One-Page Summary of Zhang Shijun's Thesis

Title: Deep Neural Network Approximation via Function Compositions

Key words: Function Composition, Deep Neural Network, Approximation Theory, Floor or ReLU Activation Function, Exponential Convergence, Polynomial Approximation

Abstract: Deep neural networks have made significant impacts in many fields of computer science and engineering, especially for large-scale and high-dimensional learning problems. This thesis focuses on the approximation theory of deep neural networks. We provide (nearly optimal) approximation error estimates in terms of the width and depth when constructing ReLU networks, via the idea of function compositions, to uniformly approximate polynomials, continuous functions, and smooth functions on a hypercube. The optimality of the approximation error estimates is discussed via connecting the approximation property to VC-dimension. Finally, we introduce a new class of networks built with either Floor (the floor function) or ReLU as the activation function in each neuron, which provides a much better approximation error than that of ReLU networks.

Main results:

Table 1: A summary of the main results in the thesis, aiming to design neural networks to approximate functions in several function spaces

	target function	activation function	width	depth (#hidden-layer)	approximation error	optimality
Lemma 4.2	$f(x) = x^2$	ReLU	$3N$	L	N^{-L}	
Theorem 4.1	polynomial $f(\boldsymbol{x}) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$	ReLU	$\mathcal{O}(N)$	$\mathcal{O}(L)$	$\mathcal{O}(N^{-L})$	
Corollary 4.7	$f \in \text{Hölder}([0,1]^d, \alpha, \lambda)$	ReLU	$\mathcal{O}(N)$	$\mathcal{O}(L)$	$\mathcal{O}(\lambda N^{-2\alpha/d} L^{-2\alpha/d})$	nearly optimal in N and L, see Section 4.4.1
Theorem 4.6	$f \in C([0,1]^d)$	ReLU	$\mathcal{O}(N)$	$\mathcal{O}(L)$	$\mathcal{O}\big(\omega_f(N^{-2/d} L^{-2/d})\big)$	
Theorem 4.11	$f \in C^s([0,1]^d)$, $s \geq 1$	ReLU	$\mathcal{O}(N \ln(N+1))$	$\mathcal{O}(L \ln(L+1))$	$\mathcal{O}(\\|f\\|_{C^s} N^{-2s/d} L^{-2s/d})$	nearly optimal in N and L, see Section 4.4.2
Corollary 5.3	$f \in \text{Hölder}([0,1]^d, \alpha, \lambda)$	Floor and ReLU	$\max\{d, 5N+13\}$	$64dL+3$	$3\lambda d^{\alpha/2} N^{-\alpha\sqrt{L}}$	
Theorem 5.1	$f \in C([0,1]^d)$	Floor and ReLU	$\max\{d, 5N+13\}$	$64dL+3$	$\omega_f(\sqrt{d}\, N^{-\sqrt{L}}) + 2\omega_f(\sqrt{d}) N^{-\sqrt{L}}$	

All approximation errors in Table 1 hold for arbitrary $N, L \in \mathbb{N}^+$ and on $[0,1]^d$ uniformly. All constants in $\mathcal{O}(\cdot)$ are explicitly estimated in the thesis. $\omega_f(\cdot)$ is modulus of continuity defined by $\omega_f(r) = \sup\{|f(\boldsymbol{x}) - f(\boldsymbol{y})| : \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq r,\ \boldsymbol{x}, \boldsymbol{y} \in [0,1]^d\}$. $\text{Hölder}([0,1]^d, \alpha, \lambda)$ denotes the space of Hölder continuous functions of order $\alpha \in (0,1]$ with a Hölder constant λ on $[0,1]^d$.