無限次元異方的平滑度を持つ関数に対す拡張畳み込みネットワークによる関数近似・推定誤差の解析

奥本 翔 (GRI 株式会社)　鈴木 大慈 (東京大学/理研 AIP)

1 背景

近年，畳み込み構造を持つニューラルネットワークが様々なタスクにおいて高いパフォーマンスを示すことが実験的に明らかになり，その理論的な研究にも関心が高まっている．それらのタスクに共通する点として，データの高次元性がある．そのため，データの次元 d がサンプルサイズ n に比べて非常に大きな場合に，次元の影響を受けずに学習できる条件が理論的に重要となる．本研究では，拡張畳み込みニューラルネットワーク (拡張 CNN) が，異方平滑性を持つ関数を，次元に依存することなく推定することが可能であることを示す．

2 研究内容

$[0,1]^\infty$-値確率変数 X_i $(i \in \mathbb{N})$ がある分布 P_X に従い，y_i は関数 $f^\circ : \mathbb{R}^\infty \to \mathbb{R}$, $\sigma > 0$ を用いて，

$$y_i = f^\circ(X_i) + \xi_i \quad (\xi_i \sim \mathrm{N}(0, \sigma^2))$$

というモデルから生成されているとする．この時，観測値 $(X_i, y_i)_{i=1}^n$ から，f° を推定する回帰問題を考える．ただし，ある集合 A に対して，$A^\infty := \{(a_1, \dots, a_i, \dots) : a_i \in A\}$ とした．また，$\psi_{l_i}(x_i) = \begin{cases} \sqrt{2}\cos(2\pi|l_i|x_i) & (l_i < 0), \\ \sqrt{2}\sin(2\pi|l_i|x_i) & (l_i > 0), \\ 1 & (l_i = 0), \end{cases}$ と $\psi_l := \prod_{i=1}^\infty \psi_{l_i}$ を用いて，$s \in \mathbb{N}_0^\infty$ に対して，$\delta_s(f)(X) := \sum_{\lceil 2^{s_i - 1} \rceil \le l_i < 2^{s_i}} \langle f, \psi_l \rangle_{L^2([0,1]^\infty)} \psi_l(X)$ で定義する．このとき，$\mathbb{N}_0^\infty$ は各要素が自然数または 0 かつ，非 0 要素数が有限の数列とする．ある $\gamma : \mathbb{N}_0^\infty \to \mathbb{R}$ に対して，γ-平滑空間は，$\|f\|_{\mathcal{F}^\gamma_{p,\theta}} := \left(\sum_{s \in \mathbb{N}_0^\infty} 2^{\theta\gamma(s)} \delta_s(f)^\theta\right)^{\frac{1}{\theta}}$ を用いて，

$$\mathcal{F}^\gamma_{p,\theta}([0,1]^\infty) := \left\{ f : [0,1]^\infty \to \mathbb{R} \mid \|f\|_{\mathcal{F}^\gamma_{p,\theta}} < \infty \right\}$$

と定義する. 各要素が正の数列 a を用いて，$\gamma(s) = \sum_{i=1}^\infty a_i s_i$, $\gamma(s) = \max\{a_i s_i\}_{i=1}^\infty$, の場合，それぞれ混合平滑，異方平滑と呼ばれる．これ以降，紙面の都合上，異方平滑の場合についてのみ議論する．ここで，a の各要素は X の各座標軸方向の滑らかさを表している．この関数空間は [1] で用いられた関数空間を一般化したものと言える．また，有限次元の場合は，これらの空間はそれぞれ混合・異方ベゾフ空間と呼ばれ，その推定精度については，[3, 2] で議論されている．このとき，次の定理が成立する：

定理 1 (多項式オーダーの下での推定誤差)**.**
$q > 1$ を用いて，$i^q \lesssim a_i$, $\gamma(s) = \max_i\{a_i s_i\}_i$ の場合，$v = \max\left\{\frac{1}{p} - \frac{1}{2}, 0\right\}$, $\tilde{a} = \sum_{i=1}^\infty \frac{1}{a_i}$, $v\tilde{a} < 1$ の下で，$f^\circ \in U(\mathcal{F}^\gamma_{p,\theta})$ $(p \ge 1,\ 1 \le \theta \le 2)$ と，ある $B_f > 0$ が存在して $\|f^\circ\|_\infty \le B_f$ が成り立つような任意の関数 f° に対して，適当な CNN の集合 $\bar{\mathcal{P}}$ の中での経験誤差最小化元 $\hat{f}$ について，

$$\mathrm{E}_{P^n}[\|\hat{f} - f^\circ\|^2_{L^2(P_X)}] \lesssim n^{-\frac{2(\frac{1}{\tilde{a}} - v)}{2(\frac{1}{\tilde{a}} - v) + 1}} (\log n)^{\frac{2}{q} + 2} \max\{(\log n)^{\frac{4}{q}}, (\log n)^4\}$$

が成立する.

この定理により，拡張 CNN を用いた異方平滑関数の推定誤差のオーダーは次元に依存せず，サンプルサイズ n と関数の異方平滑係数 a に依存することがわかる. この定理は，多くの既存研究の実験結果が示すように，拡張 CNN が超高次元データを伴う回帰問題を次元の影響を受けずに行うことができることを理論的に示している．また，同時に拡張 CNN は非線形関数回帰にも利用できることを示唆している.

参考文献

[1] Yuri Ingster and Natalia Stepanova. Estimation and detection of functions from anisotropic sobolev classes. *Electronic Journal of Statistics*, 5:484–506, 2011.

[2] T. Suzuki and A. Nitanda. Deep learning is adaptive to intrinsic dimensionality of model smoothness in anisotropic besov space. *arXiv preprint arXiv:1910.12799*, 2019.

[3] Taiji Suzuki. Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: optimal rate and curse of dimensionality. In *International Conference on Learning Representations*, 2019.