \|f_S\|_{\mathcal{H}}^2 + \|f_{S^\perp}\|_{\mathcal{H}}^2$$

が成り立つ.
$\rightsquigarrow f$をf_Sに変えると，関数値は$\lambda\|f_{S^\perp}\|^2$ だけ減少する
$\rightsquigarrow f$の最適解が存在する範囲として，部分空間Sを考えれば十分.

以上の結果は，表現定理としてまとめられる.

Th 4.10 (表現定理)

学習データを

$$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

として，関数

$$L(f(x_1) + b, \ldots, f(x_n) + b; D) + \Psi(\|f\|_{\mathcal{H}}^2) \quad (4.8)$$

を$f \in \mathcal{H}$と$b \in \mathbb{R}$に関して最小化することを考える．

- L: 任意の関数, Ψ: 単調非減少関数

$\rightsquigarrow f \in \mathcal{H}$について，

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) \quad (4.9)$$

と表せる最適解が存在する．

(4.8)による定式化

- 最適化すべきパラメータfの次元が無限次元になることもありえる

表現定理

- $n+1$次元パラメータ $(\alpha_1, \ldots, \alpha_n, b)$ の最適化問題として定式化
- カーネル関数$k(x, x')$の値が簡単に計算できるなら，最適化の計算コストはデータ数nによって決まる．

(4.8)の最適化を有限次元の問題として表す．

- $k_i := (k(x_i, x_1), \ldots, k(x_i, x_n))^{\mathrm{T}} \in \mathbb{R}^n$
- $K = (k_1, \ldots, k_n)$: グラム行列
- $\alpha := (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}} \in \mathbb{R}^n$

$\rightsquigarrow$ 関数(4.8)は

$$L(\alpha^{\mathrm{T}} k_1 + b, \ldots, \alpha^{\mathrm{T}} k_n + b; D) + \Psi(\alpha^{\mathrm{T}} K \alpha)$$

となる．

パラメータα, bについて最適化し，最適解$\hat{\alpha}, \hat{\beta}$が得られたとき，学習された関数は，

$$f(x) = \sum_{i=1}^{n} \hat{\alpha}_i k(x_i, x) + \hat{b}$$

と表される．