보조정리6의 증명

신창수, 오정훈

서울대학교
통계학과

February 7, 2023

레마 1 *(Suzuki (2018) Approximation of cardinal B-spline basis by the ReLU activation) There exists a constant* $c_{(d,m)}$ *depending on* d *and* m *such that, for all* $\varepsilon > 0$*, there exists a neural network* $\check{M} \in \Phi(L_0, W_0, S_0, B_0)$ *with* $L_0 := 3+2\lceil\log_2\left(\frac{3^{d\vee m}}{\varepsilon c_{(d,m)}}\right)+5\rceil\lceil\log_2(d\vee m)\rceil,\ W_0 := 6dm(m+2)+2d,\ S_0 := L_0W_0^2$ *and* $B_0 := 2(m+1)^m$ *that satisfies*

$$||M^d_{0,0} - \check{M}||_{L^\infty(\mathbb{R}^d)} \leq \varepsilon$$

and $\check{M}(x) = 0$ *for all* $x \notin [0, m+1]^d$*.*

증명 Mhaskar and Micchelli (1992)을 참고하면 $\mathcal{N}_m(x) = \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}(x-j)^m_+$임을 알 수 있다. Yarotsky (2017)에 의해, $D \in \mathbb{N}$와 임의의 $\varepsilon > 0$이 주어지면, 다음의 조건을 갖는 신경망 $\phi_{mult} \in \Phi(L, W, S, B)$가 존재한다. $L = \lceil\log_2\left(\frac{3^D}{\varepsilon}+5\right)\rceil\lceil\log_2(D)\rceil,\ W = 6d,\ S = LW^2$ 이고 $B = 1$이면서

$$\sup_{x\in[0,1]^d}\left|\phi_{mult}(x_1,\cdots,x_D) - \prod_{i=1}^D x_i\right| \leq \varepsilon$$

그리고 $\phi_{mult}(0,\cdots,0) = 0,\ y \in \mathbb{R}^d\ s.t.\ \prod_{j=0}^d y_j = 0$를 만족한다. 또한 임의의 $M > 0$에 대하여, 함수 $\min\{M, \max\{x,0\}\}$를 단층(single-layer) 신경망 $\phi_{(0,M)}(x) := \eta(x) - \eta(x-M)$으로 정의할 수 있으므로, 위 결과에 응용하면

$$\sup_{x\in[0,M]}\left|\phi_{mult}(\phi_{(0,1)}\left(\frac{x}{M}\right),\cdots,\phi_{(0,1)}\left(\frac{x}{M}\right) - \left\{\phi_{(0,1)}\left(\frac{x}{M}\right)\right\}^m\right|$$

가 성립함을 알 수 있다. 이제 $\mathcal{N}_m(x) = 0, \quad \forall x \notin [0, m+1]$ 이므로 다음을 확인하면 $(j = 0, 1\cdots, m+1)$

$$\left\{\phi_{(0,m+1-j)}(x-j)\right\}^m = \begin{cases} 0, & x < j \\ (x-j)^m, & j \leq x < m+1 \\ 0, & x \geq m+1 \end{cases}$$

$(x-j)_+^m \equiv \left\{\phi_{(0,m+1-j)}(x-j)\right\}^m, \quad j=0,1,\cdots,m+1$ 임을 알 수 있으므로, 따라서

$$\begin{aligned}
\mathcal{N}_m(x) &= \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}(x-j)_+^m \\
&= \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}\left\{\phi_{(0,m+1-j)}(x-j)\right\}^m \\
&= \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}\left\{(m+1)\cdot\phi_{(0,1-j/(m+1))}\left(\frac{x-j}{m+1}\right)\right\}^m
\end{aligned}$$

함수 하나를 아래와 같이 정의하자. $0 \le x \le m+1$ 사이에서

$$f(x) = \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}(m+1)^m\phi_{mult}\Bigg(\underbrace{\phi_{(0,1-\frac{j}{m+1})}\left(\frac{x-j}{m+1}\right),\cdots,\phi_{(0,1-\frac{j}{m+1})}\left(\frac{x-j}{m+1}\right)}_{m\text{-개}}\Bigg)$$

정의상 $f(x)=0, \quad \forall x \le 0$ 이고, 결과들을 종합해보면

$$\begin{aligned}
&\sup_{0\le x\le m+1}|\mathcal{N}_m(x)-f(x)| \\
=\ &\sup_{0\le x\le m+1}\Bigg|\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}(m+1)^m \\
&\cdot\Bigg[\underbrace{\phi_{mult}\left(\phi_{(0,1-\frac{j}{m+1})}\left(\frac{x-j}{m+1}\right),\cdots,\phi_{(0,1-\frac{j}{m+1})}\left(\frac{x-j}{m+1}\right)\right)-\left\{\phi_{(0,1-j/(m+1))}\left(\frac{x-j}{m+1}\right)\right\}^m}_{\le\varepsilon}\Bigg]\Bigg| \\
\le\ &\frac{1}{m!}(m+1)^m\underbrace{\sum_{j=0}^{m+1}\binom{m+1}{j}}_{=2^{m+1}}\cdot\varepsilon \\
\le\ &\frac{(m+1)^m}{\sqrt{2\pi}m^{m+\frac{1}{2}}e^{-m}}\cdot 2^{m+1}\varepsilon = \underbrace{\frac{2}{\sqrt{2\pi}}\left(1+\frac{1}{m}\right)^m}_{\le 1\cdot e}m^{-1/2}(2e)^m\varepsilon \\
\le\ &e\frac{(2e)^m}{\sqrt{m}}\varepsilon \overset{let}{=} \varepsilon'
\end{aligned}$$

의 부등식이 충분히 큰 모든 m에 대해 성립한다. 두 번째 부등식에서는 스털링 공식 $m! \ge \sqrt{2\pi}m^{m+\frac{1}{2}}e^{-m}$ 을 사용하였다. 마지막으로 $m+1$보다 큰 x에 대해서는 함수 f의 값을 아래와 같이 잡자.

$$\begin{aligned}
f(x) &= \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}(m+1)^m\phi_{mult}\left(\phi_{(0,1-\frac{j}{m+1})}\left(\frac{m+1-j}{m+1}\right),\cdots,\phi_{(0,1-\frac{j}{m+1})}\left(\frac{m+1-j}{m+1}\right)\right) \\
&\overset{let}{=} \delta', \quad \forall x > m+1
\end{aligned}$$

그러면, $|\delta'| \le \varepsilon'$임을 알 수 있는데, 왜냐하면 $x \ge m+1$일때

$$\mathcal{N}_m(x) = \frac{1}{m!}\sum_{j=0}^{m}(-1)^j\binom{m}{j}(x-j)^m = \frac{1}{m!}\sum_{j=0}^{m}(-1)^j\binom{m}{j}(x-j)^j(x-j)^{m-j} = 0$$

이므로, 위 부등식 $\sup\limits_{0\le x\le m+1} |\mathcal{N}_m(x) - f(x)| \le \varepsilon'$에 의하여 $x = m+1$이면 $|\mathcal{N}_m(x) - f(x)| = |f(m+1)| = |\delta'| \le \varepsilon'$. 추가로, $0 \le \mathcal{N}_m(x) \le 1$도 확인할 수 있다.(더 생각해봐야.) 위에서 정의된 f로부터 함수 하나를 잡는다. $g(x) := \phi_{(0,1)}\left(f(x) - \frac{\delta'}{m+1}\phi_{(0,m+1)}(x)\right)$으로 놓으면 $\sup\limits_{x\in\mathbb{R}} |g(x)| \le 1$ 이고 $g(x) = 0, \quad \forall x \notin [0, m+1]$ 임과

$$\sup_{x\in\mathbb{R}} |\mathcal{N}_m(x) - g(x)| \le 2\varepsilon'$$

가 성립한다. 먼저 g에 대한 사실부터 살펴보도록 한다. $x < 0$일때

$$g(x) = \phi_{(0,1)}\left(-\frac{\delta'}{m+1}\phi_{(0,m+1)}(x)\right) = \phi_{(0,1)}(0) = 0, \quad \forall x < 0$$

$0 \le x \le m+1$일때는

$$\begin{aligned} g(x) &= \phi_{(0,1)}\left(f(x) - \frac{\delta'}{m+1}\phi_{(0,m+1)}(x)\right), \quad 0 \le \forall x \le m+1 \\ &= \eta\left(f(x) - \frac{\delta'}{m+1}\phi_{(0,m+1)}(x)\right) - \eta\left(f(x) - \frac{\delta'}{m+1}\phi_{(0,m+1)}(x) - 1\right) \\ &= \begin{cases} 1 \\ f(x) - \frac{\delta'}{m+1}\phi_{(0,m+1)}(x) \\ 0 \end{cases} \end{aligned}$$

세 번째 등식에서는 ReLU인 η 내부의 값이 각각 양수, 음수인 경우로 나누어 생각했을 때 g가 가질수 있는 값들이다. 두 번째 값은 $|f(x) - \frac{\delta'}{m+1}\phi_{(0,m+1)}(x)| \le 1$일 때 가능하다. 이제 $x > m+1$일 때를 생각해보면

$$g(x) = \phi_{(0,1)}(\delta' - \delta') = \phi_{(0,1)}(0) = 0, \quad \forall x > m+1$$

따라서 $\sup\limits_{x\in\mathbb{R}} |g(x)| \le 1$이고 $g(x) = 0, \quad \forall x \notin [0, m+1]$임을 보였다. 다음으로 $|\mathcal{N}_m(x) - g(x)|$가 $2\varepsilon'$에 유계임을 보이자. $x \le 0$또는 $x \ge m+1$일 때에는 $\mathcal{N}_m(x) = g(x) = 0$이기 때문에 $0 < x < m+1$일 때 발생하는 경우의 상계만 찾으면 된다.

$$\begin{aligned} |\mathcal{N}_m(x) - g(x)| &= \left|\mathcal{N}_m(x) - f(x)\frac{\delta'}{m+1}\phi_{(0,m+1)}(x)\right|, \quad \forall x \in (0, m+1) \\ &\le \varepsilon' + \delta' \\ &\le 2\varepsilon' \end{aligned}$$

그러므로 $\sup\limits_{x\in\mathbb{R}} |\mathcal{N}_m(x) - g(x)| \le 2\varepsilon'$임을 보였다. 앞에서 $M^d_{0,0}(x) = \prod_{j=1}^{d} \mathcal{N}_m(x_j), \ x = (x_1, \cdots, x_d)$으로

정의하였음을 상기하자. 다시 $g(x_j)$에 대하여 Yarotsky (2017)의 ϕ_{mult} 부등식을 적용하면

$$\begin{aligned}
& \sup_{x\in[0,1]^d} |M^d_{0,0}(x) - \phi_{mult}(g(x_1),\cdots,g(x_d))| \\
\le\ & \sup_{x\in[0,1]^d} \left|\prod_{j=1}^{d}\mathcal{N}_m(x_j) - \prod_{j=1}^{d} g(x_j)\right| + \sup_{x\in[0,1]^d}\left|\prod_{j=1}^{d} g(x_j) - \phi_{mult}(g(x_1),\cdots,g(x_d))\right| \\
\le\ & \sup_{x\in[0,1]^d} \sum_{j=1}^{d} |\mathcal{N}_m(x_j) - g(x_j)| + \varepsilon \\
\le\ & 2d\varepsilon' + \varepsilon
\end{aligned}$$

임을 알 수 있다. 두 번째 부등식으로 넘어가는데에는 수학적 귀납법을 적용하면 된다. 즉

$$\begin{aligned}
|\mathcal{N}_m(x_1)\mathcal{N}_m(x_2) - g(x_1)g(x_2)| &= |\mathcal{N}_m(x_2)(\mathcal{N}_m(x_1) - g(x_1)) + g(x_1)(\mathcal{N}_m(x_2) - g(x_2))| \\
&\le |\mathcal{N}_m(x_2)|\cdot|\mathcal{N}_m(x_1) - g(x_1)| + |g(x_1)|\cdot|\mathcal{N}_m(x_2) - g(x_2)| \\
&\le |\mathcal{N}_m(x_1) - g(x_1)| + |\mathcal{N}_m(x_2) - g(x_2)|, \quad \because |\mathcal{N}_m(x)| \le 1,\ |g(x)| \le 1
\end{aligned}$$

임을 알 수 있고, d에 대하여 $|\prod_{j=1}^{d}\mathcal{N}_m(x_j) - \prod_{j=1}^{d} g(x_j)| \le \sum_{j=1}^{d}|\mathcal{N}_m(x_j) - g(x_j)|$가 참이라고 하자. $A \overset{let}{=} \prod_{j=1}^{d}\mathcal{N}_m(x_j),\ B \overset{let}{=} \prod_{j=1}^{d} g(x_j)$로 놓으면 $d+1$일 때는

$$\begin{aligned}
\left|\prod_{j=1}^{d+1}\mathcal{N}_m(x_j) - \prod_{j=1}^{d+1} g(x_j)\right| &= |A\cdot\mathcal{N}_m(x_{d+1}) - \mathcal{N}_m(x_{d+1})\cdot B + B\cdot\mathcal{N}_m(x_{d+1}) - B\cdot g(x_{d+1})| \\
&\le |\mathcal{N}_m(x_{d+1})|\cdot|A-B| + |B|\cdot|\mathcal{N}_m(x_{d+1}) - g(x_{d+1})| \\
&\le |A-B| + |\mathcal{N}_m(x_{d+1}) - g(x_{d+1})|, \quad \because |\mathcal{N}_m(x_{d+1})| \le 1,\ |B| \le 1
\end{aligned}$$

에 의해 성립.

이제 $G \coloneqq \phi_{mult}(g(x_1),\cdots,g(x_d))$에 $\phi_{(0,1)}$을 적용한 함수를 h라 하면, h는
$h = \phi_{(0,1)}(G)) = min(1, max(0,G)) = G,\ when\ 0 \le G \le 1$ 이다. 따라서, 위의 범위 내에서, h는 G와 같다. 따라서,

$$\begin{aligned}
\|M^d_{0,0} - h\|_{L^\infty(\mathbb{R}^d)} = \sup_{x\in\mathbb{R}^d}\|M^d_{0,0} - h\| = \sup_{x\in[0,m+1]^d}\|M^d_{0,0} - h\| &= \sup_{x\in[0,m+1]^d}\|M^d_{0,0} - G\| \\
&\le 2d\epsilon^{'} + \epsilon \\
&\therefore \|M^d_{0,0} - h\|_{L^\infty(\mathbb{R}^d)} \le 2d\epsilon^{'} + \epsilon \doteq \epsilon^{''}
\end{aligned}$$

맨 처음에 lemma1에 있는 $\check{M}$이 h가 되는 것이고, 우리는 $M^d_{0,0}$의 B-spline을 $\check{M}$ 이라는 신경망으로 근사시킨 것이다. L,W,S,B에 대한 조건은 h 신경망을 잘 살펴보면 알 수 있다.

For the order $m \in \mathbb{N}$ of the cardinal B-spline bases, let $J(k) = \left\{-m, -m+1, \cdots, 2^k - 1, 2^k\right\}^d$ and the quasi-norm of the coefficient $(\alpha_{k,j})_{k,j}$ for $k \in \mathbb{N}_+$ and $j \in J(k)$ be

$$\|(\alpha_{k,j})_{k,j}\|_{b^s_{p,q}} = \left\{ \sum_{k \in \mathbb{N}_+} \left\{ 2^{k(s-d/p)} \left\{ \sum_{j \in J(k)} |(\alpha_{k,j})|^p \right\}^{1/p} \right\}^q \right\}^{1/q} \quad -(1)$$

레마 2 *(Suzuki (2018) Under the $s > d(1/p - 1/r)_+$ condition and the condition $0 < s < min(m, m - 1 + 1/p)$ where $m \in \mathbb{N}$ is the order of the cardinal B-spline bases, for any $f \in B^s_{p,q}(\Omega)$, there exists f_N that satisfies*

$$\|f - f_N\|_{L^r(\Omega)} \lessapprox N^{-s/d}\|f\|_{B^s_{p,q}}$$

for $N \gg 1$, and has the following form:

$$f_N(x) = \sum_{k=0}^{K} \sum_{j \in J(k)} \alpha_{k,j} M^d_{k,j}(x) + \sum_{k=K+1}^{K^*} \sum_{i=1}^{n_k} \alpha_{k,j_i} M^d_{k,j_i}(x),$$

where $(j_i)_{i=1}^{n_k} \subset J(k), K = \lceil C_1 log(N)/d \rceil, K^* = \lceil log(\lambda N)\nu^{-1} \rceil + K + 1, n_k = \lceil \lambda N 2^{-\nu(k-K)} \rceil (k = K + 1, \cdots, K^*)$ for $\delta = d(1/p - 1/r)_+$ and $\nu = (s - \delta)/2\delta$ and the real number constants $C_1 > 0$ and $\lambda > 0$ are chosen to satisfy $\sum_{k=1}^{K}(2^k + m)^d + \sum_{k=K+1}^{K^*} n_k \leq N$ independently to N. Moreover, we can choose the coefficients $(\alpha_{k,j})$ to satisfy

$$\|(\alpha_{k,j})_{k,j}\|_{b^s_{p,q}} \lessapprox \|f\|_{B^s_{p,q}}$$

증명 먼저, $P_k(f)(x) = \sum_{j \in J(k)} \alpha_{k,j} M^d_{k,j}(x)$ 라는 것을 정의할건데, $J(k)$는 B-spline $N_{j,k}$가 $\Omega \in [0,1]^d$ 에서 사라지지 않는 index(do not vanish identically) j 들의 집합이고, $M^d_{k,j}(x) = \prod_{i=1}^{d} N_m(2^k x_i - j_i)$ 이다. $P_k(f)(x)$는 DeVore and Popov (1988)에 따르면 projection 의 의미를 갖는데, B-spline인 $N_{j,k}$로 span된 공간에 projection 하는 것이다. B-spline $N_{j,k}$는 partiton of unity이므로, 전체 space로 확장시킬 수 있음이 알려져 있다.

다음으로, $p_k(f) := P_k(f) - P_{k-1}(f), P_{-1}(f) = 0$이라 하자. $P_k(f)$가 projector이므로, 이 식은 projection 후 남은 직교공간으로 이해할 수 있다. DeVore and Popov (1988)에 따르면, $\|(p_k(f))_{k=0}^{\infty}\|_{b^s_p(L^p)} < \infty$의 수렴 조건하에서, '$f$가 Besov space에 속한다' 와 '$f$가 $f = \sum_{k=0}^{\infty} p_k(f)$로 decompose된다' 가 동치임이 알려져 있다. 즉, Besov space에 속한다면, 함수를 위에서의 $P_k(f)$처럼 정의할 수 있다. 또한, DeVore and Popov (1988)에 따르면, 만약 $f_{k k=0}^{\infty}$의 sequence가 어떤 공간 L_p의 component function이면, quasi norm의 형태로 $\|\{f_k\}\|_{b^{\beta}_{\theta}(L^p)} := \left(\sum_{k=0}^{\infty} \left\{ 2^{\beta k} \|f_k\|_p \right\}^{\theta} \right)^{1/\theta}$ 로 정의할 수 있다. 또한, Dũng (2011)에 의하면, $\|f\|_{B^2_{p,q}} \simeq \|\{(p_k(f))_{k=0}^{\infty}\}\|_{b^s_q(L^p)} := \left(\sum_{k \in \mathbb{N}_+} \left\{ 2^{sk} \|p_k\|_p \right\}^q \right)^{1/q}$ 이 성립한다. 그리고, 여기서 $P_k(f)(x)$가 선형결합이므로,

그의 차인 $p_k(f)$도 선형결합으로 다음과 같이 표현할 수 있다.

$$p_k(f) = \sum_{j \in J(k)} \alpha_{k,j} M^d_{k,j}(x),\ \text{여기서 } \alpha_{k,j}\text{는 위에서와의 값과는 다름}$$

따라서, 이를 $f = \sum_{k=0}^{\infty} p_k(f)$에 대입하면, L_p의 관점에서 수렴할 때, $f = \sum_{k=0}^{\infty} \sum_{j \in J(k)} \alpha_{k,j} M^d_{k,j}(x)$로 decompose 됨을 알 수 있다. 또한, DeVore and Popov (1988)에 의하면, $\|p_k(f)\|_{L^p} \simeq (2^{-kd} \sum_{j \in J(k)} |\alpha_{k,j}|^p)^{1/p}$임이 알려져 있다. 이를 통해, $\|f\|_{B^s_{p,q}} \simeq \|(\alpha_{k,j})_{k,j}\|_{b^s_{p,q}}$ 임을 보일 수 있는데, 왜냐하면 $\left(\sum_{k=0}^{\infty} \left\{ 2^{sk} \|p_k\|_{L_p} \right\}^q \right)^{1/q} \simeq \|f\|_{B^s_{p,q}}$ 로 부터, $\|p_k(f)\|_{L^p} \simeq (2^{-kd} \sum_{j \in J(k)} |\alpha_{k,j}|^p)^{1/p}$를 여기에 대입하면,

$$\begin{aligned}
\|f\|_{B^s_{p,q}} &\simeq \left(\sum_{k=0}^{\infty} \left(2^{sk} (2^{-kd} \sum_{j \in J(k)} |\alpha_{k,j}|^p)^{1/p} \right)^q \right)^{1/q} \\
&= \left(\sum_{k \in \mathbb{N}_+} \left(2^{k(s-d/p)} (\sum_{j \in J(k)} |\alpha_{k,j}|^p)^{1/p} \right)^q \right)^{1/q} \\
&= \|(\alpha_{k,j})_{k,j}\|_{b^s_{p,q}} (\because by (1))
\end{aligned}$$

따라서, $\|(\alpha_{k,j})_{k,j}\|_{b^s_{p,q}} \simeq \|f\|_{B^s_{p,q}}$ 임을 알 수 있다. 이제 f_N을 정의할건데, Dũng (2011)의 (3.8)식을 Suzuki (2018)에서 위의 조건들과 Dũng (2011)내의 여러 조건들에 맞게끔 $f_N(x) = \sum_{k=0}^{K} \sum_{j \in J(k)} \alpha_{k,j} M^d_{k,j}(x) + \sum_{k=K+1}^{K^*} \sum_{i=1}^{n_k} \alpha_{k,j_i} M^d_{k,j_i}(x)$ 의 형태로 변형했고, Dũng (2011)의 Thm3,1,5.4에 의해, $\|f - f_N\|_{L^r(\Omega)} \lessapprox N^{-s/d} \|f\|_{B^s_{p,q}}$ 임을 알 수 있다.

성질 1 *(Approximation ability for Besov space) Suppose that* $0 < p, q, r \leq \infty$ *and* $0 < s < \infty$ *satisfy the following condition:*

$$s > d(1/p - 1/r)_+$$

Assume that $m \in \mathbb{N}$ *satisfies* $0 < s < min(m, m - 1 + 1/p)$. *Let* $\nu = (s - \delta)/(2\delta)$. *For sufficiently large* $N \in \mathbb{N}$ *and* $\epsilon = N^{-s/d - (\nu^{-1} + d^{-1})(d/p - s)_+} \log(N)^{-1}$, *let*

$$L = 3 + 2\lceil \log_2 \left(\frac{3^{d \vee m}}{\epsilon c_{(d,m)}} \right) + 5 \rceil \lceil \log_2 (d \vee m) \rceil, \qquad W = N W_0,$$
$$S = (L-1) W_0^2 N + N, \qquad B = O(N^{(\nu^{-1} + d^{-1})(1 \vee (d/p - s)_+)}),$$

then it holds that

$$R_r(\Phi(L, W, S, B),\ B^s_{p,q}([0,1]^d)) \simeq N^{-s/d}$$

증명 레마 1과 레마 2에서 보인 B-스플라인 기저의 유한선형결합의 인덱스 집합을 E_N이라 나타내자. 그리고 레마 1에서 B-스플라인으로 근사된 신경망을 $\check{M}_{k,j}$로 쓰자. 이로부터 $f_N = \sum_{(k,j)\in E_N} \alpha_{k,j} M_{k,j}^d$와 $\check{f} := \sum_{(k,j)\in E_N} \alpha_{k,j}\check{M}_{k,j}^d$ 로 쓰자. 정의역의 값 $x \in \mathbb{R}^d$가 주어지면,

$$
\begin{aligned}
|f_N(x) - \check{f}(x)| &\leq \sum_{(k,j)\in E_N} |\alpha_{k,j}||M_{k,j}^d(x) - \check{M}_{k,j}^d(x)| \\
&\leq \epsilon \sum_{(k,j)\in E_N} |\alpha_{k,j}| I\{M_{k,j}^d(x) \neq 0\} \\
&\leq \epsilon(m+1)^d(1+K^*)2^{K^*(d/p-s)_+}||f||_{B_{p,q}^s} \\
&\lesssim \log(N)N^{(\nu^{-1}+d^{-1})(d/p-s)_+}\epsilon||f||_{B_{p,q}^s}
\end{aligned}
$$

K^*는 레마2에서 잡아온 조건의 값이다. 첫번째 부등식은 삼각부등식에 의해서, 두번째 부등식은 레마 1의 근사된 신경망 모형의 결과로부터 유도되었다. 세번째 부등식은 인덱스 (k,j)가 각각 $0 \leq k \leq K^*$이고 $j \in \{1, 2, \cdots, m+1\}$인 것과 레마 2에서 성립된 조건

$$||(\alpha_{k,j})_{k,j}||_{b_{p,q}^s} \lesssim ||f||_{B_{p,q}^s}$$

에 의해서

$$
\begin{aligned}
& \sum_{k=0}^{K^*}\left[2^{k(s-d/p)}\left(\sum_{j\in J(k)}|\alpha_{k,j}|^p\right)^{1/p}\right]^q \leq ||f||_{B_{p,q}^s} \\
\Longrightarrow \quad & 2^{k(s-d/p)q}\left(\sum_{j\in J(k)}|\alpha_{k,j}|^p\right)^{q/p} \leq ||f||_{B_{p,q}^s}^q \text{ for all } 0 \leq k \leq K^* \\
\Longrightarrow \quad & |\alpha_{k,j}|^{p\cdot 1/p} \leq \left(\sum_{j\in J(k)}|\alpha_{k,j}|^p\right)^{1/p} \leq ||f||_{B_{p,q}^s}\cdot 2^{k(d/p-s)} \text{ for all } 0 \leq k \leq K^* \\
\Longrightarrow \quad & |\alpha_{k,j}| \leq 2^{K^*(d/p-s)_+}
\end{aligned}
$$

을 유도함으로써 성립한다. 마지막으로 네번째 부등식은

$$
\begin{aligned}
K^* &= \lceil \log(\lambda N)\nu^{-1}\rceil + \lceil C_1\log(N)d^{-1}\rceil + 1 \\
&\lesssim \log(N)\cdot(\nu^{-1}+d^{-1})
\end{aligned}
$$

그리고

$$2^{K^*(d/p-s)_+} \lesssim 2^{\log(N)(\nu^{-1}+d^{-1}(d/p-s)_+)} \Longrightarrow 2^{K^*(d/p-s)_+} \lesssim N^{(\nu^{-1}+d^{-1})(d/p-s)_+}$$

를 확인함으로써 알 수 있다. 이제 $\epsilon \leq N^{-s/d}/\left(\log(N)N^{(\nu^{-1}+d^{-1})(d/p-s)_+}\right)$ 보다 작게 잡으면 $f \in \mathcal{U}\left(B_{p,q}^s([0,1]^d)\right)$

가 주어졌을 때

$$\begin{aligned} ||f - \check{f}||_{L^r} &\leq ||f - f_N||_{L^r} + ||f_N - \check{f}||_{L^r} \\ &\lesssim \log(N) N^{(\nu^{-1}+d^{-1})(d/p-s)_+} ||f||_{B^s_{p,q}} \epsilon + N^{-s/d} \\ &\lesssim N^{-s/d} \end{aligned}$$

가 자명하게 보인다.

참고문헌

DeVore, R. A. and Popov, V. A. (1988). Interpolation of besov spaces. *Transactions of the American Mathematical Society*, 305(1):397–414.

Dũng, D. (2011). Optimal adaptive sampling recovery. *Advances in Computational Mathematics*, 34(1):1–41.

Mhaskar, H. N. and Micchelli, C. A. (1992). Approximation by superposition of sigmoidal and radial basis functions. *Advances in Applied mathematics*, 13(3):350–373.

Suzuki, T. (2018). Adaptivity of deep relu network for learning in besov and mixed smooth besov spaces: optimal rate and curse of dimensionality. *arXiv preprint arXiv:1810.08033*.

Yarotsky, D. (2017). Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114.