AMCS 304: Deep Learning and Analysis

Finite Neuron Method: Programming tasks

Jinchao Xu

KAUST

jinchao.xu@kaust.edu.sa

KAUST, Nov 12th, 2024

1. L^2-fitting: Training of shallow neural networks

L^2-fitting fitting using neural network

- Consider 1D L^2-fitting problem on $\Omega = [0,1]$:

$$\min_{f_n \in \Sigma_n} \int_0^1 \frac{1}{2} |f(x) - f_n(x)|^2 dx. \tag{1}$$

- Σ_n is the spcae of ReLU shallow neural network with n neurons

$$\Sigma_n = \left\{ v(x) = \sum_{i=1}^{n} a_i \sigma(x + b_i) : a_i \in \mathbb{R}, \quad b_i \in \mathbb{R} \right\}, \quad \sigma(x) = \max(0, x).$$

- The resulting nonlinear, nonconvex optimization problem

$$\min_{a_i, b_i} \int_0^1 \frac{1}{2} |f(x) - \sum_{i=1}^{n} a_i \sigma(x + b_i)|^2 dx. \tag{2}$$

The above optimization problem (2) is usually solved by GD (or Adam).
Question

- Does the GD (or Adam) algorithm converge?
- Can we achieve the theoretical approximation rate, i.e., $\|f - f_n\|_{L^2} = O(n^{-2})$ in 1D?

Orthogonal greedy algorithm

- OGA
$$f_0 = 0, \quad g_n = \arg\max_{g \in \mathbb{D}} |\langle g, f - f_{n-1} \rangle|, \quad f_n = P_n(f), \tag{3}$$
where P_n is a projection onto $H_n = \mathrm{span}\{g_1, g_2, ..., g_n\}$
- For 1D L^2-fitting for target $f(x)$ with $f(0) = 0$, the ReLU shallow neural network dictionary can be given by
$$\mathbb{D} = \{\sigma(x + b), b \in [-1, 0]\} \tag{4}$$

Tasks

Conaider a simple target function $f(x) = sin(2\pi x)$. You may also try a function of a higher frequency, say, $f(x) = sin(10\pi x)$.

1. Solve the optimization problem using GD (or Adam). You may implement this with the help of PyTorch.
   - Record the L^2 errors for different number of neurons. Plot some numerical solutions for observation.
2. Use orthogonal greedy algorithm to train (build) a ReLU shallow neural network for fitting the target function.
   - Record the L^2 errors for different number of neurons.

The code is available on the git repo:
`https://github.com/georgexxu/ProgrammingAssignment`
Ref:

- Hong, Q., Siegel, J. W., Tan, Q., & Xu, J. (2022). On the activation function dependence of the spectral bias of neural networks. arXiv preprint arXiv:2208.04924.
- Siegel, J. W., & Xu, J. (2022). Optimal convergence rates for the orthogonal greedy algorithm. IEEE Transactions on Information Theory, 68(5), 3354-3361.
- Siegel, J. W., Hong, Q., Jin, X., Hao, W., & Xu, J. (2023). Greedy training algorithms for neural networks and applications to PDEs. Journal of Computational Physics, 484, 112084.
- Xu, J., & Xu, X. (2024). Efficient and provably convergent randomized greedy algorithms for neural network optimization. arXiv e-prints, arXiv-2407.