

協力ゲーム理論に基づく分類タスクのための特徴量重要度評価

Feature Importance Evaluation for Classification Tasks Based on Cooperative Game Theory

金丸 大樹 (Daiki Kanamaru), 本田 あおい (Aoi Honda)
九州工業大学 (Kyushu Institute of Technology)
e-mail : kanamaru.daiki739@mail.kyutech.jp

1 はじめに

近年、機械学習モデルの予測結果に対する説明可能性 (Explainability) の重要性が高まっており、特に医療、金融、自動運転などの高リスクな応用分野において、モデルの判断根拠を明らかにする手法への関心が急速に高まっている。中でも、個々の予測結果に対して各特徴量がどの程度寄与したのかを定量的に評価する「インスタンス単位の特徴量重要度評価」は、ユーザによるモデルの解釈を支援するだけでなく、バイアスの検出や信頼性の向上にも寄与するものとして注目されている。

本講演では、協力ゲーム理論の代表的概念であるシャープレイ値に基づき、分類タスクにおけるクラス感受性 (class sensitivity) を考慮した新たな特徴量の貢献度評価指標, CSPI (Class-based Shapley-Shubik Power Index) を提案する。本指標は、従来の手法では捉えにくかった「クラス予測の決定にどの特徴量が強く影響しているのか」という構造的な情報を明示的かつ定量的に捉えることを可能にする。特に本手法は多クラス分類問題におけるモデルの解釈性の向上を目的としており、回帰タスクに対する SHAP とは異なり、分類タスク固有のクラス決定構造を踏まえた特徴量貢献の定量化を実現する。

2 提案手法: Class-based Shapley-Shubik Power Index (CSPI)

提案する CSPI は、分類タスクにおける各クラスへの予測決定に対する特徴量の貢献度を定量的に評価するものである。これは SHAP をベースに用いているが、SHAP は回帰タスクに対応する手法であり、確率や回帰値の変化に基づいて特徴量の重要度を評価するのに対し、分類タスクでは「クラス決定」に貢献したかという離散的な決定が重要である。そのため、SHAP の手法に投票ゲームで用いられる Shapley-Shubik 指数の考え方を応用するものである。本手法では Shapley-Shubik で投票ゲームでプレイヤーの参画で結果が変わったかどうかを表すピポットの概念を活用した。

CSPI のアルゴリズムの枠組みは次の通り。

1) クラス決定の変化に基づく特性関数の構成

ある特徴量部分集合 S を指定し、他のインスタンスに対してターゲットインスタンスのその部分の値を代入し、再予測を行う。予測されたクラスの変化を one-hot ベクトルの差分として表現し、それを全インスタンスで平均することで、 S のクラス決定への影響をベクトル値で定義する。このベクトルを協力ゲーム理論における特性関数 $v(S)$ とみなす。

2) Shapley 値に基づく特徴量貢献度の算出

構成された特性関数 $v(S)$ に対して、Shapley 値の定義に従い、各特徴量がクラス決定に与える貢献度を算出する。これにより、どの特徴量がどのクラスに向けた予測に強く寄与しているかを定量的に評価できる。

具体的なアルゴリズムは以下の通りである。

Step 1. 対象インスタンスの選択: 任意のインスタンス $\mathbf{x}^{(\text{ref})} = \mathbf{x}^{(r)}$ をデータセット $D = \{\mathbf{x}^{(i)}\}_{i=1}^n$ から選ぶ。例えば,

$$D = \begin{pmatrix} x_1^{(1)} & \cdots & x_d^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(n)} & \cdots & x_d^{(n)} \end{pmatrix}, \quad \mathbf{x}^{(\text{ref})} = \mathbf{x}^{(r)}$$

Step 2. モデル出力の取得: 各 $\mathbf{x}^{(i)}$ についてモデル f により確率 $\mathbf{p}^{(i)} = (p_1^{(i)}, \dots, p_K^{(i)})$ を計算し, one-hot ベクトル $\mathbf{y}^{(i)}$ に変換:

$$y_k^{(i)} = \begin{cases} 1 & \text{if } k = \arg \max_j p_j^{(i)} \\ 0 & \text{otherwise} \end{cases}$$

Step 3. 部分特微量の上書きと再予測: 特微量集合 $S \subseteq \{1, \dots, d\}$ を選び, $x_{S,j}^{(i)}$ を以下のように定義する:

$$x_{S,j}^{(i)} = \begin{cases} x_j^{(\text{ref})} & j \in S, \\ x_j^{(i)} & j \notin S, \end{cases}$$

これを f に入力し再予測し, $\mathbf{y}_S^{(i)}$ を得る。

Step 4. 差分ベクトルの計算:

$$\Delta \mathbf{y}_S^{(i)} = \mathbf{y}_S^{(i)} - \mathbf{y}^{(i)}$$

(例: $\mathbf{y}^{(i)} = (0, 1, 0, \dots), \mathbf{y}_S^{(i)} = (1, 0, 0, \dots) \Rightarrow \Delta \mathbf{y}_S^{(i)} = (1, -1, 0, \dots)$)

Step 5. 差分の平均 (特性関数):

$$v(S) = \frac{1}{n} \sum_{i=1}^n \Delta \mathbf{y}_S^{(i)} = (v_1(S), \dots, v_K(S)), \quad \sum_{k=1}^K v_k(S) = 0$$

この $v(S)$ は、特微量集合 S の各クラスに対する貢献度 (成分和 0、範囲 $[-1, 1]$) を表す。

Step 6. Shapley 値による特微量ごとの貢献度 (CSPI 値) の算出: Step 5 で得た特性関数 $v(S)$ を用いて、各特微量 $j \in \{1, \dots, d\}$ の貢献度 (CSPI 値) ϕ_j を Shapley 値として次式で計算する:

$$\phi_j = \sum_{S \subset N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup \{j\}) - v(S)]$$

ここで $N = \{1, \dots, d\}$ は全特微量の集合であり、 $v(S)$ は特微量部分集合 S に対応する出力変化ベクトルの平均 (Step 5) である。CSPI 値 ϕ_j は、注目インスタンスにおいて特微量 j が正解クラスの予測に与える寄与度を表す。

3 実験

本講演では、提案手法の有効性を示すため、実際の分類タスクに用いられるデータセットを用いた具体的な事例を通して、解釈結果の特徴や有用性を紹介する。

参考文献

- [1] L.S.Shapley and M. Shubik, "A method for evaluating the distribution of power in a committee system," American Political Science Review, Vol. 48, pp. 787-792, 1954.

包除積分による LIME の拡張と画像分類モデルの高次相互作用の可視化

Extending LIME with Inclusion-Exclusion Integration and Visualizing Higher-Order Interactions in Image Classification Models

和田 健汰 (Wada Kenta)¹, 本田 あおい (Honda Aoi)¹

¹ 九州工業大学 (Kyushu Institute of Technology)

e-mail : Wada.kenta608@mail.kyutech.jp

1 概要

近年、機械学習はさまざまな分野で顕著な成功を収めている。しかしながら、精度の向上に伴うモデルの複雑化により、モデル出力の根拠を重みなどから直接解釈することが困難になってきており、これがモデルの信頼性に対する懸念を引き起こしている。特にブラックボックス型のモデルでは、その内部構造や意思決定プロセスが不透明であるため、信頼性や透明性の確保が重要な課題となっている。本研究では、任意のモデルを局所的に説明する手法である LIME と、解釈可能なモデルである包除積分モデルを組み合わせることにより、説明変数間の相互作用も考慮した局所的解釈を獲得する手法を提案する。この手法により、複雑な機械学習モデルの解釈性と信頼性が向上し、より正確かつ詳細な説明が可能となることが期待される。

2 前提知識

2.1 LIME

LIME [1] は、解釈可能なモデルを用いて、任意のブラックボックスモデルの局所的な予測を説明する手法である。ブラックボックスな予測モデルを f 、説明モデルを $g \in G$ (G は解釈可能なモデルのクラス)、 $\Omega(g)$ を説明モデルの複雑さ（例えば決定木における深さや、線形モデルにおける非ゼロ重みの数など）とする。また、 x を注目するインスタンス、 $z \in Z$ を x の摂動によって作成されたインスタンス、 $\pi_x(z)$ を x と z との近接性の尺度、 $\mathcal{L}(f, g, \pi_x)$ を π_x によって定義される局所領域において、 g が f を近似する際の不正確さの尺度とする。このとき、LIME による x の説明 $\xi(x)$ は以下の最適化問題により得られる：

$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \pi_x + \Omega(g)) \quad (1)$$

不正確さの度合い \mathcal{L} には、たとえば式 2 のように $\pi_x(z)$ で重み付けした平均二乗誤差を用いることができる。ここで z' は z のバイナリベクトル表現を表す。

$$\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z) (f(z) - g(z'))^2 \quad (2)$$

2.2 包除積分

包除積分 [2] は、古典的な加法積分を一般化した柔軟な集約フレームワークであり、要素間の高次相互作用を組み込む点に特徴がある。この拡張は、特徴の寄与が単純に加法的でなく、相乗効果や冗

長性を伴う可能性がある場合に特に有用である。

$X = \{1, 2, \dots, d\}$ を有限集合, $\mathcal{P}(X)$ をその冪集合とする. $\mu : \mathcal{P}(X) \rightarrow [0, 1]$ をファジィ測度, すなわち $A \subseteq B$ のときに常に $\mu(A) \leq \mu(B)$ が成り立つ集合関数とする. $f : X \rightarrow [0, 1]$ を X の各要素の重要度あるいは関連度を表す関数とする. さらに, f に関連する相互作用演算子として $I : [0, 1]^d \times \mathcal{P}(X) \rightarrow [0, 1]$ を定義する. この演算子は, X の各部分集合に対して, その要素の共時的な振る舞いに基づく値を割り当てるものである. I の代表的な例として, 要素間の最小演算 $I(f | A) = \bigwedge_{i \in A} x_i$ や, 代数積 $I(f | A) = \prod_{i \in A} x_i$ などが挙げられる.

f の μ に関する包除積分は, 次式で定義される:

$$\int^{IE} f d\mu := \sum_{A \in \mathcal{P}(X)} \left(\sum_{B \supseteq A} (-1)^{|B \setminus A|} I(f | B) \right) \mu(A), \quad (3)$$

また, 次のように μ の Möbius 変換 m^μ を用いて別表現も可能である:

$$\int^{IE} f d\mu = \sum_{A \in \mathcal{P}(X)} m^\mu(A) I(f | A), \quad m^\mu(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B). \quad (4)$$

3 提案手法

本研究では, LIME の説明モデルとして包除積分モデルを用いることを提案する. ブラックボックスな予測モデル f が与えられたとき, 式 5 の形の包除積分を説明モデル g に採用することで, x 付近における f の局所的近似を目指す.

$$g(z') = \int^{IE} z' d\mu := \sum_{A \in \mathcal{P}(X)} m^\mu(A) I(z' | A), \quad (5)$$

ここで, $X = \{1, 2, \dots, d\}$ は特徴量の集合, μ はファジィ測度, m^μ は μ の Möbius 変換, I は z' に作用する相互作用演算子である. $m^\mu(A)$ は部分集合 A に関連付けられた貢献度を表し, $I(z' | A)$ は与えられた z' に対する A 内の特徴量の共存的挙動を評価する. 式 4 に基づくこの定式化により, 説明モデルは特徴量間の正および負の相互作用を明示的に表現することが可能となる.

実際には, LIME の手順によって生成された摂動バイナリサンプル $(z_i, f(z_i))$ に対して, LIME と同様に式 2 に類似した重み付き最小二乗誤差を最小化することで, 説明モデル g を推定する. このアプローチにより, 説明モデル g は局所的に f に忠実であると同時に, 特徴量間の高次相互作用構造を捉えることが可能となる.

参考文献

- [1] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135 – 1144.
- [2] Honda, A., & Okazaki, Y. (2017). Theory of inclusion-exclusion integral. *Information Sciences*, **376**, 136–147.

誤差逆伝播法を用いない新たな勾配降下法構築による LSTM の予測精度向上と高速化、およびリザーコンピューティングの二重ループ学習化

Improved prediction accuracy and speed-up of LSTM by constructing a new gradient descent method without using error back propagation, and double-loop learning of reservoir computing

米田剛 (Tsuyoshi Yoneda)¹

¹ 一橋大学大学院経済学研究科 (Graduate School of Economics, Hitotsubashi University)

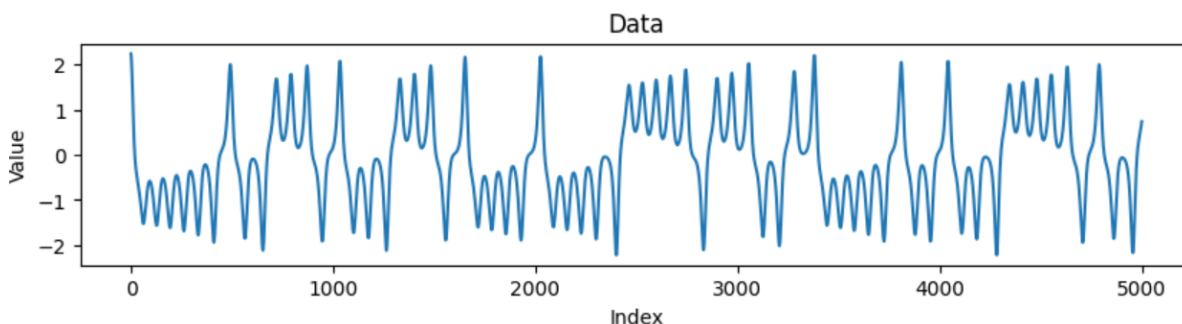
e-mail : t.yoneda@r.hit-u.ac.jp

1 概要

本講演では「現代 AI の根幹を成す誤差逆伝播法」を用いない新たな勾配降下法を構築し、新しい学習スキームとしての「二重ループ学習」を提案する。二重ループ学習とは、リザーコンピューティングにおける「出力結合重み行列に対するリッジ回帰」を踏襲しつつも、従来のハイパーパラメタも学習対象とする、というスキームである。また、誤差逆伝播法を用いない「簡易的な LSTM」も紹介し、その学習結果が（データ数 1 万未満の定常的な小規模時系列データに対して）良好であることを紹介する。この簡易版から、忘却ゲート層が「確率的なエコーステートプロパティ」を有することが理解される。

2 学習データ・学習モデルの説明

学習モデルの精度比較のために、標準化した以下の 1 次元ローレンツデータを用いる。



予測精度を比較するために固定するパラメタは以下の通りである。

- 学習可能パラメタ数 20,000 (ノード数: リザーバー ~ 150, LSTM ~ 70, GRU ~ 80)
- トレーニングデータ数 5000 (上の画像の通り)
- 予測: 100 ステップ先
- L^2 正則化パラメタ 0.0001

本講演では以下の学習モデルを比較する。

- ライブラリ TensorFlow による標準的な LSTM(勾配降下には Adam を使用)

- [3] 誤差逆伝播を用いない新たなる勾配降下による LSTM
- 誤差逆伝播を用いない新たなる勾配降下による GRU
- [3] 高速軽量 LSTM（講演時に詳述する）
- [3] 二重ループ学習のリザーコンピュティング

ここでは、TensorFlow-LSTM で使われている標準的な勾配降下法と新しい勾配降下法との違いを簡単に説明しよう。まず、全ての学習モデルにおいて、一次元データは、以下のように遅れ座標によって多次元化されたものを入力データとして用いている（遅れ座標そのものに関しては [1] を参照されたい）。

$$d_t = (data(t), data(t - \tau), data(t - 2\tau), \dots, data(t - (M - 1)\tau))$$

$data(t)$ をオリジナルの時系列データとし、 d_t が学習モデルに入力するデータの型となる。ただし、標準的な TensorFlow-LSTM とそれ以外では入出力データに対して次のような違いがある。

標準的な TensorFlow-LSTM: 入力: $\tau = 1$ かつ M は学習前に決める必要がある
 出力: スカラー値
 新たなる勾配降下: 入力: この τ, M も学習対象となる
 出力: 入力と同様の遅れ座標ベクトル

なお、「新たなる勾配降下」の学習において $\tau = 1$ が選ばれてしまった場合、 d_t と d_{t+1} がほとんど重なっている関係上、遅れ座標ベクトルのメリットが十分には発揮されない点に注意する。

新たなる勾配降下においては、従来のリザーコンピュティングのスキームと同様、リッジ回帰により、出力結合重み行列 W_{out} のみを学習し、それ以外は乱数を振る（一回目のループ）。新たなる勾配降下および二重ループ学習両方のポイントは、遅れ座標のラグ τ と遅れ次元 M を含む従来のハイパーパラメタに対してベイズ学習 (Optuna[2]) を施す点にある（二回目のループ）。

謝辞 本研究は、日本学術振興会科学研究費補助金 24H00186 によるものである。

参考文献

- [1] K. Nakai and Y. Saiki, *Machine-learning construction of a model for a macroscopic fluid variable using the delay-coordinate of a scalar observable*, Discrete Contin. Dyn. Syst. S, 14, 1079–1092 (2021).
- [2] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, *Optuna: A next generation hyperparameter optimization framework*, in Proceedings of the 25th, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery, 2019).
- [3] T. Yoneda, <https://github.com/tsuyoshi-yoneda-math/>.

半群の増大度に基づく深層学習の汎化誤差評価

Deep Generalization Analysis Based on Semigroup Growth Rate

園田 翔 (Sho Sonoda)^{1,2}, 橋本 悠香 (Yuka Hashimoto)^{3,1}, 石川 勲 (Isao Ishikawa)^{4,1}, 池田 正弘 (Masahiro Ikeda)^{5,1}

¹ 理化学研究所 (RIKEN), ² 株式会社サイバーエージェント (CyberAgent, Inc), ³ NTT 株式会社 (NTT, Inc), ⁴ 京都大 (Kyoto University), ⁵ 大阪大 (The University of Osaka)
e-mail : sho.sonoda@riken.jp

1 概要

深層学習の汎化誤差評価において、特に深さに対する依存性を精査する。古典的な解析では、深さに対して指数的に悪くなるという粗い評価しか得られず、深層学習の優位性をうまく説明できなかった。2020 年代に入って、特定のネットワークに対して多項式～対数程度の依存性が示された。本研究では、擬距離空間上で定義された一般の深層ネットワークに対し、半群の幾何学的な考察から深さに対する劣線形～対数依存性を示す。

2 定式化

多様な深層ネットワークを統一的に扱うため、可能な限り一般的な定式化を目指す。まず、入力空間はユークリッド空間に限定せず、グラフや群、関数空間など任意の擬距離空間 (\mathcal{X}, d) とする。擬距離とは、距離の公理のうち識別性「 $d(x, y) = 0 \implies x = y$ 」を落とした、一般化距離である。つまり、相異なる 2 点 $x \neq y$ の擬距離 $d(x, y)$ は 0 になることが許されている。典型例は関数 f, g の値を有限個のデータ点 x_i 上で評価して得られる経験距離 $d_n(f, g) := \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2$ である。

次に、 \mathcal{X} 上の実数値連続関数 $f : \mathcal{X} \rightarrow \mathbb{R}$ の空間を $C(\mathcal{X})$ とし、一様ノルム $\|f\|_\infty := \sup_{x \in \mathcal{X}} |f(x)|$ を入れる。 $C(\mathcal{X})$ は位相ベクトル空間である（ただし完備化は不要）。また、 \mathcal{X} の連続自己写像 $f : \mathcal{X} \rightarrow \mathcal{X}$ の空間を $C(\mathcal{X}, \mathcal{X})$ とし、一様擬距離 $d_\infty(f, g) := \sup_{x \in \mathcal{X}} d(f(x), g(x))$ を入れる。 $C(\mathcal{X}, \mathcal{X})$ は $(\mathcal{X}$ に和が定義されていないので) 一般にベクトル空間ではなく、単なる擬距離空間であることに注意されたい。

以上の準備を踏まえて、「深さ k のネットワーク」を次のような合成写像として定義する：

$$\mathcal{X} \xrightarrow{f_k} \mathcal{X} \xrightarrow{f_{k-1}} \dots \xrightarrow{f_1} \mathcal{X} \xrightarrow{h} \mathbb{R}, \quad \{f_i \mid 1 \leq i \leq k\} \subset C(\mathcal{X}, \mathcal{X}), h \in C(\mathcal{X}).$$

つまり、各中間写像 f_i と出力写像 h はそれぞれ中間層（隠れ層）と出力層に対応する。ただし具体的なネットワーク構造は特定せず、単なる連続写像で十分である。なお、慣習に従えば、上述のネットワークは「深さ $k+1$ 」と数えるが、本研究では単に深さ k と数える。

以下では、1 層の中間層のクラス $F \subset C(\mathcal{X}, \mathcal{X})$ と出力層のクラス $H \subset C(\mathcal{X})$ を固定する。

深さ k のネットワークは、より小さい深さで実現できるかもしれない。このような冗長性を排除するため、語長と語球を導入する。 $(F$ に属するとは限らない) 各要素 $g \in C(\mathcal{X}, \mathcal{X})$ に対し、 F を生成系とする g の語長 (word length) $|g|_F$ とは、 $g = f_1 \cdots f_n$ となる生成元の列 $\{f_i\}_{i \in [n]} \subset F$ の長さの最小値である： $|g|_F := \min\{n \in \mathbb{N} \mid g = f_1 \cdots f_n, f_i \in F\}$ 。ただし、そのような列が存在しないときは $|g|_F = \infty$ とする。また、 F が生成する深さ k の語球 $B(k, F)$ とは、語長 k 以下の元を全て集めた集合である：

$$B(k, F) := \{g \in C(\mathcal{X}, \mathcal{X}) \mid |g|_F \leq k\}.$$

正の数 $\varepsilon > 0$ に対し、語球 $B(k, F)$ の被覆数 N を次のように定義する:

$$N(B(k, F), d_\infty, \varepsilon) := \min\{|N| \mid N \subset C(\mathcal{X}, \mathcal{X}), \forall g \in B(k, F) \exists n \in N \text{ s.t. } d_\infty(g, n) \leq \varepsilon\}$$

また、 ε を積分消去した**周辺化増大度** $\beta(k)$ を以下で定義する:

$$\sqrt{\log \beta(k, F)} := \int_0^{\text{diam } B(k, F)} \sqrt{\log N(B(k, F), d_\infty, \varepsilon)} d\varepsilon.$$

3 主結果

主定理 (簡易版). 出力層クラス H と損失関数 L はいずれも有界かつ 1 -Lipschitz となるようにスケールされているものとする. 深さ k のネットワーク $h \in H \circ B(k, F)$ を経験誤差最小化によって学習した場合の汎化誤差は、サンプルサイズ n と深さ k に対し高確率で次のように評価される:

$$L_{\text{test}} - L_{\text{train}} \lesssim \mathfrak{R}_n(H) + \sqrt{\frac{\log \beta(k, F)}{n}} + O_p(1/\sqrt{n})$$

ただし $\mathfrak{R}_n(H)$ は出力層 H の Rademacher 複雑度である.

主定理はまず、(1) 隠れ層 F と出力層 H の影響が完全に分離できることを示している. 技術的には、 $C(\mathcal{X}, \mathcal{X})$ という一般の \mathcal{X} 値の関数族に対して Dudley 型の汎化誤差評価を与えたところが特に新しく、応用が見込める. 標準的な汎化誤差評価には Rademacher 複雑度を利用するが、これは実数値関数に対して定義されているため、一般の \mathcal{X} 値写像に対して Rademacher 複雑度を計算することはできない. そこで、Rademacher 複雑度が Dudley のエントロピー積分で評価されることを利用し、出力層の Rademacher 複雑度と隠れ層の被覆数 (位相エントロピー) による評価を導出した.

また、(2) 深さ k に対する依存性は、隠れ層の平均増大度の対数 $\log \beta(k, F)$ に依存して決まることを示している. 増大度 $\beta(k, F)$ は半群 $C(\mathcal{X}, \mathcal{X})$ の幾何学的な不変量であり、典型的には指数増大 $\beta(k, F) = \exp O(k)$ や多項式増大 $O(k^s)$ となるから、対応する汎化誤差評価は対数をとって劣線形依存 $\sqrt{\log \beta(k, F)} = O(\sqrt{k})$ や対数依存 $O(\sqrt{\log k})$ となる. 従来の深層汎化誤差解析では、Rademacher 複雑度を隠れ層の Lipschitz 係数を用いて評価し、深さに対して指数的に依存する汎化誤差評価 $\exp O(k)$ が得られていたが、このように悲観的な評価は得られにくいことを暗示している.

そして、(3) H, F はネットワークの構造を特定せずに定義されているため、深層汎化誤差の深さ依存性は半群 $C(\mathcal{X}, \mathcal{X})$ の幾何学的な不変量に由来することを示している. 従来の先行研究では、具体的なネットワーク構造を設定し、ネットワーク構造に固有の議論の末に深さ依存性が示されていたが、実は増大度のみに集中して調べればよいことが示された.

深層構造は、組み合わせ的に表現力を増すことができるので、特に表現力 (近似誤差) 評価において有利である. 主定理は、例えば、近似誤差を深さに対して指数的に減衰させつつ、推定誤差を深さに対して劣線形程度の増大に押さえるような解の存在を示唆している. 従って本定理は、深層構造の優位性 (depth-supremacy) を示す結果と言ってよいだろう.

謝辞 本研究は JSPS 科研費 24K21316, 25H01453, JST BOOST JPMJBY24E2, および JST CREST JPMJCR2015 の助成を受けたものです.

参考文献

- [1] S. Sonoda, Y. Hashimoto, I. Ishikawa, M. Ikeda, “Generalization Through Growth: Hidden Dynamics Controls Depth Dependence”, arXiv preprint: 2505.15064, (2025).