仮説集合の複雑度 2.2節, 2.3節

参考文献

- 金森, 統計的学習理論, 講談社, 2015.

Table of contents

ラデマッハ複雑度

実数値関数の集合に対して自然に定義される

$\mathcal{G} \subset \{f : \mathcal{X} \to \mathbb{R}\}$: 入力空間$\mathcal{X}$上の実数値関数からなる集合

Def 2.4 (経験ラデマッハ複雑度)

$S = \{x_1, \ldots, x_n\} \subset \mathcal{X}$: 入力点の集合

$\sigma_1, \ldots, \sigma_n$: $+1$と-1を等確率でとる独立な確率変数

このとき，$\mathcal{G}$の経験ラデマッハ複雑度$\hat{\mathfrak{R}}_S(\mathcal{G})$は，

$$\hat{\mathfrak{R}}_S(\mathcal{G}) = \mathbb{E}_\sigma \left[\sup_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i g(x_i) \right]$$

として定義される．

$\mathbb{E}_\sigma$: $\sigma_1, \ldots, \sigma_n$に関する期待値

経験ラデマッハ複雑度の直感的な解釈

- Setting
  - 2値判別
  - $\mathrm{sign}(g(x_i))$: 判別器
  - $\sigma_i \in \{+1, -1\}$: x_iのラベル を予測

$\rightsquigarrow \sigma_i g(x_i) > 0$ならば予測が正しい．
$\rightsquigarrow \sigma_i g(x_i)$が大きな値をとるとき，$g \in \mathcal{G}$によってデータ$(x_i, \sigma_i)$が十分よく学習されている．

$\rightsquigarrow$ 経験ラデマッハ複雑度は，S上のランダムなラベル付け$(x_1, \sigma_1), \ldots, (x_n, \sigma_n)$に対する，関数集合$\mathcal{G}$のデータへの適合度を平均的に測っている量である．

Def 2.5 (ラデマッハ複雑度)

入力点$S = (x_1, \dots, x_n)$が分布Dに従う確率変数のとき，$\mathcal{G}$の経験ラデマッハ複雑度の期待値

$$\mathfrak{R}_n(\mathcal{G}) = \mathbb{E}_{S \sim D}[\hat{\mathfrak{R}}_S(\mathcal{G})]$$

を$\mathcal{G}$のラデマッハ複雑度という．

経験ラデマッハ複雑度の性質を以下に示す．
入力点の集合Sについて期待値をとれば，ラデマッハ複雑度についても成立する．

Th 2.6

$\mathcal{G}, \mathcal{G}_1, \ldots, \mathcal{G}_k$: 実数値関数の集合

1. $\mathcal{G} \subset \mathcal{G}_2 \Rightarrow \hat{\mathfrak{R}}_S(\mathcal{G}_1) \leq \hat{\mathfrak{R}}_S(\mathcal{G}_2)$
2. $\forall c \in \mathbb{R},\ \hat{\mathfrak{R}}_S(c\mathcal{G}) = |c|\hat{\mathfrak{R}}_S(\mathcal{G})$
3. $\hat{\mathfrak{R}}_S(\mathcal{G}) = \hat{\mathfrak{R}}_S(\mathrm{conv}\mathcal{G})$
4. [Talagrand's lemma]
   $\phi : \mathbb{R} \to \mathbb{R}$: リプシッツ連続，$L$: リプシッツ定数
   $\rightsquigarrow \hat{\mathfrak{R}}_S(\phi \circ \mathcal{G}) \leq L\hat{\mathfrak{R}}_S(\mathcal{G})$
   - $\phi \circ \mathcal{G} = \{x \mapsto \phi(f(x)) \mid f \in \mathcal{G}\}$

Th 2.6 (続き)

5. $\hat{\mathfrak{R}}_S(\sum_{i=1}^k \mathcal{G}_i) \leq \sum_{i=1}^k \hat{\mathfrak{R}}_S(\mathcal{G}_i)$
   - $\sum_{i=1}^k \mathcal{G}_i = \{\sum_{i=1}^k g_i \,|\, g_i \in \mathcal{G}_i,\ i = 1, \dots, k\}$
6. $\mathcal{Y}$: 有限集合, $\mathcal{G} \subset \{f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}\}$, $\mathcal{G}_y = \{f(\cdot, y) : \mathcal{X} \to \mathbb{R} \,|\, f \in \mathcal{G}\}$
   $\rightsquigarrow \hat{\mathfrak{R}}_S(\mathcal{G}) \leq \sum_{y \in \mathcal{Y}} \hat{\mathfrak{R}}_S(\mathcal{G}_y)$
7. $\mathcal{G}_1, \dots, \mathcal{G}_k \subset \{f : \mathcal{X} \to \mathbb{R}\}$
   $\mathcal{G} = \{\max\{f_1, \dots, f_k\} \,|\, f_1 \in \mathcal{G}_1, \dots, f_k \in \mathcal{G}_k\}$
   $\rightsquigarrow \hat{\mathfrak{R}}_S(\mathcal{G}) \leq \sum_{\ell=1}^k \hat{\mathfrak{R}}_S(\mathcal{G}_\ell)$

証明は略 (別資料に載せる予定).

ラデマッハ複雑度とVC次元の関連

- $\mathcal{H}$: 2値ラベル$\{+1,-1\}$に値をとる仮説集合．$\mathrm{VCdim}(\mathcal{H})=d$.
- $A := \{(h(x_1),\ldots,h(x_n)) \in \{+1,-1\}^n \mid h \in \mathcal{H}\}$

$\rightsquigarrow$ Lemma 2.1より，$n \geq d$のとき

$$|A| = \Pi_{\mathcal{H}}(x_1,\ldots,x_n) \leq \left(\frac{en}{d}\right)^d$$

となる．このとき，$S=\{x_1,\ldots,x_n\}$における$\mathcal{H}$の経験ラデマッハ複雑度は，

$$\hat{\mathfrak{R}}_S(\mathcal{H}) = \frac{1}{n}\mathbb{E}_\sigma\left[\sup_{z\in A}\sum_{i=1}^n \sigma_i z_i\right] \leq \sqrt{\frac{2d}{n}\log\frac{en}{d}} \quad (2.1)$$

となる．$|S| \geq d$となる任意のSで(2.1)が成り立つので，ラデマッハ複雑度$\mathfrak{R}_n(\mathcal{H})$についても同じ不等式が成立する．

例2.4 (有限集合)

関数の有限集合$\mathcal{G} \subset \{g : \mathcal{Z} \to \mathbb{R}\}$の経験ラデマッハ複雑度を計算する．
集合$\{(g(z_1), \ldots, g(z_n)) \in \mathbb{R}^n \mid g \in \mathcal{G}\}$に対して，マサールの補題を用いると，

$$\hat{\mathfrak{R}}_S(\mathcal{G}) = \mathbb{E}_\sigma \left[\max_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i g(z_i) \right] \leq \max_{g \in \mathcal{G}} \sqrt{\sum_{i=1}^{n} g(z_i)^2} \cdot \frac{\sqrt{2 \log |\mathcal{G}|}}{n}$$

ここで，有界性$\|g\|_\infty \leq r$，$g \in \mathbb{G}$を仮定すると，

$$\hat{\mathfrak{R}}_S(\mathcal{G}) \leq r \sqrt{\frac{2 \log |\mathcal{G}|}{n}}$$

となる．ラデマッハ複雑度$\mathfrak{R}_n(\mathcal{G})$についても同じ上界が得られる．

一様大数の法則

- 一様大数の法則
  - VC次元を用いて予測判別誤差を評価したTh 2.2を拡張
  - ラデマッハ複雑度は，一様大数の法則における誤差に相当
  - 有界な関数の集合に対して成立
    - c.f. Th 2.2, 0-1損失

Th 2.7 (一様大数の法則)

集合$\mathcal{Z}$から有界区間$[a, b]$への実数値関数の集合を$\mathcal{G} \subset \{f : \mathcal{Z} \to [a, b]\}$とする．
また，$Z_1, \dots, Z_n \overset{\text{i.i.d.}}{\sim} D$, $Z \sim D$とする．
$\rightsquigarrow \forall\delta \in (0, 1)$に対して，分布$D^n$の下で$1 - \delta$以上の確率で次式が成立．

Th 2.7 (続き)

$$\sup_{g\in\mathcal{G}}\left\{\mathbb{E}[g(Z)]-\frac{1}{n}\sum_{i=1}^{n}g(Z_i)\right\}\leq 2\mathfrak{R}_n(\mathcal{G})+(b-a)\sqrt{\frac{\log(1/\delta)}{2n}}$$

同様の不等式が左辺の符号を逆転したものについても成立する．
よって，絶対誤差については，分布D^nの下で$1-\delta$以上の確率で

$$\sup_{g\in\mathcal{G}}\left|\mathbb{E}[g(Z)]-\frac{1}{n}\sum_{i=1}^{n}g(Z_i)\right|\leq 2\mathfrak{R}_n(\mathcal{G})+(b-a)\sqrt{\frac{\log(2/\delta)}{2n}}$$

が成立する．

証明は略 (別資料に載せる予定)．

ラデマッハ複雑度を用いて予測判別誤差の確率的上界を求める例

- $\mathcal{H} \subset \{h : \mathcal{X} \to \{+1, -1\}\}$: 2値判別のための有限仮説集合
    - $h_0 \in \mathcal{H}$と仮定 (h_0: ベイズ規則)
- $\mathcal{G} = \{(x, y) \mapsto \mathbf{1}[h(x) \neq y] \mid h \in \mathcal{H}\}$ (2.5)

$\rightsquigarrow$ 例2.4と$|\mathcal{G}| = |\mathcal{H}|$より，$\mathcal{G}$のラデマッハ複雑度は，

$$\mathfrak{R}_n(\mathcal{G}) \leq \sqrt{\frac{2 \log |\mathcal{H}|}{n}}$$

となる．

一様大数の法則より，学習データの分布の下で，$1-\delta$以上の確率で

$$\max_h |R_{\mathrm{err}}(h) - \hat{R}_{\mathrm{err}}(h)| \leq 2\sqrt{\frac{2\log|\mathcal{H}|}{n}} + \sqrt{\frac{\log(2/\delta)}{2n}}$$

が成り立つ．よって，

$$R_{\mathrm{err}}(h_S) \leq R_{\mathrm{err}}(h_0) + O_p\left(\sqrt{\frac{\log|\mathcal{H}|}{n}}\right)$$

が成り立つ．

Remark

VC次元を用いた評価 (例 2.1) では対数因子$\log(n/\log|\mathcal{H}|)$があった．