修士論文 要旨

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

数理情報学専攻 48196208 奥本 翔
指導教員 鈴木 大慈 准教授

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．また，音声認識や関数データ解析の分野では，与えられるデータが無限次元の信号とみなすことができる．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) というモデルが，異方平滑と呼ばれる滑らかさを持つ関数を，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 問題設定

2.1 回帰問題

$[0,1]^\infty$-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，

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

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．

2.2 γ-平滑空間

$$\psi_{l_i}(x_i) = \begin{cases} \sqrt{2}\cos(2\pi|l_i|x_i) & (l_i < 0), \\ \sqrt{2}\sin(2\pi|l_i|x_i) & (l_i > 0), \\ 1 & (l_i = 0), \end{cases}$$

と $\psi_l := \prod_{i=1}^\infty \psi_{l_i}$ を用いて，$s \in \mathbb{N}_0^\infty$ に対して，

$$\delta_s(f)(X) := \sum_{\lceil 2^{s_i-1}\rceil \le l_i \le 2^{s_i}} \langle f, \psi_l\rangle \psi_l(X)$$

で定義する．このとき，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．

2.3 拡張 CNN モデル

全結合 ReLU ニューラルネットワークは，$A_l \in \mathbb{R}^{d_{l+1}\times d_l}$, $b_l \in \mathbb{R}^{d_l}$, $\eta(x) = \max\{x, 0\}$ を用いて，$x \in \mathbb{R}^{d_1}$ に対して，

$$\mathrm{FNN}(x) := (A_L\eta(\cdot) + b_L)\circ\cdots\circ(A_l\eta(\cdot)+b_l)\circ\cdots\circ(A_1x+b_1)$$

で定義される．また，$C \in \mathbb{N}$, $X \in \mathbb{R}^{C\times\infty}$ が与えられたとき，$T \in \mathbb{N}$, $w \in \mathbb{R}^{C\times T}$, $h \in \mathbb{N}$ を用いて，$w \star_h X : \mathbb{R}^{C\times\infty} \to \mathbb{R}^\infty$ を，$(w\star_h X)_k = \sum_{i=1}^C \sum_{j=1}^T w_{i,j}x_{i,h(j-1)+k}$ と定義し，$w \star_h X$ を間隔 h，幅 T の拡張畳み込みと呼ぶ．$h = 1$ の場合は通常の畳み込みという．また，フィルター $F \in \mathbb{R}^{C'\times C\times T}$ が与えらえたとき，$\mathrm{Conv}_{h,F} : \mathbb{R}^{C\times\infty} \to \mathbb{R}^{C'\times\infty}$ を，

$$\mathrm{Conv}_{h,W}(X) = \begin{pmatrix} F_{1,:,:} \star_h X \\ \vdots \\ F_{C',:,:} \star_h X \end{pmatrix}$$

と定義する．L', $W' \in \mathbb{N}$, $l \in [L']$, $C_l \in \mathbb{N}$, $F_l \in \mathbb{R}^{C_{l+1},\times C_l\times W'}$，FNN をある全結合 ReLU ニューラルネットワークとして，

$$\Big(\mathrm{FNN}\circ\mathrm{Conv}_{W'^{L'-1},F_{L'}}\circ\cdots\circ \mathrm{Conv}_{W'^{l},F_l}\circ\cdots\circ\mathrm{Conv}_{1,F_1}\circ X\Big)_1$$

の形で表されるモデルを拡張 CNN と呼ぶ．ただし，FNN は無限列の各要素に適用されるものとする．また，畳み込み間隔が 1 のものを単に CNN と呼ぶ．

2.4 既存研究

有限次元説明変数を用いた場合の回帰問題の解析はすでに行われている [1, 2, 3, 4]．例えば，高次元データが与えられた場合であっても，それが低次元に埋め込まれるのであれば推定誤差が次元の影響を受けずに済むことが，いくつかの研究で示されている [1, 2]．また，$d \ll n$ の設定において，混合・異方性を持つ関数を，次

元の影響を受けずに推定することが可能であることが示されている [3, 4].

3 拡張 CNN による推定誤差

3.1 滑らかさが多項式オーダーで増大する場合

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，$f^\circ \in U(\mathcal{F}^\gamma_{p,\theta})$ $(p \geq 1,\ 1 \leq \theta \leq 2)$ と，ある $B_f > 0$ が存在して $\|f^\circ\|_\infty \leq B_f$，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，

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

が成立する.

この定理では，滑らかさがインデックスに関して多項式オーダーで増大する場合，CNN による推定誤差が次元の影響を全く受けないことを示している．推定誤差を決定しているのは，次元ではなく滑らかさに依存する量 $\tilde{a}$ であることが分かる．また，多項式オーダーで滑らかさが増大するならば，通常の CNN を用いれば十分であることが分かる.

3.2 滑らかさにスパース性がある場合

a を正の要素により構成される数列とする，$(i_1, i_2, \dots)$ を $a_{i_1} < a_{i_2} < \cdots$ となるようなインデックスとする．$0 < q < \infty$，$\|a\|_{wl^q} := \sup_j j^q a_{i_j}^{-1}$ を用いて，

$$\|a\|_{wl^q} < \infty$$

であるとき，a はスパース性を持つという．これを用いて，次の定理が成り立つ.

定理 2 (スパース性がある場合の推定誤差).

$\gamma(s) = \max_i\{a_i s_i\}$ の場合，$\|a\|_{wl^q} \leq 1$，$\log i \lesssim a_i$，$v\tilde{a} < 1$ であれば，$f^\circ \in U(\mathcal{F}^\gamma_{p,\theta})$ $(p \geq 1,\ 1 \leq \theta \leq 2)$ と，ある $B_f > 0$ が存在して $\|f^\circ\|_\infty \leq B_f$，が成り立つような任意の関数 f° に対して，適当な拡張 CNN，$\bar{\mathcal{P}}$ の中での経験誤差最小化元 $\hat{f}$ について，

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

が成立する.

滑らかさにスパース性がある場合，滑らかさを決定する列 a に長期依存性が発生する．結果として，広い範囲のインデックスを取得でき，かつパラメータ数が増大しないモデルが求められる．この定理では，拡張 CNN を用いれば，それが実現でき，多項式オーダーの場合と同様に滑らかさにより決定される量，$\tilde{a}$ により決定される推定誤差の上界が得られることを示している.

4 数値実験による検証

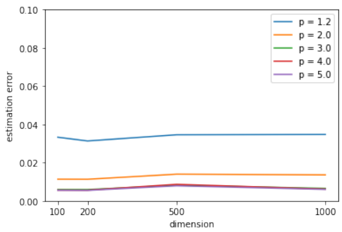

図 1. 入力の次元と深層学習による推定精度の関係. p は滑らかさ a_i の増大度合いを表すオーダー ($i^p \lesssim a_i$).

異方平滑性を持つ関数を用いて，本理論の数値的な検証を行った．図 1 より，推定誤差は，滑らかさを決定する変数 p によって決定され，次元に非依存であること確認できる.

5 結論

本研究により，拡張 CNN を用いれば，関数の滑らかさに依存し，次元に非依存であるような推定誤差を達成できることを理論・実験的に示した．これらの結果は，拡張 CNN が高次元データに有効であることを示すものである.

参考文献

[1] C. Minshuo, J. Haoming, L. Wenjing, and Z. Tuo. Efficient approximation of deep ReLU networks for functions on low dimensional manifolds. In *Advances in Neural Information Processing Systems*, volume 32, pages 8174–8184, 2019.

[2] R. Nakada and M. Imaizumi. Adaptive approximation and generalization of deep neural network with intrinsic dimensionality. *Journal of Machine Learning Research*, 21(174):1–38, 2020.

[3] T. 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.

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