第 I 部

Introduction and Statement

1 Introduction

1.1 Abstract

最初に論文に記載されている内容を説明する. 数学的に示していることをざっくりとした (数学的には不正確な) 日本語で説明すると以下の二点である.

- 多様体仮説がある多様体に対して, 成り立つと仮定する. この時, 実際に多様体仮説が成り立つ多様体 M を数学的に構成した.
- 上の多様体 M を具体的に求めるアルゴリズムを構成した.

(この論文) では, 入力となるデータに関する性質を記述できていないので, 多様体仮説が成り立つことは確認できない. だが, 多様体仮説が成り立つかはわからないが, 多様体仮説が成り立つならば, 求めたい多様体 (といっていいもの) が定義でき, さらにそれを実際に求めるアルゴリズムを構成できたので, 実際に実験してみればよいと主張している. もし数学的に多様体仮説を示すのであれば, 入力データの偏りに相当する仮定が必要になる. また, 多様体仮説を聞いて, 多様体の形に基づいた, クラスタリングが定義できるのではないかと予想していたが, この論文では特に記述はなかった. (例えば, 連結成分等でうまく定義できて欲しいものだが...)

結果を具体的に記述するため, 最低限の数学的な用語を定義する.

1.2 Notation and Definition

Definition 1.1. 有限次元とは限らない $\mathcal{H}$ を $\mathbb{R}$-ベクトル空間が以下を満たす時, ヒルベルト空間という.

- 内積が定義できる. すなわち, 以下を満たす写像 $<\cdot,\cdot>: \mathcal{H}\times\mathcal{H}\to\mathbb{R}$ が存在する.
  1. 任意の $x_1, x_2, y \in \mathcal{H}$ に対し,$<x_1+x_2, y> = <x_1, y> + <x_2, y>$
  2. 任意の $x, y \in \mathcal{H}, a \in \mathbb{R}$ に対し,$< ax, y >= a < x, y >$
  3. 任意の $x, y \in \mathcal{H}$ に対し,$< x, y >=< y, x >$
  4. 任意の $x \in \mathcal{H}$ に対し,$<x, x> \geq 0$ であり,$<x, x> = 0$ と $x = 0$ は同値である.
- 内積が定める距離について完備である.

Remark. 上で定めた内積のことを正定値対象双線型形式という.

Remark. $x, y \in \mathcal{H}$ に対し,$d(x,y) := \sqrt{<x-y, x-y>}$ とすると距離の公理を満たす. 距離の公理とは以下の *3* つのこと.

1. $d(x,y) = d(y,x)$
2. $d(x,y) = 0$ は $x = y$ と同値
3. $d(x,y) + d(y,z) \geq d(x,z)$

Remark. 距離空間 X が完備とは, 任意のコーシー列 $\{x_n\}_{n \in \mathbb{Z}_{\geq 0}}$ が, ある $x \in X$ に対する収束列となること.

Example 1.2. $\mathbb{R}^n$ は標準内積により, ヒルベルト空間になる.

この論文で使われる"多様体"の定義をする.

Definition 1.3 (reach)**.** M を $\mathcal{H}$ の部分集合とし.$x \in \mathcal{H}$ に対し,$d(x,M)$ を $\inf_{y \in M} d(x,y)$ で定める. 実数 τ が M の ***reach*** とは,$d(x,M) < \tau$ となる任意の x に対し, $y \in M$ が存在し,y と異なる任意の $z \in M$ に対し,$d(x,z) > d(x,M) = d(x,y)$ となること.

Example 1.4. 半径 r の円周 $\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\}$ の *reach* は r になる.

Remark. *reach* の定義は初めて見た. 恐らく多様体をクラスタリングするために, 定義されているのだと思う.

Definition 1.5. 位相空間 X の部分集合 Y が**稠密**とは Y の閉包が X に一致することをいう. 位相空間 M が稠密な加算部分集合を持つ時, **可分**という.

Example 1.6. 実数体 $\mathbb{R}$ に通常のユークリッド位相を定めた時, 有理数体 $\mathbb{Q}$ は稠密になる. また $\mathbb{Q}$ は加算集合のため,$\mathbb{R}$ は可分である.

Definition 1.7 (Tangent Space)**.** H を可分ヒルベルト空間とする. 閉集合 $A \subset \mathcal{H}$ と $a \in A$ に対し,$Tan^0(a,A)$ を以下で定める.

$$\{v \in \mathcal{H} \mid 任意の\epsilon に対し, ある b \in A が存在し,0 < |b-a| < \epsilon に対し |\frac{v}{|v|} - \frac{b-a}{|b-a|} < \epsilon.\}$$

$Tan(a,A)$ を $\{x \in H | x - a \in Tan^0(a,A)\}$ で定め,a での ***Tangent Space*** という.

Propostion 1.8. A を $\mathbb{R}^n$ の閉部分集合とする. この時

$$reach(A)^{-1} = \sup\{2|b-a|^{-2} d(b, Tan(a,A)) | a.b \in A\}$$

Definition 1.9. $V \subset \mathcal{H}$ が d 次元**アフィン空間**とは, ある $a \in \mathcal{H}$ が存在し,$\{x \in \mathcal{H} \mid x + a \in V\}$ が d 次元のベクトル空間になること.

Definition 1.10 (C^r-submfd)**.** ヒルベルト空間 $\mathcal{H}$ の閉集合 M は以下を満たす時, d 次元 C^r 級部分多様体という.

- 任意の $p \in M$ に対し,$x \in U$ となる $\mathcal{H}$ の開部分集合と C^r 級写像 $\phi : U \to \mathcal{H}$ が存在する.
- $\phi|_{U \cap M}$ は終域を像に制限すると微分同相写像である
- d 次元アフィン空間の族 $\{V_i\}_{i \in I}$ に対し,$\phi(U \cap M) = \cap_{i \in I} V_i \cap \phi(U)$ が成り立つ.

$B_{\mathcal{H}} := \{x \in \mathcal{H} \mid d(x,x) \leq 1\}$ とする. $\mathcal{G}(d,V,\tau)$ を体積 V 以下であり,*reach* が τ 以上となる $B_{\mathcal{H}}$ に含まれる d 次元 (境界のない)C^r 級多様体全体のなす集合とする.

Remark. 多様体は普通の数学書では上記のようには定義されない. $\mathcal{H}$ を固定して考えたいがために, 上記のように定義したのだと思われる.

Remark. この論文では, 後でもう一つ多様体と呼ぶものが出てくる.C^r 級部分多様体が必ずしも後で出てくる多様体とは限らないことに注意せよ.

Definition 1.11. Ω を集合とする. F を Ω 上の σ-加法族, μ_P を F 上の非負測度で,$\mu_P(\Omega) = 1$ とする. この時,(Ω, F, μ_P) を確率空間という.

Definition 1.12. $X : \Omega \to [-\infty, \infty]$ が確率変数とは任意の $a \in [-\infty, \infty]$ で,$\{x \in \Omega \mid X(x) \leq a\} \in F$ となること.

Definition 1.13. X を確率変数とした時,$P(x) = \mu_P(X < x)$ と書ける時,P を確率分布関数という.

1.3 Manifold Hypothesis

この論文で考える多様体仮説とそのために必要な用語を定義する. 可分なヒルベルト空間を $\mathcal{H}$, その単位球を $B_{\mathcal{H}}$, $B_{\mathcal{H}}$ 上の確率分布関数を P とする.

$$\mathcal{L}(M,P) := \int_{x \in B_{\mathcal{H}}} d(x,M)^2 \frac{dP(x)}{dx} dx.$$

とする. 確率分布関数 P による i.i.d(独立試行) が行われた時, 以下が成り立つことを多様体仮説と呼ぶ.

HYPOTHESIS 1. ある $M \in \mathcal{G}(d, CV, \tau/C)$ が存在し,$\mathcal{L}(M,P) \leq C\epsilon$ となる.

Remark. 論文中でこれを多様体仮説と直接呼んではいないが, 内容から判断して多様体仮説と呼ぶことにした.

Remark. この時点でもわかるように,d, C, V, τ としてどのような値を取るべきかは不明.

1.4 Main Theorem 1

この論文の 1 つめの主定理を述べる.

Main Thoerem 1. ある $r > 0$ に対し,

$$U_{\mathcal{G}(1/r)} := CV(\frac{1}{\tau^d} + \frac{1}{(\tau r)^{d/2}})$$

とする. また,

$$s_{\mathcal{G}}(\epsilon, \delta) := C(\frac{U_{\mathcal{G}}(1/\epsilon)}{\epsilon^2}(\log^4(\frac{U_{\mathcal{G}}(1/\epsilon)}{\epsilon})) + \frac{1}{\epsilon^2}\log\frac{1}{\delta})$$

とする. $s \geq s_{\mathcal{G}}(\epsilon, \delta)$ とし,$x = \{x_1, \ldots, x_s\}$ を確率分布 P による独立試行により, 得られた集合とする. P_X を X 上の一様確率分布関数とする. $M_{erm}(x) \in \mathcal{G}(d, V, \tau)$ を

$$L(M_{erm}(x), P_X) - \inf_{M \in \mathcal{G}(d,V,\tau)} L(M, P_X) < \frac{\epsilon}{2}$$

を満たす $\mathcal{G}(d, V, \tau)$ で, 以下を最小にするものとする.

$$\sum_{i=1}^{s} d(x_i, M)^2$$

この時,

$$\mathbb{P}[L(M_{erm}(x), P) - \inf_{M \in \mathcal{G}(d,V,\tau)} L(M, P) < \epsilon] > 1 - \delta$$

となる.

Remark. s 次元の直積測度で考えた時, 上の条件を満たす x 全体の集合をその測度に代入すると $1 - \delta$ より大きいという意味.

Remark. この定理が何を示しているかというと,$L(M, P)$ の極小値に十分近い値を取る, 多様体 $M_{erm}(x)$ を構成できることが言えた.$M_{erm}(x)$ は最適化問題の解として得られるので, 極小値よりもかなり具体的になっている. 特に,$M_{erm}(x)$ が確率を計算しなてくも求められるのがよい. 基本的に P_X は未知であることを前提に問題と解くので,P_X に依存しないのは強い結果である.

Remark. 仮定からわかるように,$\epsilon.\delta$ の値にも強く依存する. 私はまだ, これらとして何を取るべきかはわかっていない. 計算してみて, どのぐらい大きい s が必要か推測 *(*計算*)* したい.

2 Preparation for Main Theorem 1

主定理 1 を示すための準備を行う. 多様体を考察することで以下を示す.

Claim 1. *$M \in \mathcal{G}(d, V, \tau)$ とし,Π_x を $\mathcal{H} \to \mathrm{Tan}(x, M)$ への射影とする. 十分大きい (control された?) 定数 C に対し,*

$$U := \{y \mid |y - \Pi_x y| \leq \tau/C\} \cap \{y \mid |x - \Pi_x y| \leq \tau/C\}$$

とする. この時,$C^{1,1}$ 級関数 $F_{x,U} : \Pi_x(U) \to \Pi_x^{-1}(\Pi_x(0))$ で以下を満たすものが存在する.

1. *$F_{x,U}$ の Lipschitz constant of the gradient が C 以下である.*
2. *$M \cap U = \{y + F_{x,U}(y) \mid y \in \Pi_x(U)\}$*

2.1 Proof of Claim 1

$\mathcal{H}$ を (無限次元でもよい) ヒルベルト空間とする. D-plane を $\mathcal{H}$ の D 次元部分ベクトル空間とする. DPL で D-plane 全体のなす空間を表す. $\Pi, \Pi' \in DPL$ に対し,$T : \mathcal{H} \to \mathcal{H}$ を直交変換 (内積が不変な変換) であって,$T(\Pi) = \Pi'$ とを満たすとする.T 全体のなす集合を $A_{\Pi,\Pi'}$ とする. この時,

$$\mathrm{dist}(\Pi, \Pi') := \inf_{T \in A_{\Pi,\Pi'}} ||T - I||$$

で定める. $(DPL, dist)$ は距離空間になる.

D-palne Π に対し,

$$\Psi : B_\Pi(0, r) \to \Pi^\perp$$

を $C^{1,1}$ 級関数 (定義不明) で $\Psi(0) = 0$ とする. この時,a patch of radius r over Π centered at 0 とは,

$$\Gamma = \{x + \Psi(x) \mid x \in B_\Pi(0, r)\}$$

をさす. さらに

$$||\Gamma||_{C^{1,1}(B_\Pi(0,r))} := \sup_{x \neq y \in B_\Pi(0,r)} \frac{\nabla\Psi(x) - \nabla\Psi(y)}{||x - y||}$$

とする. ここで $\nabla\Psi(x) : \Pi \to \Pi^\perp$ は接ベクトル空間の間の射 $T_x\Pi \to T_x\Pi^\perp$ のこと (としか考えられない). ただし,$\Pi \sim T_x\Pi, \Pi^\perp \sim T_x\Pi^\perp$ を使って同一視している. 実質 x での gradient.

もし $\nabla\Psi(0) = 0$(0 写像) が成り立っている場合,Γ を a patch of radius r tangent to Π at its center 0 という.

Lemma 2.1. *Γ_1 を Π_1 上の半径 r_1 で中心 z_1 で Π_1 に接している patch とする. $z_2 \in \Gamma_1$ で $||z_2 - z_1|| < c_0 r_1$ を満たすとする.*

$$||\Gamma_1||_{C^{1,1}(B_\Pi(z_1,r_1))} \leq \frac{c_0}{r_1}$$

とする.$\Pi_2 \in DPL$ で $\mathrm{dist}(\Pi_2, \Pi_1) < c_0$ とする．この時,Π_2 上の半径 $c_1 r_1$ で中心 z_2 の *patch* Γ_2 で以下を満たすものが存在する．

$$||\Gamma_2||_{C^{1,1}(B_\Pi(0,c_1r_1))} \geq \frac{200c_0}{r_1}$$

と

$$\Gamma_2 \cap B_{\mathcal{H}}(z_2, \frac{c_1r_1}{2}) = \Gamma_1 \cap B_{\mathcal{H}}(z_2, \frac{c_1r_1}{2})$$

Proof. わからん…. 恐らく陰関数定理を使いたいのだろうが, 条件を満たす Γ の一意性が言えないような… □

Definition 2.2. *$M \subset \mathcal{H}$ が"compact imbedded D-manifold" (for short, just a "manifold") とは, 以下が成り立つことである．*

- *M が compact*
- *任意の $z \in M$ に対し, ある $T_zM \in DPL$ と,T_zM 上の半径 r_1, 中心 z,z で $T_z(M)$ に接する patchΓ が存在し, $M \cap B_{\mathcal{H}}(z, r_2) = \Gamma \cap B_{\mathcal{H}}(z, r_2)$ となる．*

We say that M has infinitesimal reach ρ if for every $\rho' < \rho$, there is a choice of $r_1 > r_2 > 0$ such that for every $z \in M$ there is a patch Γ over $T_z(M)$ of radius r_1, centered at z and tangent to $T_z(M)$ at z which has $C^{1,1}$-norm at most $1/\rho'$

Lemma 2.3 (Growing Patch)**.** *Let M be a manifold and let r_1, r_2 be as in the definition of a manifold. Suppose M has infinitesimal reach ≥ 1. Let $\Gamma \subset M$ be a patch of radius r centered at 0, over T_0M. Suppose r is less than a small enough constant $\hat{c}$ determined by D. Then there exists a patch Γ^+ of radius $r + cr_2$ over T_0M, centered at 0 such that $\Gamma \subset \Gamma^+ \subset M$*

Proof. $\mathcal{H}$ はヒルベルト空間なので,$\mathcal{H} = T_0M \oplus T_0M^\perp$ に直和分解する．M が D 次元なので,$T_0M \simeq \mathbb{R}^D$ となる．patch Γ を $\Gamma = \{(x, \Psi(x) \mid x \in B_{\mathbb{R}^D}(0,r)\}$ と書く, ここで,$C^{1,1}$ 写像 $\Psi : B_{\mathbb{R}^D}(0,r) \to \mathcal{H}^\perp$ は $\Psi(0) = 0, \nabla\Psi(0) = 0$ となり,$||\Psi||_{C^{1,1}(B_{\mathbb{R}^D}(0,r))} \leq C_0$ となる．r を十分小さくとり,$y \in B_{\mathbb{R}^D(0,r)}$ に対し, $|\nabla\Psi(y)| \leq C_0||y|| \leq C_0$ とできる. この時, 上の lemma より (本当はきちんと条件を確認したい. $\Pi_1 = \Pi_2$ でとっているのと平行移動していること, そもそもこの取り方してるならもっと弱い補題でいけるのでは?)

$$\Psi_y : B_{\mathbb{R}^D}(y, c'r_2) \to \mathcal{H}^\perp$$

で $||\nabla\Psi_y(z)||, ||\Psi_y||_{C^{1,1}}$ が有界で,

$$M \cap B_{\mathcal{H}}((y, \Psi(y)), c''r_2) = \{(z, \Psi_y(z)) \mid z \in B_{\mathbb{R}^D}(y, c'r_2)\} \cap B_{\mathcal{H}}((y, \Psi(y)), c''r_2)$$

□

第 II 部

Main theorem 1の証明

Main theorem 1 の証明の概略をのべ, 一部を実際に示す.

Remark. *1 点は論理的に間違いがあり, 証明できていない (と思う)*

3 証明のフロー

Main Theorem 1 の証明の方針を述べる.

1. JL の補題による確率に関する不等式 (4.8)
2. fat shattering dimensin による条件を満たす場合に一様確率分布の場合の期待値と実際の期待値との誤差が十分小さくなる.(4.9)
3. fat shatterig dimension を特別な確率変数 (族) の場合に評価する (4.10)
4. 確率変数 (族) として多様体 M との距離を取る. この確率変数族は上の特別な確率変数族とはならないが, 多様体に関するある条件のもとで, 上の条件を満たす確率変数 (族) から期待値の評価ができる.(4.11)
5. $\mathcal{G}$ の多様体は上記の条件を満たすことを Claim 1 を用いて示す.(4.13)
6. 上で示した, 一様確率分布の場合と実際の期待値の間に差がないことは多様体の選び方によらないことと一様確率分布の場合は距離を小さくする多様体を取ることで, 期待値を十分小さくできることから定理は示せる.

4 TOOLS FROM EMPIRICAL PROCESS

証明に使う用語を定義する.

Definition 4.1 (metric entropy)**.** 距離空間 $(Y;\rho)$ と $Z \subset Y$,$\eta \in \mathbb{R}$ に対し, 任意の $y \in Y$ に対し,$\rho(y,z) < \eta$ となる $z \in Z$ が存在する時,Y を *η-net of Y* という. 距離空間 X 上の確率測度 P に対し,$F \subset \mathrm{Map}_{cont}(X,\mathbb{R})$ は L^2-ノルムにより距離空間となる. この時 $N(\eta, F, L^2(P))$ を F の *η-net* となる有限集合 Y の位数の最小値で定め, これを ***metric entropy*** という.

Definition 4.2 (Fat-shattering dimension)**.** *Let $F \subset \mathrm{Map}_{cont}(\mathbb{R},\mathbb{R}$. We say that a set of points $(x_1,\ldots,x_k)$ is γ-shattered by F if there is $t =$*

$(t_1, \ldots, t_k) \in \mathbb{R}^k$ *such that for all* $b = (b_1, \ldots, b_k) \in \pm 1^k$*, each* $i \in [s] = \{1; \ldots, s\}$*, there is a function* $f_{b,t}$ *satisfying,*

$$b_i(f_{b,t}(xi) - t_i) > \gamma$$

Fat-shattering dimension $\mathrm{Fat}_\gamma(F)$ を γ*-shattering set* の位数の極大値で定める.

Remark. F の取り方に依存する.F として $\mathrm{Map}_{cont}(\mathbb{R}, \mathbb{R})$ を取ると常に *Fat-shattering dimension* は ∞ となる.

Definition 4.3 (VC dimension)**.** *Let* Λ *be a collection of measurable subsets of* $\mathbb{R}^m$*. For* $x_1, \ldots, x_k \in \mathbb{R}^m$*, let* $\#\{(x1, \ldots, x_k) \cap H \mid H \in \Lambda\}$ *be denoted the shatter coefficient* $N(x_1, \ldots, x_k)$*. The VC dimension* VC_Λ *is the largest integer* k *such that there exists* $(x_1, \ldots, x_k)$ *such that* $N(x_1, \ldots, x_k) = 2^k$*.*

Remark. $(x_1, \ldots, x_k)$ のうち欲しいものだけを含む H が必ず存在するのが *VC dimension.*

VC dimension には以下が成り立つ.

Theorem 4.4. *Let* Λ *be the class of halfspaces in* $\mathbb{R}^g$*. Then* $VC = g + 1$*.*

Lemma 4.5. *For any* $(x_1, \ldots, x_k) \in \mathbb{R}^g$*,* $N(x_1, \ldots, x_k) \leq \sum_{i=0}^{VC_\Lambda} \binom{k}{i}$

Johson-Lindenstartususs の補題を述べる.

Lemma 4.6 (Johnson-Lindenstrauss)**.** *Let* $\epsilon \in (0, 1/2)$*. Let* $Q \subset \mathbb{R}^d$ *be a set of* n *points and* $k = \frac{20 \log n}{\epsilon^2}$ *There exists a Lipshcitz mapping* $f : \mathbb{R}^d \to \mathbb{R}^k$ *such that for all* $u, v \in Q$*:*

$$(1 - \epsilon)||u - v||^2 \leq ||f(u) - f(v)||^2 \leq (1 + \epsilon)||u - v||^2$$

これを確率論に適用したものがある

Lemma 4.7 (JL 確率版)**.** *A related lemma is the distributional JL lemma. This lemma states that for any* $0 < \epsilon, \delta < 1/2$ *and positive integer* d*, there exists a distribution over* $\mathbb{R}^{kd}$*, from which the matrix* A *is drawn such that for* $k = O(\epsilon - 2 log(1/\delta))$ *and for any unit-length vector* $x \in \mathbb{R}^d$*, the claim below holds.*

$$P(|\|Ax\|_2^2 - 1| > \varepsilon) < \delta$$

One can obtain the JL lemma from the distributional version by setting $x = (u - v)/\|u - v\|_2 x = (u - v)/\|u - v\|_2$ and $\delta < 1/n^2 \delta < 1/n^2$ for some pair u, v both in X. Then the JL lemma follows by a union bound over all such pairs.

本論文ではその corollary として特に，以下を用いる.

Lemma 4.8 (JL)**.** *Let $y_1, \ldots, y_l$ be points in the unit ball in $\mathbb{R}^m$ for some finite m. Let $\mathbb{R}$ be an orthogonal projection onto a random g-dimensional subspace (where $g = C\frac{\log}{\gamma^2}$) for some $\gamma > 0$ and an absolute constant C). Then,*

$$\mathbb{P}[\sup_{i,j \in \{1,\ldots,g\}}]|\frac{m}{g}Ry_i \cdot Ry_j - y_i \cdot y_j| > \frac{\gamma}{2}] < \frac{1}{2}$$

JL を使い fat-shattering dimension を用いて確率の評価をしたのが以下である.

Lemma 4.9. *Let μ be a measure supported on X, $F \subset \mathrm{Map}(X, \mathbb{R})$. Let $x_1, \ldots, x_s$ be independent random variables drawn from μ and s be the uniform measure on $x := (x_1, \ldots, x_s)$. If*

$$s \geq \frac{C}{\epsilon^2}((\int_{c\epsilon}^{\infty} \sqrt{\mathrm{fat}_\gamma(F)}\mathrm{d}\gamma)^2 + \log 1/\delta)$$

then,

$$\mathbb{P}[\sup_{f \in F}|\mathbb{E}_{\mu_s}f(x_i) - \mathbb{E}_\mu f| \geq \epsilon] \leq \delta.$$

Remark. 定理の主張がおそらく間違っていたので修正した *(*論文では $1-\delta$ となっていた*)*.

Proof. Let $g = C_1(4\gamma^{-2}\log(s + kl))$ for a sufficiently large universal(意味不明) constant C_1. g 次元部分空間 $V \subset \mathrm{span}(XcupV)$ に対し，$R : \mathrm{span}(XcupV) \to V$ を自然な射影とする. この時 JL より $y_i, y_j \in \{x_1, \ldots, x_s, v_1 1, \ldots, v_{kl}\}$ を用い

$$\mathbb{P}[\sup_{i,j \in \{1,\ldots,s+kl\}}|\frac{m}{g}(Ry_i) \cdot (Ry_j) - y_i \cdot y_j| > \gamma] < \frac{1}{2}$$

となる. 余事象の確率と sup の定義から,

$$\mathbb{P}[\sup_{i,j \in \{1,\ldots,s+kl\}}|\frac{m}{g}(Ry_i) \cdot (Ry_j) - y_i \cdot y_j| \leq \gamma] \geq \frac{1}{2}$$
$$\mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r) - v_{ij} \cdot x_r| \leq \gamma] \geq \frac{1}{2}$$

となる. …中断 □

Lemma 4.10. *Let P be a probability distribution supported on $B_{\mathcal{H}}$. Let $F_{k,l} = \{f \in \mathrm{Map}(B_{\mathcal{H}}, \mathbb{R}) \mid f(x) = \max_j \min_i (v_{ij} \cdot x)\}$*

$$\mathrm{fat}_\gamma(F_{k,l}) \leq \frac{Ckl}{\gamma^2}\log^2\frac{Ckl}{\gamma^2}$$
$$\textit{If} \quad s \geq \frac{C}{\epsilon^2}(kl\log^4(kl/\epsilon^2) + \log 1/\delta),$$
$$\textit{then} \quad \mathbb{P}[\sup_{f \in F_{k,l}}|\mathbb{E}_{\mu_s}f(x_i) - \mathbb{E}_\mu f| \geq \epsilon] \leq \delta$$

Proof. fat-shattering dimension を上から bound する．$X = (x_1, \dots, x_s)(x_i \in B_{\mathcal{H}})$ を γ-shattered とする． この時, 任意の $A \subset X$ に対し,$f_A \in F_{k.l}, t = (t_1, \dots, t_s)$ が存在し，以下が成り立つ.

$$\begin{aligned} x_i \in A \text{ の時} \quad & f_A(x_i) - t_i > \gamma \\ x_i \notin A \text{ の時} \quad & f_A(x_i) - t_i < -\gamma \end{aligned}$$

今後の都合に合わせ (本当に?)$X \setminus A$ を A とする. つまり

$$\begin{aligned} x_i \in A \text{ の時} \quad & f_A(x_i) - t_i < -\gamma \\ x_i \notin A \text{ の時} \quad & f_A(x_i) - t_i > \gamma \end{aligned}$$

とする．f_A の定義から, ある $v_{ij} \in B$ を用いて, $f_A(x) = \mathrm{max}_j \mathrm{min}_i (v_{ij} \cdot x)$ とかける．これを用いると，上の条件は以下のようになる．

$$\begin{aligned} & x_r \in A \text{ の時, 任意の } j \text{ に対し, ある } i \text{ が存在し，} v_{ij} \cdot x_r < t_r - \gamma \\ & x_r \in A \text{ の時, ある } j \text{ が存在し, 任意の } i \text{ に対し，} v_{ij} \cdot x_r > t_r + \gamma \end{aligned}$$

Let $g = C_1(4\gamma^{-2}\log(s + kl))$ for a sufficiently large universal(意味不明) constant C_1. g 次元部分空間 $V \subset \mathrm{span}(X \cup V)$ に対し，$R : \mathrm{span}(X \cup V) \to V$ を自然な射影とする.

$x_r \in A$ の時，任意の j に対し，ある i が存在し，以下が成り立つ.

$$\begin{aligned} \mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r) - v_{ij} \cdot x_r \leq \gamma] & \geq \mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r)| - |v_{ij} \cdot x_r| \leq \gamma] \\ & \geq \mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r)| - t_r + \gamma \leq \gamma] \\ & = \mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r)| \leq t_r] \end{aligned}$$

$x_r \notin A$ の時，ある j が存在し，任意の i に対し，いかが成り立つ.

$$\begin{aligned} \mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r) - v_{ij} \cdot x_r \leq \gamma] & \geq \mathbb{P}[-|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r)| + |v_{ij} \cdot x_r| \leq \gamma] \\ & \geq \mathbb{P}[-|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r)| + t_r + \gamma \leq \gamma] \\ & = \mathbb{P}[|\frac{m}{g}(Rv_{ij}) \cdot (Rx_r)| \geq t_r] \end{aligned}$$

$J := \mathrm{Span}RX$ とし，$t^J J \to \mathbb{R}$ をある i が存在し,$y = Rx_i$ と書ける時，$t_j(y)$ をその最小の i で定め，それ以外の時 0 とする．$F_{J,k,l}$ をある相異なる kl 個の $w_{ij} \in J$ に対し,

$$\{z \mid \mathrm{max}_j \mathrm{min}_i \begin{pmatrix} w_{ij} \\ 1 \end{pmatrix} \cdot \begin{pmatrix} z \\ -t^J(z) \end{pmatrix} \leq 0\}$$

□

となる集合全体とする．この時以下が成り立つ

Claim 2. *Let* $y_1, \cdots, y_s \in J$*. Then, the number* $W(s, F_{J,k.l})$ *of distinct sets* $\{y_1, \cdots, y_s\} \cap \iota, \iota \in F_{J,k,l}$ *is less or equal to* $s^{O((g+2)k)}$

Remark. O をそこに使う??

Proof. VC 次元の性質 (theorem 4.4) より,$(z, -t^J(z))$ で生成される半平面の VC 次元は高々$g+2$ これより，$W(s, F_{J,1,1}) \leq \sum_{i=0}^{g+2} \binom{s}{i} \leq$ となり,(一般に $s \geq g$ ではない．$s < g$ の時は $\binom{s}{g} = 0$ とする.) $F_{J,k,l}$ の元は $F_{J,1,1}$ の元の k 個の和集合の l 個の共通部分となるので, 位数は $N = \binom{W(s, F_{J,1,1})}{k}$ とすると, $\binom{N}{l}$ 個以下となり，特に $W(s, F_{J,1,1})^{kl}$ 以下となる．

Remark. もっと強い評価ができそう．

確率が 1/2 以上なので,$RX \cap \iota$ の個数は 2^{s-1} 以上 (と書いてある.) (正しくない．RX の位数が X の位数と異なる場合は上記は成り立たない．今回は上記が成り立たない X を取るとする) この時,Claim より $2^{s-1} \geq W(s, F_{J,k,l}) \geq s^{kl(g+2)}$ となる．よって, 示された． □

s の上界を求める．$s \leq kl$ なら既に求まるため,$s \geq kl$ と仮定する．$g = C_1(\gamma^{-2}\log(s+kl))$ を代入し, 式を整理すると, 以下となる．

$$\frac{s}{\log^2(s)} \leq O(\frac{kl}{\gamma^2})$$

となる．これより以下が成り立つ

$$s \geq O(\frac{kl}{\gamma^2}\log^2(\frac{kl}{\gamma}))$$

Remark. 本当に?? 不等式評価の向きを間違えているのでは?

B_H 上の点なので, 任意の 2 点の距離は 2 以下となる. そのため，$\gamma \geq 2$ の時 fat-shattering dimension は 0 となる. これと上の lemma より,

$$s \geq \frac{C}{\epsilon^2}(\int_{c\epsilon}^{2} \frac{\sqrt{kl\log^2(kl/\gamma)}}{\gamma} d\gamma)^2 + \log 1/\delta$$

Remark. 論文に誤植あり．

ならば

$$\mathbb{P}[\sup_{f \in F}|\mathbb{E}_{\mu_s} f(x_i) - \mathbb{E}_\mu f| \geq \epsilon] \leq \delta.$$

がわかる．s の条件に出てくる不等式は $t = \log(\sqrt{kl}\gamma.)$ により置換積分すればよい．これから

$$s \geq \frac{C}{\epsilon^2}(kl\log^4(kl/\epsilon^2) + \log 1/\delta)$$

なら,

$$\mathbb{P}[\sup_{f\in F}|\mathbb{E}_{\mu_s}f(x_i)-\mathbb{E}_{\mu}f|\geq\epsilon]\leq\delta.$$

がわかる.

上記の多様体の距離に関する期待値に適用する.

Lemma 4.11. $\epsilon,\delta\in[0,1]$,$U_{\mathcal{G}}:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ とする. 任意の $M\in\mathcal{G}(d,V,\tau)$ が任意の $r>0$ に対し,$U_{\mathcal{G}}(1/r)$ 個の半径 $\frac{\sqrt{r\tau}}{16}$ の *open ball* の和集合で被覆されるとする. もし,

$$s\geq C(\frac{U_{\mathcal{G}(1/\epsilon)}}{\epsilon^2})(\log^4\frac{(U_{\mathcal{G}}(1/\epsilon)}{\epsilon}))+\frac{1}{\epsilon^2}\log\frac{1}{\delta}$$

ならば,

$$\mathbb{P}[\sup_{M\in\mathcal{G}(d,V,\tau)}|\frac{\sum_{i=1}^s d(x_i,M)^2}{s}-\mathbb{E}_P d(x,M)^2|<\epsilon]>1-\delta$$

この仮定を実際に満たすこと Main Theorem を示すことにしている. Main Thoerem の主張では $\mathcal{G}(d,V,\tau)$ で, 以下を最小にするものとしていた.

$$\sum_{i=1}^s d(x_i,M)^2$$

ただ, 論文を読む限り極限をとった議論を使っており, 最小を構成ししているようにも読めなかった. (論文の一番最後の帰結を導く, 多様体の存在は示せるが最小が存在するのかは不明)

近似を用いて Px に対して $d(x,M_{erm})$ と $d(x,M)$ の差を小さくできることと

上の仮定を満たすことは以下からわかる.

Definition 4.12. *Let* (X,d) *be a metric space, and* $r>0$*. We call that* $Y\subset X$ *is an* r*-net of* X *if for each* $x\in X$*, there is a* $y\in Y$ *such that* $d(x,y)<r$*.*

Corllary 4.13. *Let*

$$U_{\mathcal{G}}:\mathbb{R}^+\to\mathbb{R}$$

be given by

$$U_{\mathcal{G}}(1/r)=CV(\frac{1}{\tau^d}+\frac{1}{(\tau r)^{d/2}}).$$

Let $M\in\mathcal{G}$*, and* M *be equipped with the metric* $d_{\mathcal{H}}$ *of the* $\mathcal{H}$*. Then, for any* $r>0$*, there is an* $\sqrt{\tau r}$*-net of* M *consisting of no more than* $U_{\mathcal{G}}(1/r)$ *points.*

Remark. なぜ、逆数で与える。違和感しか, 感じない.

Proof. $\mathcal{G}$ はコンパクトなので、$s < r$ とした時、$B(x,s)$ を x を中心として半径 s の開球とすると、$\cup_x B(x,s)$ は有限部分被覆 $\{B(x_i.s)\}$ を持つ。この時、$Y = \{x_i\}$ は有限で、任意の $x_i \in Y$ に対し、$\text{mini}|x_i - x_j| < \epsilon$ となる. Y が M の r-net になることは上より、明らか。

これもうまく証明できていない.(おそらく有限個の全体がほとんど重複しない球による有限被覆が存在することを示し, Claim 1 から球と M の共通部分の体積が十分あることを示している.) □