第 I 部

内容

1 start

第 1 回で紹介したいもの

- 最低限の周辺知識＋記号
- 前半の主定理 (Thm1) と意味
- Claim1 の証明を含めた論文の１３章

Theorem 1.1.

第 II 部

論文そのもの

1 Introduction

多様体仮説について解説する。

参照性を高めるため、章立てはや定理番号は論文と合わせる。必要な範囲で数学的な用語、意味論をまとめて紹介したい。この論文は一言で言うならば、"In this paper, we take a "worst case" viewpoint of the Manifold Learning problem."となる。

最初にこの論文で使う記号の定義を説明する。H を可分な Hilbert Space(おそらく $\mathbb{R}$ ベクトル空間) $|\cdot|:H\to\mathbb{R}$ をヒルベルト空間のノルムとし、$d(x,y):=|x-y|$ で距離を定める。$B_H:=\{x\in H||x|\leq 1\}$ とし、P を B_H 上の確率測度とする。M を B_H の閉部分集合とし、$x\in B_H$ に対し、$d(x,M):=\inf_{y\in M}|x-y|$ とし、$\mathcal{L}(M,P):=intd(x,M)^2dP(x)$ とする。確率測度 P で i.i.d(互いに独立) に分布するとする。しかし、P は未知とする。This is a worst-case view in the sense that no prior information about the data generating mechanism is assumed to be available or used for the subsequent development

In order to state the problem more precisely, we need to describe the class of manifolds within which we will search for the existence of a manifold which satisfies the manifold hypothesis. Let M be a submanifold of H. The reach $\tau>0$ of M is the largest number such that for any $0<r<\tau$, any point at a distance r of M has a unique nearest point on M. Let $\mathcal{G}_e==\mathcal{G}_e(d;V;\tau)$ を d 次元の B_H の C^2 部分多様体であって、体積が V 以下で、

reach が τ 以上のもの。P を確率分布とする。(測度との関係は？)$(x_1, x_2, \dots)$ を iid な分布とする。(H は無限次元であってもよい。)

The test for the Manifold Hypothesis answers the following affirmatively: Given error ϵ, dimension d, volume V, reach and confidence $1-\delta$, is there an algorithm that takes a number of samples depending on these parameters and with probability $1-\delta$ distinguishes between the following two cases (as least one must hold):

- Whether there is a

  $$M \in \mathcal{G}_e = \mathcal{G}_e(d; CV; \tau/C)$$

  such that

  $$\int d(M; x)^2 dP(x) < C\epsilon$$

- Whether there is no manifold

  $$M \in \mathcal{G}_e(d, V/C; C\tau)$$

  such that

  $$\int d(M, x)^2 dP(x) < \epsilon/C$$

ただし、C は d のみ依存する。

Empirical Loss$L_{emp}(M)$ を

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

とする。***************** あとで埋める***************** M_A の定義がわからなかったので、次の subusection に移動した。

1.1 Definitions

Definition 1.1 (reach)**.** *Let M be a submanifold of H. The reach $\tau > 0$ of M is the largest number such that for any $0 < r < \tau$, any point at a distance r of M has a unique nearest point on M.*

Definition 1.2 (Tanget Space)**.** H を可分ヒルベルト空間とする。閉集合 $A \subset H$ と $a \in A$ に対し、$Tan^0(a, A)$ を $v \in H$ で、任意の ϵ に対し、ある $b \in A$ が存在し、$0 < |a-b| < \epsilon$ と、$|\frac{v}{|v|} - \frac{b-a}{|b-a|} < \epsilon.$ を表す。$Tan(a, A)$ を $\{x \in H | x - a \in Tan^0(a, A)\}$ とする。

Propostion 1.3. 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 1.4 (C^r-submfd)**.**

We assume that $\tau < 1$ and $r = 2$. ****************** あとで埋める****************** 燃え尽きました。

2 Sample complexity of manifold fitting

we show that if instead of estimating a least-square optimal manifold using the probability measure, we randomly sample sufficiently many points and then find the least square fit manifold to this data, we obtain an almost optimal manifold. つまり、確率が一番小さい多様体というものを見つけなくても、適当に十分な点を取れば、だいたい欲しい図形を取れる。最小化という作業が不要。(つまり学習のタスクが少ない) を主張したい。

Definition 2.1 (Sample Complexity)**.** $\epsilon, \delta \in \mathbb{R}$,$X$ を位相空間、F を $f : X \to \mathbb{R}$ 全体のなす集合 $s = S(\epsilon, \delta, F)$ を以下が成り立つ最小の実数とする。ある $A : X^s \to F$ が存在し、X 上の任意の確率分布に対し、$(x_1 cdots, x_x) \in X^s$ が P の $i.i.d$ な列であって．$f_{out} := A(x_1, \ldots, x_s)$ が以下を満たすとする。

$$\mathbb{P}[\mathbb{E}_{x|?P} f_{out}(x) < (\inf_{f \in F} E_{x|P} f) + \epsilon] > 1 - \delta$$

Theorem 2.2. 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、$x = \{x_1, \ldots, x_s\}$ を確率測度 P による独立試行によるエられた集合とする。P_X を X 上の一様確率測度とする。*(どの元が出る確率が等しい)* M_{erm} を $\mathcal{G}(d, V, \tau)$ の元であって、

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

を満たし、以下を最初にするものとする。

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

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。

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. $$M \cap U = \{y + F_{x,U}(y) \mid y \in \Pi_x(U)\}$$

3 Proof of CLAIM 1

3.1 Constans

D is a fixed integer. Constants $c, C, C\prime$ etc depend only on D. These symbols may denote different constants in different occurrences, but D always stays fixed.

3.2 D-planes

$\mathcal{H}$ をヒルベルト空間とする．(無限次元でもよい) 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，直交変換 (内積が不変な変換)$T : \mathcal{H} \to \mathcal{H}$ であって，$T(\Pi) = \Pi'$ とする．これ全体のなす集合を $A_{\Pi,\Pi'}$ とする (自分オリジナル)．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．(未確認)

3.3 Patches

D-palne Π に対し，

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

を $C^{1,1}$(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 $T_x\Pi\to T_x\Pi^\perp$ のこと．ただし，$\Pi\sim T_x\Pi$ を使って同一視している．

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 3.1. 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 $||z_2-z_1||<c_0r_1$ 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_1r_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… □