カーネル法の基礎 4.5 - 4.6節

参考文献

- 金森, 統計的学習理論, 講談社, 2015.

Table of contents

- RKHSのラデマッハ複雑度
- 普遍カーネル

RKHSのラデマッハ複雑度

Th 4.11

$\mathcal{X}$上の再生核Hilbert空間を$(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$とし，対応する再生核を$k : \mathcal{X}^2 \to \mathbb{R}$とする．
関数集合$\mathcal{G} \subset \mathcal{H}$は，

$$\mathcal{G} \subset \{f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \leq a\} \ (a > 0)$$

を満たすとする．このとき，入力点の集合$S = \{x_1, \ldots, x_m\} \subset \mathcal{X}$に対して経験ラデマッハ複雑度は，

$$\hat{\mathfrak{R}}_S(\mathcal{G}) \leq \frac{a}{m}\left(\sum_{i=1}^{m} k(x_i, x_i)\right)^{1/2}$$

となる．

(証明略)

カーネル関数が有界ならば，さらに次の不等式が得られる．

Corollary 4.12

Th 4.11の条件に加えて，カーネル関数が$\sup_{x\in\mathcal{X}} k(x,x) \leq \Lambda^2$を満たすとき，

$$\hat{\mathfrak{R}}_S(\mathcal{G}) \leq \frac{a\Lambda}{\sqrt{m}}$$

が成り立つ．

同様にして，入力の分布に関して期待値をとったラデマッハ複雑度$\mathfrak{R}_m(\mathcal{G}) = \mathbb{E}_S[\hat{\mathfrak{R}}_S(\mathcal{G})]$に対して，

$$\mathfrak{R}_m(\mathcal{G}) \leq \frac{a\sqrt{\mathbb{E}[k(x,x)]}}{\sqrt{m}} \leq \frac{a\Lambda}{\sqrt{m}}$$

が成り立つ．

普遍カーネル

Def 4.13 普遍カーネル

- $\mathcal{X}$: コンパクト距離空間，$C(\mathcal{X})$: $\mathcal{X}$上の連続関数の集合
- k: $\mathcal{X}$上の連続なカーネル関数，$\mathcal{H}$: kに対応するRKHS
  任意の$g \in C(\mathcal{X})$と$\varepsilon > 0$に対して，$f \in \mathcal{H}$が存在して

$$\|f - g\|_\infty := \sup_{x \in \mathcal{X}} |f(x) - g(x)| \leq \varepsilon$$

となるとき，kを普遍カーネルという．

Remark

- 普遍カーネルはコンパクト集合上で連続 $\rightsquigarrow$ 有界
  - 対応する再生核Hilbert空間の元は連続関数からなる (from Th 4.7)

$\rightsquigarrow \mathcal{H} \subset C(\mathcal{X})$

普遍カーネルの例

$\mathbb{R}^d$のコンパクト集合$\mathcal{X}$上の普遍カーネルの代表例

- ガウシアンカーネル
- 指数カーネル

$$k(x, x') = e^{x^{\mathrm{T}} x'}, \quad x, x' \in \mathcal{X}$$

- 2項カーネル

$$k(x, x') = (1 - x^{\mathrm{T}} x')^{-\alpha}, \quad x, x' \in \mathcal{X} \subset \{x \in \mathbb{R}^d \mid \|x\| < 1\}, \ \alpha > 0$$

Remark

- 多項式カーネルに対応するRKHSは有限次元なので，$\mathcal{X} = \infty$のときは普遍カーネルではない．

普遍カーネルから生成されるRKHSを用いれば，連続関数をよく近似できる (一般の可測関数の近似は保証されていない).
一方で，ベイズ誤差

$$R^*_{\mathrm{err}} = \inf_{f:\text{可測}} \mathbb{E}[\mathbf{1}[\mathrm{sign}(f(X)) \neq Y]]$$

は可測関数の集合上の下限で定義される.
$\rightsquigarrow$ RKHS $\mathcal{H}$上での予測判別誤差の下限がR^*_{err}に一致するなら，$\mathcal{H}$の要素で十分精度が高い判別が可能.

一般の予測損失について，可測関数集合上の下限とRKHS上の下限の関係について考察する．

$f_1, \dots, f_L$: $\mathcal{X}$上の可測関数の組 が与えられたとき，データ$(x, y) \in \mathcal{X} \times \mathcal{Y}$に対する損失を考える．

判別問題を想定し，$\mathcal{Y}$を有限集合とする．

L: 一般の自然数(2値判別では，$L = 1$，多値判別では$L = |\mathcal{Y}|$)．

Th 4.14

- $\mathcal{H}$: $\mathcal{X}$上の普遍カーネルから定義されるRKHS
- $\mathcal{Y}$: 有限集合

非負値をとる関数$\ell : \mathbb{R}^L \times \mathcal{Y} \to \mathbb{R}_+$に対して次の条件を仮定．

1. 任意の$y \in \mathcal{Y}$に対して$t \mapsto \ell(t, y)$は$\mathbb{R}^L$上で連続．
2. 単調非減少関数$h : \mathbb{R}_+ \to \mathbb{R}_+$が存在し，任意の$(t, y) \in \mathbb{R}^L \times \mathcal{Y}$に対して $\ell(t, y) \leq h(\|t\|_1)$が成立．($\|t\|_1$は$t \in \mathbb{R}^L$の1-ノルム)

このとき，$\mathcal{X} \times \mathcal{Y}$上の任意の確率分布に対して

$$
\begin{aligned}
&\inf_{f_1,\ldots,f_L:\text{可測}} \mathbb{E}[\ell(f_1(X), \ldots, f_L(X), Y)] \\
= &\inf_{f_1,\ldots,f_L \in \mathcal{H}} \mathbb{E}[\ell(f_1(X), \ldots, f_L(X), Y)]
\end{aligned}
$$

が成り立つ．