輪講資料 Understanding Machine Learning: From Theory to Algorithms Part V

Masanari Kimura

June 25, 2019

Abstract

本資料は書籍”Understanding Machine Learning: From Theory to Algorithms”[1] の輪講資料です．本資料は該当書籍の Chapter6 の内容を含みます．

1 無限サイズの仮説集合の学習可能性

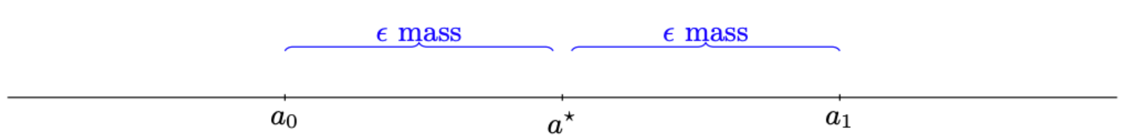
前までの章で，有限サイズの仮説集合が学習可能であることを示してきた．では仮説集合のサイズが無限である場合はどのような議論ができるだろうか？以下に，学習可能な無限サイズの仮説集合の例を示す．

例 1. $\mathcal{H}$ を実数上の閾値関数の集合とする，つまり，$\mathcal{H} = \{h_a : a \in \mathbb{R}\}$ であり閾値関数 $h_a : \mathbb{R} \to \{0, 1\}$ は $h_a(x) = \mathbb{1}_{[x<a]}$ となるような関数である．自明に $\mathcal{H}$ は無限サイズの仮説集合であるにも関わらず，以下の補題は $\mathcal{H}$ が *ERM* アルゴリズムを用いて *PAC* 学習可能であることを示す．

補題 1. 閾値関数の仮説集合 $\mathcal{H}$ は *ERM* アルゴリズムを用いることで，サンプル複雑性 $m_{\mathcal{H}}(\epsilon, \delta) \leq [\log(2/\delta)/\epsilon]$ で *PAC* 学習可能である．

Proof. a^* を仮説 $h^*(x) = \mathbb{1}_{[x<a^*]}$ が $L_{\mathcal{D}(h^*)}()h^* = 0$ を達成するような閾値とする．$a_0 < a^* < a_1$ を以下のように設定する，

$$\mathbb{P}_{x\sim\mathcal{D}_x}[x \in (a_0, a^*)] + \mathbb{P}_{x\sim\mathcal{D}_x}[x \in (a^*, a_1)] + \epsilon \tag{1}$$



学習データセット S が与えられた時,$b_0 = \max\{x : (x, 1) \in S\}$,$b_1 = \min\{x : (x, 0) \in S\}$ とする．$b_S \in (b_0, b_1)$ を仮説 h_S に対応する閾値とすると，$L_D(h_S) \leq \epsilon$ を満たすために必要な条件は $b_0 \geq a_0$ かつ $b_1 \leq a_1$ になる．

$$\mathbb{P}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}} > \epsilon] \leq \mathbb{P}_{S\sim\mathcal{D}}[b_0 < a_0 \vee b_1 > a_1], \tag{2}$$

一様バウンドを用いて，

$$\mathbb{P}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}} > \epsilon] \leq \mathbb{P}_{S\sim\mathcal{D}}[b_0 < a_0] + \mathbb{P}_{S\sim\mathcal{D}}[b_1 > a_1], \tag{3}$$

$b_0 < a_0$ となるのは S のサンプルが全て (a_0, a^*) に含まれない場合なので，

$$\mathbb{P}_{S\sim\mathcal{D}^m}[b_0 < a_0] = \mathbb{P}_{S\sim\mathcal{D}^m}[\forall(x, y) \in S, x \notin (a_0, a^*)] = (1 - \epsilon)^m \leq e^{-\epsilon m} \tag{4}$$

ここで $m > \log(2/\delta)/\epsilon$ と仮定するとこの式はたかだか $\delta/2$ になる.同様に $\mathbb{P}_{S\sim\mathcal{D}^m}[b_1 > a_1] \leq \delta/2$ も示せるため，3と組み合わせて，補題を証明できる．

□

2 VC 次元

前述のように，仮説集合の有限性は十分条件ではあるが必要条件ではない．ここでは，VC 次元と呼ばれる，仮説集合の学習可能性を正しく特徴付ける特性を紹介する．

定義 **1.** （*VC* 次元）仮説集合 $\mathcal{H}$ によって完全に分離できる集合 $C \in \mathcal{X}$ の最大のサイズを仮説集合 $\mathcal{H}$ の *VC* 次元といい，$VCDim(\mathcal{H})$ と書く．もし $\mathcal{H}$ が任意のサイズの集合を分離できる場合，$\mathcal{H}$ は無限の *VC* 次元を持つという．

定理 **1.** $\mathcal{H}$ を無限の *VC* 次元を持つ仮説集合とする．この時 $\mathcal{H}$ は *PAC* 学習不可能となる．

3 The Fundamental Theorem of PAC learning

定理 **2.**（*The Fundamental Theorem of Statistical Learning*）$\mathcal{H}$ をドメイン $\mathcal{X}$ から $\{0, 1\}$ への写像の仮説集合とし，損失関数を $0 - 1$ 損失とする．ここで，以下は等価となる：

- $\mathcal{H}$ は一様バウンドを持つ
- 任意の *ERM* アルゴリズムは $\mathcal{H}$ について *agnostic PAC* 学習可能な学習器を獲得できる
- $\mathcal{H}$ は *agnostic PAC* 学習可能である
- 任意の *ERM* アルゴリズムは $\mathcal{H}$ について *PAC* 学習可能な学習器を獲得できる

- $\mathcal{H}$ は有限の VC 次元を持つ

また，VC 次元は PAC 学習可能性のみならず，サンプル複雑性も特徴づける．

定理 **3.** （*The Fundamental Theorem of Statistical Learning - Quantitative Version*）$\mathcal{H}$ をドメイン $\mathcal{X}$ から $\{0,1\}$ への写像の仮説集合とし，損失関数を $0-1$ 損失とする．また，$VCdim(\mathcal{H}) = d < \infty$ と仮定する．ここで，以下を満たすような定数 C_1, C_2 が存在する：

- $\mathcal{H}$ は以下のサンプル複雑性で一様バウンドを満たす．

$$C_1 \frac{d + \log(1/\delta)}{\epsilon^2} \leq m_{\mathcal{H}}^{UC})(\epsilon, \delta) \leq C_2 \frac{d + \log(1/\delta)}{\epsilon^2} \tag{5}$$

- $\mathcal{H}$ は以下のサンプル複雑性で *agnostic PAC* 学習可能である．

$$C_1 \frac{d + \log(1/\delta)}{\epsilon^2} \leq m_{\mathcal{H}}(\epsilon, \delta) \leq C_2 \frac{d + \log(1/\delta)}{\epsilon^2} \tag{6}$$

- $\mathcal{H}$ は以下のサンプル複雑性で *PAC* 学習可能である．

$$C_1 \frac{d + \log(1/\delta)}{\epsilon} \leq m_{\mathcal{H}}(\epsilon, \delta) \leq C_2 \frac{d \log(1/\epsilon) + \log(1/\delta)}{\epsilon} \tag{7}$$

References

[1] shai shalev shwartz and shai ben david. *understanding machine learning: from theory to algorithms.* cambridge university press, 2014.