

主成分分析の汎化誤差などについて

On Generalization Errors of Principal Component Analysis

赤間陽二^{1*} 上野康隆¹
Yohji Akama¹ and Yasutaka Uwano¹

¹ 東北大学 大学院理学研究科 数学専攻
¹ Mathematical Institute, Tohoku University

Abstract: We are concerned with principal component analyses (PCA), which reduce the dimensionality of data, and we formulate them as empirical risk minimizations of Vapnik where the class of the loss functions of the empirical risk minimization is unbounded and non-negative. The classes of the loss functions of PCA induce natural classes of geometric concepts. In order to evaluate the generalization errors of PCA, we first evaluate the VC dimensions of the classes by using techniques of discrete geometry, such as Voronoi diagrams. Then, from any unbounded class of non-negative loss functions, we derive a distribution-independent sample complexity of the risk minimization, which corresponds to the sample complexity of PAC-learning.

1 はじめに

統計的学習理論 [7] の枠組では、学習の一般的なモデルの構成要素は、(i) 空間 Z から得られる訓練データ z が従う未知の確率分布関数 $F(z)$ 、(ii) 表現クラス Λ 、および (iii) 損失関数 $Q : Z \times \Lambda \rightarrow \mathbb{R}$ である。ただし $Q(z, \alpha)$ は訓練データ z と表現 α に関する損失を表す。学習の目標とは、リスク

$$R(\alpha) = \int Q(z, \alpha) dF(z), \quad \alpha \in \Lambda,$$

を最小化する

$$\alpha_0 \in \Lambda \text{ s.t. } R(\alpha_0) = \min_{\alpha \in \Lambda} R(\alpha)$$

を、 F に独立に従う訓練データ

$$z_1, \dots, z_l \quad (1)$$

から推定することである。

例えば、教師つき学習では各 z_i はデータ x_i とそれに対する教師の反応 y_i の組であり、学習機械は適当な関数族 $f(x, \alpha), \alpha \in \Lambda$ を実現することができる。特に $f(x, \alpha) \in \{0, 1\}$ かつ、損失関数 $Q((x, y), \alpha)$ が $y = f(x, \alpha)$ ならば 1 さもなければ 0 であるとき、パターン識別問題に対応する。

学習方法のひとつである経験リスク最小化原理とは、リスクのかわりに、経験データ (訓練データ)(1) に基づ

く経験リスク

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^l Q(z_i, \alpha)$$

を最小化する

$$\alpha_l \in \Lambda \text{ s.t. } R_{\text{emp}}(\alpha_l) = \min_{\alpha \in \Lambda} R_{\text{emp}}(\alpha)$$

で α_0 を近似しようというものである。この原理は一般的であり、さまざまな統計的推論を含む。パターン識別、最尤法、回帰分析などである。

汎化誤差

$$R(\alpha_l) - R(\alpha_0)$$

の l に関する上界は、訓練データの分布関数 $F(z)$ と独立に与えることができるが、その上界は損失関数族の VC 次元 (大まかにいうと表現力の強さ・繊細さ) [6] により支配されている。汎化誤差の上界はサンプル複雑度を導き、これは言語クラスの PAC 学習におけるサンプル複雑度¹ [1] に対応する。

そこで、我々は統計的推論における PRINCIPAL COMPONENT ANALYSIS (主成分分析, PCA) をリスク最小化原理として定式化する (3 節)。PCA に現れる損失関数族の VC 次元 (の下界) を、Voronoi 図などによる議論により求める (4 節)。これにより主成分分析の学習問題の本質的難しさがわかることになる。言語の PAC 学習の場合はいわば損失関数族が $\{0, 1\}$ 値であるが、PCA

*東北大学 大学院理学研究科 数学専攻
〒980-8578 仙台市青葉区 荒巻字青葉 6 番 3 号
E-mail: akama@math.tohoku.ac.jp

¹データの分布に独立に、 $1 - \eta$ 以上の確率で、誤差が高々 ε であるような仮説を出力するために十分な入力データの個数のことであり、言語クラスの VC 次元で支配される [1, 定理 2.1].

の場合は損失関数族が非有界非負値である。損失関数族が非有界非負値の場合一般に対して、構成的な分布独立な汎化誤差上界からサンプル複雑度を求める (5 節)。最後の節では今後の課題を述べる。

2 VC 次元, 成長関数に関する準備

$S \subseteq Z, \mathcal{C} \subseteq \wp(Z)$ に対して $\Pi_{\mathcal{C}}(S) := \{A \cap S \mid A \in \mathcal{C}\} \subset \wp(S)$ とする。 S が \mathcal{C} によって細分されるというのは $\Pi_{\mathcal{C}}(S) = \wp(S)$ が成り立つことと定義される。集合 A に対して $[A]_l := \{Y \subset A \mid \#Y = l\}$, ($l \in \mathbb{N}$) とする。

\mathcal{C} を決めると、いくらでも要素数の多い適当な有限集合 S があって \mathcal{C} によって細分できるか、あるいは、適当な数 $d \in \mathbb{N}$ があって、 d より大きな濃度の有限集合は \mathcal{C} ではどれも細分できないかのいずれかである [7, 定理 4.3a, 146 頁]。後者の場合、 \mathcal{C} の VC 次元 $\text{VCdim}(\mathcal{C})$ は、そのような d の最小値として定義され、前者の場合、 $\text{VCdim}(\mathcal{C})$ を ∞ と定義される。

実数値関数族 \mathcal{F} の VC 次元とは、集合族 $\mathcal{C}_{\mathcal{F}} := \{z \in Z; f(z) \geq \beta\}; f \in \mathcal{F}, \beta \in \mathbb{R}\}$ の VC 次元 h である。 \mathcal{F} の成長関数 $G_{\mathcal{F}}(l) := \log \sup_{S \in [Z]_l} \#\Pi_{\mathcal{C}_{\mathcal{F}}}(S)$ について

$$G_{\mathcal{F}}(l) \leq h(\log \frac{l}{h} + 1) \quad (l \geq h). \quad (2)$$

が成立する [4]。

3 損失関数による PCA の定式化

以下 d 次元ユークリッド空間を \mathbb{E}^d で表す。

PCA の定義は [3, 111 頁] に従う。 PCA は、データ $x_1, \dots, x_l \in \mathbb{E}^d$ を $d' \in [0, d] \cap \mathbb{Z}$ 次元 Affine 空間 $\{Tz + \mathbf{b}; z \in \mathbb{E}^{d'}\}$ ($T \in M_{d,d'}(\mathbb{E}), \text{rank}T = d', \mathbf{b} \in \mathbb{E}^d$) で近似することである。

そこで、我々は

$$\Lambda_{d'}^d = \{(T, \mathbf{b}) \in M_{d,d'}(\mathbb{R}) \times \mathbb{E}^d; \text{rank}T = d'\}, \\ (d' \in [1, d] \cap \mathbb{Z});$$

$$\Lambda_0^d = \mathbb{E}^d, \quad (d' = 0)$$

とおき、 $d' \in [1, d] \cap \mathbb{Z}$ の場合、 $\alpha = (T, \mathbf{b}) \in \Lambda_{d'}^d$ と x に関する損失 $Q_{d'}^d(x, \alpha)$ を、 x とその Affine 空間 $\{Tz + \mathbf{b}; z \in \mathbb{E}^{d'}\}$ との距離の自乗で定義し、 $d' = 0$ の場合、 $\alpha = \mathbf{b} \in \mathbb{E}^d = \Lambda_0^d$ と x に関する損失 $Q_0^d(x, \alpha)$ を、 x と \mathbf{b} との距離の自乗で定義する。 すなわち

$$Q_{d'}^d(x, (T, \mathbf{b})) = \left\| \left(T(T^{\top}T)^{-1}T^{\top} - E \right) (x - \mathbf{b}) \right\|^2, \\ (d' \in [1, d] \cap \mathbb{Z});$$

$$Q_0^d(x, \mathbf{b}) = \|x - \mathbf{b}\|^2, \quad (d' = 0).$$

ここで右上添字の \top は転置を表す。 PCA の損失関数族は

$$\mathcal{F}_{d'}^d = \{Q_{d'}^d(z, \alpha); \alpha \in \Lambda_{d'}^d\}$$

となる。

4 PCA の損失関数族の VC 次元

PCA の損失関数族 $\mathcal{F}_{d'}^d$ の VC 次元は次の集合族の VC 次元となる：

定義 4.1 $\mathcal{C}_{d'}^d := \mathcal{C}_{\mathcal{F}_{d'}^d}$, すなわち、

$$\left\{ \begin{array}{l} \{x \in \mathbb{E}^d; x \text{ と } H \text{ の距離は } \beta \text{ 以上} \} \\ ; H \text{ は } \mathbb{E}^d \text{ の } d' \text{ 次元 Affine 空間, かつ } \beta \geq 0 \end{array} \right\}$$

とする。特に \mathcal{C}_{d-1}^d は「平行」な半空間の和集合全体となる。

一般に、集合族 \mathcal{C} の要素ごとの補集合全体 $\mathcal{C}^c := \{Z \setminus C; C \in \mathcal{C}\}$ の VC 次元は \mathcal{C} のそれと等しいから、 $\text{VCdim}(\mathcal{C}_{d-1}^d)$ は $\{\mathbb{E}^d \text{ の適当な厚さの板}\} \cup \{\emptyset\}$ の VC 次元と等しい。

一般に、 $v \geq \text{VCdim}(\mathcal{C})$ を示すときには、 v より大きな要素数のどんな集合 S も \mathcal{C} では細分できないという強い主張を証明する必要があるので、離散幾何や組み合わせ幾何の命題を使う。

次の命題は、代数幾何学で Veronese mapping (または線形化) として知られている手法による。この手法は \mathbb{E}^d の Voronoi 図を \mathbb{E}^{d+1} の凸多面体に変換するときにも使われる。

命題 4.1 ([5, 命題 10.3.2]) $\mathbb{R}[x_1, \dots, x_d]_{\leq D}$ を、次数が高々 D の d 変数実多項式全体とする。

$$\left\{ \{x \in \mathbb{E}^d; p(x) \geq 0\}; p \in \mathbb{R}[x_1, \dots, x_d]_{\leq D} \right\}$$

の VC 次元は $\binom{d+D}{d}$ 以下。

PCA の損失関数は、 $x = (x_1, \dots, x_d)^{\top}$ の次数が高々 2 の d 変数実多項式だから

命題 4.2 $\text{VCdim}(\mathcal{C}_e^d) \leq \binom{d+2}{d} \quad (0 \leq e < d)$ 。

特に $e = 0$ のときは \mathbb{E}^d から \mathbb{E}^{d+1} への Veronese mapping が取れるため次が言える。

命題 4.3 ([5, 245 頁]) $\text{VCdim}(\mathcal{C}_0^d) \leq d + 2$ 。

VC 次元の次の下界たちは初等的に証明できる。

定理 4.1 1. $\text{VCdim}(\mathcal{C}_0^1) = 2, \text{VCdim}(\mathcal{C}_1^2) = 5$ 。

2. $\text{VCdim}(\mathcal{C}_i^{d+1}) \geq \text{VCdim}(\mathcal{C}_i^d) + 1 \quad (d > i = 0, 1)$ 。

3. $\text{VCdim}(C_{e+1}^{d+1}) \geq \text{VCdim}(C_e^d) + 1 \quad (d > e \geq 0).$
4. $\text{VCdim}(C_e^d) \geq d + 1 \quad (d > e \geq 0).$
5. $\text{VCdim}(C_e^d) \geq d + 3 \quad (d > e \geq 1).$

また, $\text{VCdim}(C_0^2)$ の正確な値を求めるときは Voronoi 図の詳しい性質を用いる.

定理 4.2 ([2, 定理 7.4]) \mathbb{E}^2 の離散集合 P の Voronoi 図 $\text{Vor}(P)$ について以下が成り立つ.

1. 点 $v \in \mathbb{E}^2$ が $\text{Vor}(P)$ の vertex である.
 \Leftrightarrow 点 v の largest empty circle (v を中心とする, 内部に P の点を含まない最大半径の円盤) はその境界に点を 3 点以上含む.
2. 点 $p_1, p_2 \in P$ について $\overline{p_1 p_2}$ (p_1 と p_2 を結ぶ線分) の垂直二等分線の一部が $\text{Vor}(P)$ の edge として現れている.
 \Leftrightarrow ある点 $q \in \mathbb{E}^2$ が存在して, 点 q の largest empty circle の境界には点 p_1, p_2 しか含まれない.

まず正三角形の頂点を考えれば, $\text{VCdim}(C_0^2) \geq 3$.

いま $S \in [\mathbb{E}^2]_4$ が C_0^2 で細分できると仮定して矛盾を導く.

点集合 S の Voronoi 図 $\text{Vor}(S)$ を考える. このとき定理 4.2(2) から, 2 点 $p, q \in S$ が C_0^2 で切り出されるならば \overline{pq} の垂直二等分線の一部として edge が現れる. なぜなら, いま 2 点 $p, q \in S$ が円盤 C で切り出されたとすると, 中心に関して縮小し, p または q と交わったら交わった点を中心として縮小することで境界に p, q を含む円盤 C' が得られる. この円盤 C' は内部に S の点を含まない. この事実と定理 4.2(2) を組み合わせればよい. このことに注意すると, 円盤全体が S の任意の 2 点を切り出せることより, $\text{Vor}(S)$ には edge が $\binom{4}{2} = 6$ 個必要である. このとき平面グラフに対する Euler の公式から $\text{Vor}(S)$ の vertex は 3 個である.

ここで, この vertex を頂点にもつ三角形の内部にある S の点を q , その他の S の点を p_i ($i = 1, 2, 3$) とおく. ただし p_1, p_2, p_3 は反時計回りでこの順に並んでいるとする. すると Voronoi 図の性質から, q と各 p_i を結ぶ線分は q を囲む 3 個の edge とそれぞれ垂直に交わる. ここでその交点を p'_i ($i = 1, 2, 3$) とおくと, q は三角形 $p'_1 p'_2 p'_3$ の内部に存在する. なぜなら, r を図 1 の位置の vertex としたとき, 四角形 $p'_1 r p'_2 q$ において $\angle r p'_1 q = \angle q p'_2 r = \pi/2$ より, $\angle p'_1 r p'_2 + \angle p'_2 q p'_1 = \pi$ である. ここで $\angle p'_1 r p'_2$ は三角形の内角だから $0 < \angle p'_1 r p'_2 < \pi$. したがって $0 < \angle p'_2 q p'_1 < \pi$. 以上より q は三角形 $p'_1 p'_2 p'_3$ の内部に存在することがわかる. 三角形 $p_1 p_2 p_3$ はその内部に三角形 $p'_1 p'_2 p'_3$ を含んでいるので, q は特に三角形 $p_1 p_2 p_3$ の内部に存在する. すると $\{p_1, p_2, p_3\}$ は円盤で切り出せないことになり矛盾. ゆえに $\text{VCdim}(C_0^2) < 4$.

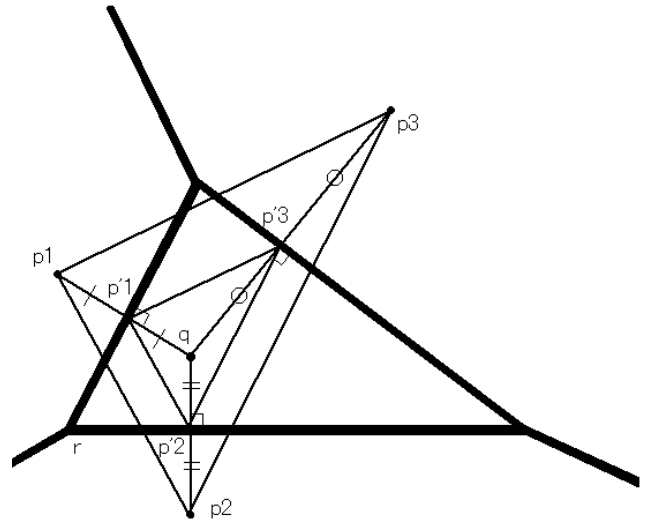


図 1: Voronoi 図 $\text{Vor}(\{p_1, p_2, p_3, q\})$ (太線), 三角形 $p_1 p_2 p_3$, 三角形 $p'_1 p'_2 p'_3$.

定理 4.3 $\text{VCdim}(C_0^2) = 3$.

一般に VC 次元が高くなると, PAC 学習では必要となるサンプル数は大きくなり, Vapnik 流の統計学習では構造リスク最小化を無視すれば, すなわち, リスク最小化原理に基づけば, 汎化誤差はサンプル数 l の増大に対して減少する. 従って, C_d^d の PAC 学習のサンプル複雑度, あるいは, PCA の Vapnik 流の統計学習の収束速度の目安が得られる.

5 損失関数族が非有界非負値の場合のサンプル複雑度

PCA の汎化誤差の構成的上界およびサンプル複雑度を調べるために, より一般的な枠組みで調べる.

5.1 サンプル複雑度の導出

損失関数族が非有界非負値の場合のサンプル複雑度を求める準備として次の定理を用いる.

定理 5.1 ([7, 定理 5.4(1)]) $p > 2$,

$$\mathcal{F} := \{Q(z, \alpha); \alpha \in \Lambda\},$$

が非有界 (i.e., $\neg \exists \beta \forall z \in Z \forall f \in \mathcal{F}. f(z) \leq \beta$), かつ $Q(z, \alpha) \geq 0$ とする. このとき高々

$$4 \exp \left\{ \left(\frac{G_{\mathcal{F}}(2l)}{l} - \frac{\varepsilon^2}{4} \right) l \right\}$$

の確率で

$$\sup_{\alpha \in \Lambda} \frac{\int Q(z, \alpha) dF(z) - \frac{1}{l} \sum_{i=1}^l Q(z_i, \alpha)}{(\int Q(z, \alpha)^p dF(z))^{1/p}} > \varepsilon a(p)$$

が成立する. ただし

$$a(p) := \sqrt[p]{\frac{1}{2} \left(\frac{p-1}{p-2} \right)^{p-1}}.$$

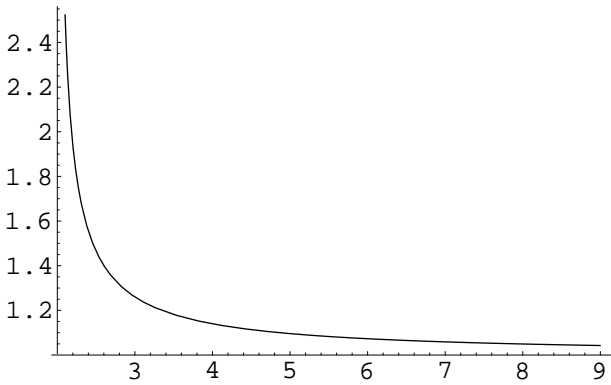


図 2: $a(p) = \sqrt[p]{\frac{1}{2} \left(\frac{p-1}{p-2} \right)^{p-1}}$ のグラフ. $\lim_{p \rightarrow 2+0} a(p) = \infty, \lim_{p \rightarrow \infty} a(p) = 1$ である単調減少関数である.

これにより, 損失関数族が非有界非負値である場合の分布独立な構成的汎化誤差上界を導くことができる:

定理 5.2 損失関数族 $\mathcal{F} := \{Q(z, \alpha); \alpha \in \Lambda\}$ が非有界非負値で, 確率変数 $Q(z, \alpha)$ の $p (> 2)$ 次のモーメントと期待値周りの p 次モーメントのそれぞれの p 乗根を, 期待値で割ったものが, ともに $\tau > 0$ で上から抑えられるとする. このとき少なくとも $1 - \eta$ の確率で

$$R(\alpha_l) - R(\alpha_0) \leq \left(\frac{R(\alpha_0) \tau \left(a(p) \sqrt{\mathcal{E}_{\mathcal{F}, \eta}(l)} + \sqrt{\frac{4}{l\eta}} \right)}{1 - \tau a(p) \sqrt{\mathcal{E}_{\mathcal{F}, \eta}(l)}} \right)_{\infty}$$

が成立する [7, (5.51)]. ただし

$$\begin{aligned} \mathcal{E}_{\mathcal{F}, \eta}(l) &:= 4 \frac{G_{\mathcal{F}}(2l) - \log(\eta/8)}{l} \\ \left(\frac{a}{b} \right)_{\infty} &:= \begin{cases} \frac{a}{b}, & b > 0; \\ \infty, & \text{otherwise.} \end{cases} \end{aligned}$$

$\mathcal{E}_{\mathcal{F}, \eta}(l)$ は $l \geq h/2$ のとき, (2) より次で上から抑えられる:

$$\check{\mathcal{E}}_{h, \eta}(l) := 4 \frac{h \left(\log \left(\frac{2l}{h} \right) + 1 \right) - \log \left(\frac{\eta}{8} \right)}{l}.$$

$$\begin{array}{ccccccc} 0 & \dots & \frac{h}{2} \sqrt{\frac{\eta}{8}} & \dots & \frac{h}{2} & \dots & \infty \\ -\infty & \nearrow & 8 / \sqrt{\frac{\eta}{8}} & \searrow & 8 \left(1 + \frac{1}{h} \log \frac{8}{\eta} \right) & \searrow & 0 \end{array}$$

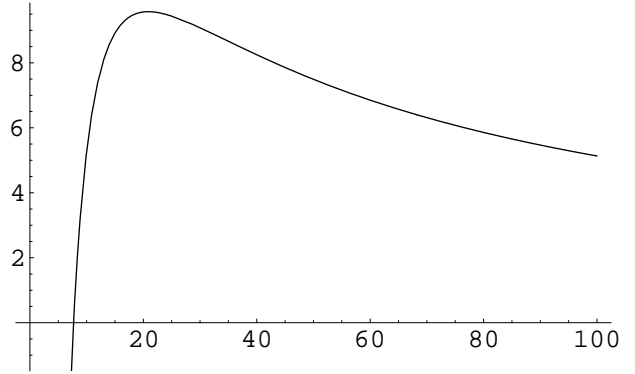


図 3: $\check{\mathcal{E}}_{h, \eta}(l) = 4 \frac{h \left(\log \left(\frac{2l}{h} \right) + 1 \right) - \log \left(\frac{\eta}{8} \right)}{l}$ の l に関する増減表 (上図), および, VC 次元 $h = 50$, 信頼度 $\eta = 0.001$ の場合の概形 (下図).

すると定理 5.2 の $\mathcal{E}_{\mathcal{F}, \eta}(l)$ を $\check{\mathcal{E}}_{h, \eta}(l)$ に置き換えたものが成立する.

そこからサンプル複雑度に関する次の定理が従う.

定理 5.3

$$l_{h, \eta}(c) = \inf \{ l' > 0; \check{\mathcal{E}}_{h, \eta}(l) < c \ (\forall l \geq l') \}.$$

とすると以下が成立する.

1. $R(\alpha_0) = 0$ の場合.

$$P \{ R(\alpha_l) = 0 \} \geq 1 - \eta$$

の十分条件は $l > l_1$, ただし

$$l_1 := \max \left\{ \frac{h}{2}, l_{h, \eta} \left(\frac{1}{\tau^2 a(p)^2} \right) \right\}.$$

2. $R(\alpha_0) > 0$ の場合.

$$P \left\{ \frac{R(\alpha_l) - R(\alpha_0)}{R(\alpha_0)} \leq \varepsilon \right\} \geq 1 - \eta$$

の十分条件は $l > l_2$, ただし l_2 は $h/2$ と

$$l_{h,\eta} \left(\frac{1}{\tau^2 a(p)^2} \frac{\varepsilon^2}{\left((1+\varepsilon) + \frac{1}{a(p)} \sqrt{\frac{4}{\eta}} \right)^2} \right)$$

の最大値であり, l_1 以上である.

$l_{h,\eta}(c)$ は, $\check{\mathcal{E}}_{h,\eta}(l) = c$ に解があればより大きいほうを返し, そうでなければ 0 を返す関数である. $\check{\mathcal{E}}_{h,\eta}(l)$ の概形 (図 3) から分かるとおり, c に関して単調減少である.

5.2 サンプル複雑度のパラメータへの依存

我々の導出したサンプル複雑度は, 非有界非負値な損失関数族の場合であるが, 参考となるものとして言語の PAC 学習における Blumer によるサンプル複雑度が挙げられる.

定理 5.4 ([1, 定理 2.1]) 言語のクラス \mathcal{C} の VC 次元が $d < \infty$ ならば, $1 - \eta$ 以上の確率で, 誤差が高々 ε であるような仮説を出力するために十分な入力データの個数 l_b は $\max\left(\frac{4}{\varepsilon} \log \frac{2}{\eta}, d \frac{8}{\varepsilon} \log \frac{13}{\varepsilon}\right)$.

l_1, l_2, l_b の信頼度 η , 誤差 ε , VC 次元 h への依存は定性的には共通している. 実際

$$\eta < 8e^{h(1-\frac{1}{8a(p)^2\tau^2})} \Rightarrow \frac{\partial l_2}{\partial \eta} < \frac{\partial l_1}{\partial \eta} < 0, \\ \frac{\partial l_2}{\partial h} > \frac{\partial l_1}{\partial h} > 0, \\ \frac{\partial l_2}{\partial \varepsilon} < 0$$

が証明できる. またこれらの式から, l_2 の方が l_1 より η と h への依存が著しいことがわかる (これは手元に正解がない学習問題の困難さを示唆している).

しかし l_1 と l_2 は前述のとおり損失関数族が非有界であるため, データ分布の (期待値周りの) モーメントにも依存する. それは今の場合 τ や p というパラメータを通して現れている.

p が大きいときは, z の確率分布が高次のモーメントまで持つことになり, 原点から遠いところでの z の分布は問題を起こさないし, $a(p)$ は小さい.

τ が小さいときは, z の分布は原点から遠い部分の寄与は小さく, また, z の確率密度関数の歪度も尖度も小さいため対称に近く尖りも少ない.

$a(p)$ が大きいときは z の分布は原点から遠いところでも寄与が重く, また, τ が大きいときは分布の対称性

も弱いかがりが強くなる. このように特異性が強くなるため訓練データが多く必要になる. 実際,

$$\eta < 8e^{h(1-\frac{1}{8a(p)^2\tau^2})} \Rightarrow \frac{\partial l_2}{\partial p} < 0, \quad \frac{\partial l_1}{\partial p} < 0, \\ \frac{\partial l_2}{\partial \tau} > 0, \quad \frac{\partial l_1}{\partial \tau} > 0$$

が証明できる.

6 今後の課題

PCA の汎化誤差の分析に離散幾何的なアプローチを行ったが, 今後は VCdim (\mathcal{C}_0^d) をはじめ VCdim (\mathcal{C}_e^d) に対する更なる評価を試み, また FISHER DISCRIMINANT ANALYSIS (フィッシャー判別分析, FDA) も同様の取り扱いができないか吟味する予定である. PCA も FDA も Rayleigh 商の最大化で計算されるが, その第一固有値が他の固有値に近接しているときは計算の条件 (誤差に対する頑健性など) が悪くなることが知られている. そのことと, 分布独立な構造的汎化誤差の上界 (5 節) との関連も考えていく.

参考文献

- [1] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. Assoc. Comput. Mach.*, 36(4):929–965, 1989.
- [2] Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. *Computational geometry*. Springer-Verlag, Berlin, 1997. Algorithms and applications.
- [3] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000. パターン識別, 尾上守夫監訳, 新技術者コミュニケーションズ, 2001.
- [4] R. M. Dudley. A course on empirical processes. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 1–142. Springer, Berlin, 1984.
- [5] Jiří Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.
- [6] Vladimir Vapnik. *Estimation of dependences based on empirical data*. Springer Series in Statistics. Springer-Verlag, New York, 1982. Translated from the Russian by Samuel Kotz.
- [7] Vladimir N. Vapnik. *Statistical learning theory*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons Inc., New York, 1998. A Wiley-Interscience Publication.