Estimation of the Mean Function of Functional Data via Deep Neural Networks

Guanqun Cao

Spring School on Models and Data 2021
University of South Carolina
April 10th, 2021
Acknowledgement

Collaborators

- Mr. Shuoyang Wang, Auburn University
- Dr. Zuofeng Shang, New Jersey Institute of Technology

Partially funded by NSF award DMS-1736470

ENAR Distinguished Student Paper Award 2021
1 **Introduction**
- Functional regression model
- Deep neural networks

2 **Methods**
- Estimation of mean function via deep neural networks
- Non-asymptotic convergency rate

3 **Real data analysis**
Brain Imaging

Progression of Alzheimer’s Disease

Healthy Brain  Mild Alzheimer’s Disease  Severe Alzheimer’s Disease

Source: https://www.caring.com/caregivers/alzheimers/
Positron Emission Tomography (PET) Images
Functional regression model

\[ Y_{ij} = f_0(X_j) + \eta(X_j) + \epsilon_i(X_j), \quad i = 1, 2, \ldots, n, j = 1, 2, \ldots, N, \]

- \( f_0 : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \mathbb{E}(Y_{ij}) = f_0(X_j); \quad X_j \in \mathbb{R}^d; \)
- \( \eta(\cdot) \): individual curve variations; zero mean Gaussian process;
- \( \epsilon_i(\cdot) \): zero mean measurement error;
- \( n \): sample size;
- \( N \): number of observations for each subject.
How to estimate mean function $f_0(\cdot)$?
Deep neural networks

**Definition**

\[
f(x) = W_L \sigma(W_{L-1} \ldots \sigma(W_2 \sigma(W_1 x + v_1) + v_2) \ldots + v_{L-1}),
\]

- \( d = d_0 \rightarrow d_1 \rightarrow \ldots, \ldots \rightarrow d_L \rightarrow d_{L+1} = 1 \);
- \( \sigma(x) = \max(x, 0) \): ReLU activation function;
- \( W_\ell : p_\ell \times p_{\ell+1} \) weight matrix;

**Sparse network space:**

\[
\mathcal{F}_{DNN}(L, p, s) = \left\{ f : \max_{\ell=0,\ldots,L} \|W_\ell\|_\infty + |v_\ell|_\infty \leq 1, \sum_{\ell=0}^{L} \|W_\ell\|_0 + |v_\ell|_0 \leq s \right\}
\]
Structured compositions of Hölder Functions

- \( g_i : [a_i, b_i]^{d_i} \rightarrow [a_{i+1}, b_{i+1}]^{d_{i+1}}, \quad g_i = (g_{ij})_{j=1,\ldots,d_{i+1}} \), ambient

- Each component \( g_{ij} \) is \( \beta_i \)-Hölder function with at most \( t_i \)-variate:
  \[ \left\{ g_{ij} \in C_{t_i}^{\beta_i} \left([a_i, b_i]^{t_i}, K_i\right), \ |a_i|, |b_i| \leq K_i \right\} \], intrinsic

- True underlying function space: \( G \left(q, \{d_i, t_i, \beta_i, K_i\}_{i \in [q]}\right) \) consists of
  \( f = g_q \circ g_{q-1} \circ \cdots \circ g_1 \circ g_0 \)

- **Smoothness** of \( f_i = g_q \circ g_{q-1} \circ \cdots \circ g_i \)
  \( \beta_i^* := \beta_i \prod_{k=i+1}^{q} (\beta_k \wedge 1) \)
Empirical risk minimization

\[ \hat{f} = \arg \min_{f \in \mathcal{F}_{DNN}} \frac{1}{N} \sum_{j=1}^{N} \left\{ \overline{Y}_j - f(x_j) \right\}^2, \]

where \( \overline{Y}_j = n^{-1} \sum_{i=1}^{n} Y_{ij}, X_j = (X_{j1}, \ldots, X_{jd}) \)
Theorem 1

Under mild assumptions, with probability greater than 
\((1 - \frac{2}{nN^q})^{\log(nN^q)+1} \to 1\), we have

\[
\|\hat{f} - f_0\|^2_N \leq c(nN^q)^{-\frac{\theta}{\theta+1}} \log^6(nN^q),
\]

where \(q \geq 0\), \(\theta = \min_{i=0,...,q} \frac{2\beta_i^*}{t_i}\) and \(c\) depend on true function class of \(f_0\).

- \((nN^q)^{-\frac{\theta}{\theta+1}} = (nN^q)^{-\alpha}\) and \(\alpha = \min_{i=0,...,q} \frac{2\beta_i^*}{2\beta_i^*+t_i}\)

- If \(q = 0\), \(\|\hat{f} - f_0\|^2_N \leq cn^{-\frac{\theta}{\theta+1}} \log^6(n)\)
79 patients from the AD group.
- 33 females
- 46 males
79 patients from the AD group.
- 33 females
- 46 males

reoriented into $79 \times 95 \times 69$ voxels.

each patient has 69 sliced 2D images with $79 \times 95$. 
Recovery \((79 \times 95)\) from 3D scans

20-th

40-th

60-th
Wang, S., Cao, G. and Shang, Z. (2020) Estimation of the Mean Function of Functional Data via Deep Neural Networks. *arxiv.org*

**R Package**
https://github.com/FDASTATAUBURN/FDADNN.
Assumptions

(A1) The true regression function $f_0$ has a composition structure.

Deep and wide neural networks

(A2) $\hat{f} \in \mathcal{F}(L, p, s)$, s.t.
- Depth: $L \asymp \log(nN^\varrho)$, $\varrho \geq 0$;
Assumptions

(A1) The true regression function \( f_0 \) has a composition structure.

Deep and wide neural networks

(A2) \( \hat{f} \in \mathcal{F}(L, p, s) \), s.t.

- Depth: \( L \asymp \log(nN^q) \), \( \varrho \geq 0 \);
- Width: \( \min_{l=1,\ldots,L} p_l \asymp (nN^q)^{\frac{1}{\theta+1}} \), where \( \theta = \min_{i=0,\ldots,q} \frac{2\beta_i^*}{t_i} \).
Assumptions

(A1) The true regression function $f_0$ has a composition structure.

Deep and wide neural networks

(A2) $\hat{f} \in \mathcal{F}(L, p, s)$, s.t.

- Depth: $L \asymp \log(nN^q)$, $q \geq 0$;
- Width: $\min_{l=1,\ldots,L} p_l \asymp (nN^q)^{\frac{1}{\theta+1}}$, where $\theta = \min_{i=0,\ldots,q} \frac{2\beta_i^*}{t_i}$;
- Sparsity: $s \asymp (nN^q)^{\frac{1}{\theta+1}}$;
Assumptions

(A1) The true regression function $f_0$ has a composition structure.

Deep and wide neural networks

(A2) $\hat{f} \in \mathcal{F}(L, p, s)$, s.t.

- Depth: $L \asymp \log(nN^\varrho)$, $\varrho \geq 0$;
- Width: $\min_{l=1, \ldots, L} p_l \asymp (nN^\varrho)^{\frac{1}{\vartheta+1}}$, where $\vartheta = \min_{i=0, \ldots, q} \frac{2\beta_i^*}{t_i}$;
- Sparsity: $s \asymp (nN^\varrho)^{\frac{1}{\vartheta+1}}$. 

Guanqun Cao (Auburn University)
(A3) The maximal eigenvalue of the kernel matrix is $O(N^{-\varphi})$ for some constant $\varphi \geq 0$. 