第 3 回 Math-iine learning Learning Functions: When Is Deep Better Than Shallow

1/28

Contents

1 Introduction **1**

2 Previous Work **1**

3 Compositional functions **1**

4 Main results **1**

4.1 Deep and shallow neural netwokrs 2

1 Introduction

この論文では,one-hidden layer のニューラルネットワークと deep network を比較する．

2 Previous Work

3 Compositional functions

4 Main results

この章では，shallow network,deep network の 2 つの場合に近似定理を述べる．2 つとは，ReLU による deep network と deep Gaussian network である．*degree of approximation* は以下で定義される．

$$\mathrm{dist}(f, V_n) = \inf_{P \in V_n} \|f - P\| \tag{1}$$

4.1 Deep and shallow neural netwokrs

$I^d := [-1,1]^d, \mathbb{X} = C(I^d, \mathbb{R})$ とし，$\|f\| = \max_{x\in I^d}|f(x)|$ とする．S_n を n 個の *unit* を持つ *shallow netowork* のなす集合とする．すなわち，

$$S_n := \{f : \mathbb{R}^d, \mathbb{R}| \text{ある } w_k^i n\mathbb{R}^d, b_k.a_k \in \mathbb{R} \text{ が存在し}, f(x) = \sum_{k=1}^n a_k\sigma(w_k x + b_k)\}$$

この時，訓練パラメータが $(d+2)n$ 個存在する.(メタ的で数学的ではない). $W_{r,d}^{NN}$ で r 回連続偏微分可能であって, $\|f\| + \sum_{1\le|k|_1\le r}\|D^k f\| \le 1$ を満たすもの全体とする．また，$W_{H,r,2}^{NN}$ を以下で定義する．

$$W_{H,r,2}^{NN} := \{h|h = f_{11}\circ\cdots\circ f_{k2^k}(f_{ij}\in W_{r,2}^{NN})\}$$

$\mathcal{D}_n$ を S_n に属する関数の合成で書けるもの全体とする．上の書き方，かなりまずいけど，$f_11(f_21, f_22)$ で表せるもの？つまり，d が実質 2 のものということですかね．この時はパラメーターの個数が $d = 2^m$ とした時に，$(d+2)m(1+2+\cdots+2^{m-1}) = (d+2)m(d-1)$ となる．

Theorem 4.1. $\sigma : \mathbb{R}\to\mathbb{R}$ を無限回微分可能であって，$\mathbb{R}$ の任意の開区間上で，多項式でないとする．この時以下が成り立つ．

1. 任意の $f \in W_{r,d}^{NN}$ に対し，
$$\operatorname{dist}(f, S_n) = O(n^{-r/d}) \tag{2}$$

2. 任意の $f \in W_{H,r,d}^{NN}$ に対し，
$$\operatorname{dist}(f, \mathcal{D}_n) = O(n^{-r/2}) \tag{3}$$

Proof. 1 つめの主張は他の論文にて示した．2 つめの主張を示す．f が無限回微分可能な時，特にリプシッツ連続である．よって，$f(g_1, g_2) - f(P_1, P_2) \le M|g_1 - P_1||g_2 - P_2|$ となる．これより，

$$\begin{aligned}|f(g_1, g_2) - P_0(P_1, P_2)| &\le |f(g_1, g_2) - f(P_1, P_2)| + |f(P_1, P_2) - P_0(P_1, P_2)| \\ &\le M|g_1 - P_1||g_2 - P_2| + \operatorname{dist}(f, S_n)\end{aligned}$$

となる．$|g_1 - P_1||g_2 - P_2| \le O(n^{-r})$ となるので．$f(g_1, g_2) - P_0(P_1, P_2) = O(n^{-r/2})$ となる．これを inductive に続けていけばよい． □

Remark. オーダとしてはこれが限界であることが示されている．