Week 3 Assignment

$\boxed{\text{Lemma 3.1}}$

3.1. Univariate polynomials

We first describe how to approximate univariate polynomials of any degree with tanh neural networks. We introduce the pth order central finite difference operator δ_h^p for any $f \in C^{p+2}([a,b])$ for some $p \in \mathbb{N}$ by

$$\delta_h^p[f](x) = \sum_{i=0}^{p} (-1)^i \binom{p}{i} f\left(x + \left(\frac{p}{2} - i\right)h\right). \tag{15}$$

Next we define for any $p \in \mathbb{N}$, $q \in 2\mathbb{N}-1$ and $M > 0$ the monomials $f_p : [-M, M] \to \mathbb{R}$ and the tanh neural networks $\hat{f}_{q,h} : [-M, M] \to \mathbb{R}$ as

$$f_p(y) := y^p \quad \text{and} \quad \hat{f}_{q,h}(y) := \frac{\delta_{hy}^q[\sigma](0)}{\sigma^{(q)}(0) h^q}. \tag{16}$$

We first prove that these neural networks are accurate approximations to monomials with odd degree.

(15) is the p-th order central finite difference operator

For example:

$\boxed{p=1}$ $$f'(x) \approx \frac{\delta_h^1[f](x)}{h} = \frac{f(x+\frac{1}{2}h) - f(x-\frac{1}{2}h)}{h}$$

$\boxed{p=2}$ $$f''(x) \approx \frac{\delta_h^2[f](x)}{h^2},$$

where $$\delta_h^2[f](x) = \left[f((x+\tfrac{h}{2})+\tfrac{h}{2}) - f((x+\tfrac{h}{2})-\tfrac{h}{2})\right] - \left[f((x-\tfrac{h}{2})+\tfrac{h}{2}) - f((x-\tfrac{h}{2})-\tfrac{h}{2})\right]$$
$$= f(x+h) - 2f(x) + f(x-h)$$

$$\Rightarrow f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

$\boxed{p=3}$ $$f'''(x) \approx \frac{\delta_h^3[f](x)}{h^3},$$

where $$\delta_h^3[f](x) = \left[f((x+\tfrac{h}{2})+h) - 2f(x+\tfrac{h}{2}) + f((x+\tfrac{h}{2})-h)\right]$$
$$- \left[f((x-\tfrac{h}{2})+h) - 2f(x-\tfrac{h}{2}) + f((x-\tfrac{h}{2})-h)\right]$$
$$= f(x+\tfrac{3h}{2}) - 3f(x+\tfrac{h}{2}) + 3f(x-\tfrac{h}{2}) - f(x-\tfrac{3h}{2})$$

$$\Rightarrow f'''(x) \approx \frac{f(x+\frac{3h}{2}) - 3f(x+\frac{h}{2}) + 3f(x-\frac{h}{2}) - f(x-\frac{3h}{2})}{h^3}$$

By induction, we can get (15)

If we do Taylor expansion of f at x to $(q+1)$-th order:

$$f(x+(\tfrac{q}{2}-i)h) = f(x) + f'(x)((\tfrac{q}{2}-i)h) + \frac{f''(x)}{2!}((\tfrac{q}{2}-i)h)^2 + \cdots +$$

$$+\frac{f^{(q+1)}(x)}{(q+1)!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{q+1}+\frac{f^{(q+2)}(\xi_x)}{(q+2)!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{q+2}$$

$$=\sum_{l=0}^{q+1}\frac{f^{(l)}(x)}{l!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{l}+\frac{f^{(q+2)}(\xi_x)}{(q+2)!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{q+2}$$

then we plug this expansion into (15), we have

$$\delta_h^q[f](x)=\sum_{i=0}^{q}(-1)^i\binom{q}{i}f\left(x+\left(\tfrac{q}{2}-i\right)h\right)$$

$$=\sum_{i=0}^{q}(-1)^i\binom{q}{i}\sum_{l=0}^{q+1}\frac{f^{(l)}(x)}{l!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{l}+\sum_{i=0}^{q}(-1)^i\binom{q}{i}\frac{f^{(q+2)}(\xi_x)}{(q+2)!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{q+2}$$

$$=\sum_{l=0}^{q+1}\frac{f^{(l)}(x)}{l!}\cdot h^{l}\underbrace{\sum_{i=0}^{q}(-1)^i\binom{q}{i}\left(\tfrac{q}{2}-i\right)^{l}}_{\substack{\parallel\\ q!\text{ if } l=p;\ 0\text{ o.w. from Katsuura (2009)}}}+\sum_{i=0}^{q}(-1)^i\binom{q}{i}\frac{f^{(q+2)}(\xi_x)}{(q+2)!}\left(\left(\tfrac{q}{2}-i\right)h\right)^{q+2}$$

$$=\frac{f^{(q)}(x)}{\cancel{q!}}h^{q}\cdot\cancel{q!}+h^{q+2}\underbrace{\sum_{i=0}^{q}(-1)^i\binom{q}{i}\frac{f^{(q+2)}(\xi_x)}{(q+2)!}\left(\tfrac{q}{2}-i\right)^{q+2}}_{:=R(x,h)}$$

$$\Rightarrow\frac{\delta_h^q[f](x)}{h^q}=f^{(q)}(x)+h^2R(x,h)$$

$$\left(\sigma(x)=\tanh(x)=\frac{e^{2x}+1}{e^{2x}-1}\right.$$

If we change the step length from h to yh, then we have

$$\delta_{yh}^q[f](x)=y^q\left(f^{(q)}(x)h^q+h^{q+2}R(x,h)\right)\Rightarrow\frac{\delta_{yh}^q[f](x)}{(yh)^q}=f^{(q)}(x)+h^2R(x,h)$$

Next, take f as σ:

$$\frac{\delta_{yh}^q[\sigma](x)}{(yh)^q}=\sigma^{(q)}(x)+h^2R(x,h)$$

$$\Rightarrow\frac{\delta_{yh}^q[\sigma](0)}{h^q}=y^q\sigma^{(q)}(0)+h^2R(x,h)$$

$$\Rightarrow\frac{\delta_{yh}^q[\sigma](0)}{\sigma^{(q)}(0)h^q}=y^q+O(h^2)$$

That's where (16) comes from.

Lemma 3.1. *Let $k \in \mathbb{N}_0$ and $s \in 2\mathbb{N} - 1$. Then it holds that for all $\epsilon > 0$ there exists a shallow tanh neural network $\Psi_{s,\epsilon} : [-M, M] \to \mathbb{R}^{\frac{s+1}{2}}$ of width $\frac{s+1}{2}$ such that*

$$\max_{\substack{p \le s, \\ p \text{ odd}}} \left\| f_p - (\Psi_{s,\epsilon})_{\frac{p+1}{2}} \right\|_{W^{k,\infty}} \le \epsilon, \tag{17}$$

$(\frac{p+1}{2})$th element of output $= y^p$

Moreover, the weights of $\Psi_{s,\epsilon}$ scale as $O\left(\epsilon^{-s/2}(2(s+2)\sqrt{2M})^{s(s+3)}\right)$ for small ϵ and large s.

Proof. Let $p \le s$ be odd and let $0 < h < 2/pM$ (i). Let $0 \le m \le \min\{k, p+1\}$ (ii). Then Taylor's theorem guarantees the existence of

The lemma states there always exists a (shallow) 1-hidden layer neural network $\Psi_{s\epsilon}$ with width $\frac{s+1}{2}$ that approximates monomials (f_p) y^p (p is odd) in interval $[-M, M]$ under any required error ϵ in Sobolov space with the weights of neural network are in a controllable manner.

In which, we can say in the $W^{k,\infty}$ sense, we aim to approximate the error of not only function value itself but also k-th derivative in L^∞ space.

(i) First, since we want to evaluate the derivative by the nodes $(\frac{p}{2} - i)hx$, $i = 0, \dots, p$. $x \in [-M, M]$, and $\max_i |(\frac{p}{2} - i)hx| \le \frac{p}{2} h|x| \le \frac{p}{2} hM$

if we have $h < \frac{2}{pM}$, then all nodes are in the interval $[-1, 1]$

(ii) In $W^{k,\infty}$ norm, $m \le k$ is clear.

| 為了將誤差分成2段估計

$\xi_{x,i}$ such that

$$\begin{aligned}
\frac{d^m}{dx^m}\delta^p_{hx}[\sigma](0) &= \sum_{i=0}^{p}(-1)^i\binom{p}{i}\left(\frac{p}{2}-i\right)^m h^m \cdot \sigma^{(m)}\left(\left(\frac{p}{2}-i\right)hx\right) \\
&= \sum_{i=0}^{p}(-1)^i\binom{p}{i}\left(\frac{p}{2}-i\right)^m h^m \\
&\quad \left(\sum_{l=m}^{p+1}\frac{\sigma^{(l)}(0)}{(l-m)!}\left(\frac{p}{2}-i\right)^{l-m}(hx)^{l-m}\right) \\
&\quad + \sum_{i=0}^{p}(-1)^i\binom{p}{i}\left(\frac{p}{2}-i\right)^m h^m \\
&\quad \frac{\sigma^{(p+2)}(\xi_{x,i})}{(p+2-m)!}\left(\frac{p}{2}-i\right)^{p+2-m}(hx)^{p+2-m}.
\end{aligned}$$

Recall $\delta^p_{hx}[\sigma](0) = \sum_{i=0}^{p}(-1)^i\binom{p}{i}\sigma\left(\left(\frac{p}{2}-i\right)hx\right)$

$$\Rightarrow \frac{d^m}{dx^m}\delta^p_{hx}[\sigma](0) = \sum_{i=0}^{p}(-1)^i\binom{p}{i}\underbrace{\sigma^{(m)}\left(\left(\frac{p}{2}-i\right)hx\right)}_{(*)}\cdot\left(\frac{p}{2}-i\right)^m h^m$$

$(*)$ expansion at $(\frac{p}{2}-i)hx = 0$ to $(p+1)$th-order

$$\begin{aligned}
\text{where } (*) = \sigma^{(m)}(0) + \Big(&\sigma^{(m+1)}(0)\cdot\left(\left(\tfrac{p}{2}-i\right)hx\right) \\
&+ \frac{\sigma^{(m+2)}(0)}{2!}\left(\left(\tfrac{p}{2}-i\right)hx\right)^2 + \dots \\
&+ \frac{\sigma^{(p+1)}(0)}{(p+1-m)!}\left(\left(\tfrac{p}{2}-i\right)hx\right)^{p+1-m}\Big) \\
&+ \frac{\sigma^{(p+2)}(\xi_{x,i})}{(p+2-m)!}\left(\left(\tfrac{p}{2}-i\right)hx\right)^{p+2-m}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \frac{d^m}{dx^m}\delta^p_{hx}[\sigma](0) = \sum_{i=0}^{p}(-1)^p\binom{p}{i}\left(\frac{p}{2}-i\right)^m h^m\Bigg[&\sum_{l=m}^{p+1}\frac{\sigma^{(l)}(0)}{(l-m)!}\left(\frac{p}{2}-i\right)^{l-m}(hx)^{l-m} \\
&+ \frac{\sigma^{(p+2)}(\xi_{x,i})}{(p+2-m)!}\left(\frac{p}{2}-i\right)^{p+2-m}(hx)^{p+2-m}\Bigg]
\end{aligned}$$

$$= \sum_{i=0}^{p} (-1)^p \binom{p}{i} \left(\frac{p}{2}-i\right)^m h^m \left(\sum_{l=m}^{p+1} \frac{\sigma^{(l)}(0)}{(l-m)!} \left(\frac{p}{2}-i\right)^{l-m} (hx)^{l-m} \right)$$

$$+ \sum_{i=0}^{p} (-1)^p \binom{p}{i} \left(\frac{p}{2}-i\right)^m h^m \frac{\sigma^{(p+2)}(\xi_{x,i})}{(p+2-m)!} \left(\frac{p}{2}-i\right)^{p+2-m} (hx)^{p+2-m}$$

From Katsuura (2009, Theorem 1) it follows that

$$\sum_{i=0}^{p} (-1)^i \binom{p}{i} \left(\frac{p}{2}-i\right)^l = p!\,\delta(l-p) = \begin{cases} p!, & (l=p), \\ 0, & (l \neq p). \end{cases} \qquad (18)$$

for $l = 0, \ldots, p$. We observe that (18) remains true also for $l = p+1$, since all summands change sign when i is replaced by $p - i$. Using this fact, we can then rewrite the first term as

$$\sum_{i=0}^{p} (-1)^i \binom{p}{i} \left(\frac{p}{2}-i\right)^m h^m \left(\sum_{l=m}^{p+1} \frac{\sigma^{(l)}(0)}{(l-m)!} \left(\frac{p}{2}-i\right)^{l-m} (hx)^{l-m} \right)$$
$$= h^m \sum_{l=m}^{p+1} \frac{\sigma^{(l)}(0)}{(l-m)!} (hx)^{l-m} \sum_{i=0}^{p} (-1)^i \binom{p}{i} \left(\frac{p}{2}-i\right)^l$$
$$= \left\{ \begin{matrix} h^m \frac{\sigma^{(p)}(0)}{(p-m)!} (hx)^{p-m} p!, & 0 \le m \le p \\ 0, & m = p+1 \end{matrix} \right\} = h^p \sigma^{(p)}(0) f_p^{(m)}(x). \qquad (19)$$

From Katsuura (2009) we already have (18) holds for $l = 0, \ldots, p$

Moreover, when $l = p+1$, (18) also holds since

$$\sum_{i=0}^{p} (-1)^i \binom{p}{i} \left(\frac{p}{2}-i\right)^{p+1} \qquad \text{let } j = p - i$$

$$= \sum_{j=0}^{p} (-1)^{p-j} \binom{p}{p-j} \left(\frac{p}{2} - (p-j)\right)^{p+1}$$

$$= \sum_{j=0}^{p} (-1)^{p-j} \binom{p}{j} \left(j - \frac{p}{2}\right)^{p+1}$$

$$= \sum_{j=0}^{p} (-1)^{p-j+p+1} \binom{p}{j} \left(\frac{p}{2} - j\right)^{p+1}$$

$$= (-1)^{2p+1} \sum_{j=0}^{p} (-1)^{-j} \binom{p}{j} \left(\frac{p}{2} - j\right)^{p+1}$$

$$= -\sum_{j=0}^{p} (-1)^j \binom{p}{j} \left(\frac{p}{2} - j\right)^{p+1} \Rightarrow \text{(18) also holds.}$$

Therefore, the first term of above

$$\sum_{i=0}^{p} (-1)^p \binom{p}{i} \left(\frac{p}{2}-i\right)^m h^m \left(\sum_{l=m}^{p+1} \frac{\sigma^{(l)}(0)}{(l-m)!} \left(\frac{p}{2}-i\right)^{l-m} (hx)^{l-m} \right)$$

$$= h^m \sum_{l=m}^{p+1} \frac{\sigma^{(l)}(0)}{(l-m)!} (hx)^{l-m} \sum_{i=0}^{p} (-1)^p \binom{p}{i} \left(\frac{p}{2}-i\right)^l$$

If $0 \le m \le p$, l can always pass $l = p$ through the sum

$\therefore$ the first term $= h^m \frac{\sigma^{(p)}(0)}{(p-m)!} (hx)^{p-m} \cdot p!$, otherwise $= 0$

and notice that $f_p^{(m)}(x) = \frac{d^m}{dx^m} x^p = \frac{p!}{(p-m)!} x^{p-m}$ for $0 \le m \le p$, and $f_p^{(p+1)}(x) = 0$

$\Rightarrow$ it can be written as $h^p \sigma^{(p)}(0) f_p^{(m)}(x)$. That is (19).

Combining the previous results, it thus follows that we have

$$\hat{f}_{p,h}^{(m)}(x) - f_p^{(m)}(x) = \sum_{i=0}^{p}(-1)^i \binom{p}{i} \frac{1}{(p+2-m)!} \frac{\sigma^{(p+2)}(\xi_{x,i})}{\sigma^{(p)}(0)} \left(\frac{p}{2}-i\right)^{p+2} h^2 x^{p+2-m}.$$

Recall $\hat{f}_{p,h}(x) = \dfrac{\delta^p_{hx}[\sigma](0)}{\sigma^{(p)}(0)h^p}$

$$\Rightarrow \hat{f}^{(m)}_{p,h}(x) = \frac{d^m}{dx^m}\left(\frac{\delta^p_{hx}[\sigma](0)}{\sigma^{(p)}(0)h^p}\right) = \frac{\frac{d^m}{dx^m}\delta^p_{hx}[\sigma](0)}{\sigma^{(p)}(0)h^p}$$

$$\Rightarrow \hat{f}^{(m)}_{p,h}(x) - f^{(m)}_p(x) =$$

$$\frac{h^p\sigma^{(p)}(0)f_p^{(m)}(x) + \sum_{i=0}^{p}(-1)^p\binom{p}{i}\left(\frac{p}{2}-i\right)^m h^m \frac{\sigma^{(p+2)}(\xi_{x,i})}{(p+2-m)!}\left(\frac{p}{2}-i\right)^{p+2-m}(hx)^{p+2-m}}{\sigma^{(p)}(0)h^p} - f_p^{(m)}(x)$$

$$= \cancel{f_p^{(m)}(x)} + \sum_{i=0}^{p}(-1)^p\binom{p}{i}\left(\frac{p}{2}-i\right)^{p+2} h^{\cancel{p}+2} x^{p+2-m} \cdot \frac{\sigma^{(p+2)}(\xi_{x,i})}{\sigma^{(p)}(0)\cancel{h^p}} \cdot \frac{1}{(p+2-m)!} - \cancel{f_p^{(m)}(x)}$$

$$= \sum_{i=0}^{p}(-1)^p\binom{p}{i}\frac{1}{(p+2-m)!}\frac{\sigma^{(p+2)}(\xi_{x,i})}{\sigma^{(p)}(0)}\left(\frac{p}{2}-i\right)^{p+2}h^2x^{p+2-m}$$

Together with the lower and upper bounds on the derivatives of σ from Lemmas A.1 and A.4, this yields for $m \le \min(k, p+1)$:

$$\left|f_p - \hat{f}_{p,h}\right|_{W^{m,\infty}} \le \sum_{i=0}^{p}\binom{p}{i}\frac{|\sigma^{(p+2)}(\xi_{x,i})|}{|\sigma^{(p)}(0)|}\left|\frac{p}{2}-i\right|^{p+2}h^2M^{p+2}$$
$$\le 2^p\frac{(2(p+2))^{p+3}}{1}\left(\frac{p}{2}\right)^{p+2}h^2M^{p+2} \qquad (20)$$
$$\le (2(p+2)pM)^{p+3}h^2.$$

If $k \le p+1$, then this shows that

$$\left\|f_p - \hat{f}_{p,h}\right\|_{W^{k,\infty}} \le (2(p+2)pM)^{p+3}h^2. \qquad (21)$$

$$|f|_{W^{m,\infty}(\Omega)} = \max_{|\alpha|=m}\|D^\alpha f\|_{L^\infty(\Omega)} \quad \text{for } m = 0, \ldots, k. \qquad (9)$$

Lemma A.1. *It holds for every $n \in \mathbb{N}$ that $\left|\tanh^{(2n-1)}(0)\right| \ge 1$.*

Lemma A.4. *Let $m \in \mathbb{N}$. Then it holds that*

$$\left|\sigma^{(m)}(x)\right| \le (2m)^{m+1}\min\{\exp(-2x), \exp(2x)\} \quad \text{for all } x \in \mathbb{R}. \qquad (A.15)$$

Now, we can evaluate the seminorm for $m \le \min(k, p+1)$

$$|f_p - \hat{f}_{p,h}|_{W^{m,\infty}} = \max_{|\alpha|=m}\|D^\alpha(f_p - \hat{f}_{p,h})\|_{L^\infty(\Omega)} = \max\|f_p^{(m)} - \hat{f}^{(m)}_{p,h}\|_{L^\infty(\Omega)}$$

$$= \max_{x\in[-M,M]}\left(\sum_{i=0}^{p}(-1)^p\binom{p}{i}\frac{1}{(p+2-m)!}\frac{\sigma^{(p+2)}(\xi_{x,i})}{\sigma^{(p)}(0)}\left(\frac{p}{2}-i\right)^{p+2}h^2x^{p+2-m}\right)$$

$\le \left|\frac{p}{2}-i\right|^{p+2} \le \left|\frac{p}{2}\right|^{p+2}$

$(\because m \le p+1 \Rightarrow (p+2-m)! \ge 1 \Rightarrow \frac{1}{(p+2-m)!} \le 1$

$\le |x|^{p+2-m} \le M^{p+2-m} \le M^{p+2}$

$$\le \sum_{i=0}^{p}\left|(-1)^p\right|\left|\binom{p}{i}\right| \cdot 1 \cdot \frac{|\sigma^{(p+2)}(\xi_{x,i})|}{|\sigma^{(p)}(0)|}\left|\frac{p}{2}\right|^{p+2}h^2M^{p+2}$$

From Lemma A.1 we have $|\sigma^{(p)}(0)| = |\tanh^{(p)}(0)| \ge 1$

A.4 $\quad |\sigma^{(p+2)}(x)| \le (2(p+2))^{p+2+1}\min\{e^{-2x}, e^{2x}\} \ \forall x \in \mathbb{R}$

$\le (2(p+2))^{p+3} \cdot 1$

$$\Rightarrow |f_p^{(m)} - \hat{f}_{p,h}^{(m)}|_{W^{m,\infty}} \leq \sum_{i=0}^{p} \binom{p}{i} \frac{(2(p+2))^{p+3}}{1} \cdot \left|\frac{p}{2}\right|^{p+2} h^2 M^{p+2}$$

$$= (2(p+2))^{p+3} \cdot \left(\frac{p}{2}\right)^{p+2} h^2 M^{p+2} \sum_{i=0}^{p} \binom{p}{i}$$

$$= (2(p+2))^{p+3} \cdot \left(\frac{p}{2}\right)^{p+2} h^2 M^{p+2} \cdot 2^p$$

$$= (2(p+2))^{p+3} \cdot \frac{p^{p+2}}{4} h^2 M^{p+2}$$

$$\leq (2(p+2))^{p+3} \cdot p^{p+3} h^2 M^{p+3} = (2p(p+2)M)^{p+3} h^2 \quad \text{for } m=0,\dots,k$$

If $k \leq p+1$, then $m \leq p+1$, consider the norm

$$\|f_p - \hat{f}_{p,h}\|_{W^{k,\infty}} = \max_{0\leq m\leq k} |f_p^{(m)} - \hat{f}_{p,h}^{(m)}|_{W^{m,\infty}} \leq (2p(p+2)M)^{p+3} h^2$$

If $k > p+1$, let $p+2 \leq m \leq k$. In this case, $f_p^{(m)} = 0$, therefore it suffices to bound $\hat{f}_{p,h}^{(m)}$. We see that for $0 < h < 1$,

$$\begin{aligned} \left|\hat{f}_{p,h}^{(m)}(x)\right| &= \left|\frac{1}{h^p \sigma^{(p)}(0)} \sum_{i=0}^{p} (-1)^i \binom{p}{i} \left(\frac{p}{2} - i\right)^m h^m \cdot \sigma^{(m)}\left(\left(\frac{p}{2} - i\right) hx\right)\right| \\ &\leq 2 \sum_{i=0}^{p} \binom{p}{i} \left|\frac{p}{2} - i\right|^m h^2 (2m)^{m+1} \\ &\leq 2^{p+1} \left(\frac{p}{2}\right)^k (2k)^{k+1} h^2 \leq (2pk)^{k+1} h^2. \end{aligned} \tag{22}$$

Next, we need to consider when $k > p+1$, i.e. $p+2 \leq m \leq k$. In this case $f_p^{(m)} = 0$.

Therefore, we only need to consider

$$|\hat{f}_{p,h}^{(m)}(x)| = \left|\frac{\frac{d^m}{dx^m}\left(\delta_{hx}^{p}[\sigma](0)\right)}{\sigma^{(p)}(0)\, h^p}\right| \overset{(P.18)}{=}$$

$$\left|\frac{1}{\sigma^{(p)}(0)\, \cancel{h^p}} \sum_{i=0}^{p} (-1)^i \binom{p}{i} \sigma^{(m)}\left(\left(\frac{p}{2} - i\right)hx\right) \cdot \left(\frac{p}{2} - i\right)^m h^{m}\right| \quad m-p \geq 2$$

$$\leq \frac{1}{1} \sum_{i=0}^{p} \binom{p}{i} (2m)^{m+1} \left|\frac{p}{2} - i\right|^m \cdot h^2 = (2m)^{m+1} h^2 \sum_{i=0}^{p} \binom{p}{i} \left(\frac{p}{2}\right)^k$$

$$\leq (2m)^{m+1} h^2 2^p \cdot \left(\frac{p}{2}\right)^k \leq (2k)^{k+1} h^2 \cdot 2^p \cdot \left(\frac{p}{2}\right)^k \leq h^2 \cdot (2k)^{k+1} \cdot 2^k \cdot \left(\frac{p}{2}\right)^k$$

$$\leq h^2 \cdot \left(2k \cdot 2 \cdot \frac{p}{2}\right)^{k+1} = (2kp)^{k+1} h^2$$ (Actually there's better approximation, just for simple)

We thus obtain, for arbitrary $k \in \mathbb{N}$:

$$\left\|f_p - \hat{f}_{p,h}\right\|_{W^{k,\infty}} \leq \left((2(p+2)pM)^{p+3} + (2pk)^{k+1}\right) h^2 =: \epsilon. \tag{23}$$

Therefore, we have considered m-th derivative for $0 \leq m \leq k$, the error:

$$\|f_p - \hat{f}_{p,h}\|_{W^{k,\infty}} \leq \left((2(p+2)pM)^{p+3} + (2kp)^{k+1}\right) h^2 := \varepsilon \quad \text{for arbitrary } k \in \mathbb{N}$$

Furthermore observe that the weights scale as $O\left(\max_i \binom{p}{i} h^{-p}\right)$. For $\epsilon \to 0$ and large p, it holds that $O(h^{-p}) = O\left(\epsilon^{-p/2}((p+2)\sqrt{2M})^{p(p+3)}\right)$, where the implied constant depends on k. Next, we find using Stirling's approximation that for $0 \le i \le p$ it holds that

$$\binom{p}{i} \le \binom{p}{\frac{p-1}{2}} \le \frac{ep^{p+1/2}}{2\pi\left(\frac{p-1}{2}\right)^{\frac{p}{2}}\left(\frac{p+1}{2}\right)^{\frac{p}{2}+1}} = O\left(\frac{2^p}{\sqrt{p}}\right). \quad (24)$$

The weights therefore scale as $O\left(\epsilon^{-p/2}(2(p+2)\sqrt{2M})^{p(p+3)}\right)$.

From

$$\hat{f}_{p,h}(y) = \frac{1}{\sigma^{(p)}(0)h^p}\sum_{i=0}^{p}(-1)^i\binom{p}{i}\sigma\left(\left(\frac{p}{2}-i\right)hy\right)$$

we know the weight $= \frac{1}{\sigma^{(p)}(0)h^p}(-1)^i\binom{p}{i}$ for $i=0,\dots,p$

, which $\le \max_i\left(\binom{p}{i}h^{-p}\right) = O\left(\max_i\binom{p}{i}h^{-p}\right)$

From (23) we have $\left[\left((2(p+2)pM)^{p+3}+(2kp)^{k+1}\right)h^2\right]^{-\frac{p}{2}} = \varepsilon^{-\frac{p}{2}}$

$$\Rightarrow h^{-p} = \varepsilon^{-\frac{p}{2}}\cdot\left[(2(p+2)pM)^{p+3}+(2kp)^{k+1}\right]^{\frac{p}{2}}$$

$$\underset{\text{(when } p \text{ is large)}}{\approx} \varepsilon^{-p/2}\left(\sqrt{2p(p+2)M}\right)^{p(p+3)}$$

$$\le \varepsilon^{-p/2}\left(\sqrt{2(p+2)^2M}\right)^{p(p+3)}$$

$$= \varepsilon^{-p/2}\left((p+2)\sqrt{2M}\right)^{p(p+3)}$$

$$= O\left(\varepsilon^{-p/2}((p+2)\sqrt{2M})^{p(p+3)}\right)$$

Next, we are going to approximate $\binom{p}{i}$ by Stirling's approximation:

$$\sqrt{2\pi}\, e^{-n} n^{n+\frac{1}{2}} < n! \le e n^{n+\frac{1}{2}} e^{-n}$$

for both n,k large

$$\Rightarrow \binom{p}{i} \le \binom{p}{\frac{p-1}{2}} = \frac{p!}{\left(\frac{p-1}{2}\right)!\left(\frac{p+1}{2}\right)!} \le \frac{p^{p+\frac{1}{2}}e^{1-p}}{2\pi\cdot e^{-p}\left(\frac{p-1}{2}\right)^{\frac{p}{2}}\left(\frac{p+1}{2}\right)^{\frac{p}{2}+1}}$$

$$= \frac{ep^{p+\frac{1}{2}}}{2\pi\left(\frac{p-1}{2}\right)^{\frac{p}{2}}\left(\frac{p+1}{2}\right)^{\frac{p}{2}+1}}$$

where $\left(\frac{p-1}{2}\right)^{\frac{p}{2}}\left(\frac{p+1}{2}\right)^{\frac{p}{2}+1}$

$$= \left(\frac{p^2-1}{4}\right)^{\frac{p}{2}}\cdot\frac{p+1}{2} = \left(\frac{p^2}{4}\left(\frac{p^2-1}{p^2}\right)\right)^{\frac{p}{2}}\cdot\frac{p+1}{2} = \left(\frac{p}{2}\right)^p\left(1-\frac{1}{p^2}\right)^{\frac{p}{2}}\cdot\frac{p+1}{2} > \left(\frac{p}{2}\right)^{p+1}c_0$$

which $\left(1-\frac{1}{p^2}\right)^{\frac{p}{2}} \ge c_0$ for some $c_0 > 0$

$$\Rightarrow \binom{p}{i} \le \frac{ep^{p+\frac{1}{2}}}{2\pi c_0\left(\frac{p}{2}\right)^{p+1}} = \frac{e}{2\pi c_0}\cdot\frac{2^{p+1}\cdot p^{p+\frac{1}{2}}}{p^{p+1}} = \frac{e}{\pi c_0}\cdot\frac{2^p}{\sqrt{p}} = O\left(\frac{2^p}{\sqrt{p}}\right)$$

Therefore, the weight scale is $O\left(\frac{2^p}{\sqrt{p}}\cdot\varepsilon^{-p/2}((p+2)\sqrt{2M})^{p(p+3)}\right)$

Then, we can choose $O(\varepsilon^{-p/2}(2(p+2)\sqrt{2M})^{p(p+3)})$ for a looser bound but more elegant.

Finally, we can define shallow tanh neural network $\Psi_{s,\varepsilon}$ by $(\Psi_{s,\varepsilon})_p = \hat{f}_{p,h}$ with $\frac{s+1}{2}$ neurons in hidden layer. □

$\boxed{\text{Lemma 3.2}}$

Here we want to do the same thing as in Lemma 3.1, but here for even degree. First we observe the relation between odd and even degree monomials.

$$y^{2n} = \frac{1}{2\alpha(2n+1)}\left((y+\alpha)^{2n+1} - (y-\alpha)^{2n+1} - 2\sum_{k=0}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}y^{2k}\right). \quad (25)$$

For example, if $n=1$:

$$\begin{aligned}
(y+\alpha)^3 &= y^3 + 3\alpha y^2 + 3\alpha^2 y + \alpha^3 \\
-)\ (y-\alpha)^3 &= y^3 - 3\alpha y^2 + 3\alpha^2 y - \alpha^3 \\
\hline
(y+\alpha)^3 - (y-\alpha)^3 &= 2\cdot 3\alpha y^2 + 2\cdot\alpha^3 \\
y^2 &= \left((y+\alpha)^3 - (y-\alpha)^3 - 2\alpha^3\right)\frac{1}{2\cdot 3\alpha}
\end{aligned}$$

if $n=2$:

$$\begin{aligned}
(y+\alpha)^5 &= y^5 + \binom{5}{1}\alpha y^4 + \binom{5}{2}\alpha^2 y^3 + \binom{5}{3}\alpha^3 y^2 + \binom{5}{4}\alpha^4 y + \alpha^5 \\
-)\ (y-\alpha)^5 &= y^5 - \binom{5}{1}\alpha y^4 + \binom{5}{2}\alpha^2 y^3 - \binom{5}{3}\alpha^3 y^2 + \binom{5}{4}\alpha^4 y - \alpha^5 \\
\hline
(y+\alpha)^5 - (y-\alpha)^5 &= 2\binom{5}{1}\alpha y^4 + 2\binom{5}{3}\alpha^3 y^2 + 2\alpha^5 \\
y^4 &= \left((y+\alpha)^5 - (y-\alpha)^5 - \sum_{k=0}^{1}\binom{5}{2k}\alpha^{2(2-k)+1}y^{2k}\right)\cdot\frac{1}{2(2\cdot 2+1)\alpha}
\end{aligned}$$

We can see the key point is to eliminate odd terms by $(y+\alpha)^{2n+1} - (y-\alpha)^{2n+1}$ and other even terms y^{2k} with $2k<2n$ can be written together. Thus we can attain y^{2n} term recursively.

Next, we want to find $\Psi_{s,\varepsilon}(y)$ which p-th output $= \hat{f}_{p,h}(y)$ for p odd
$= (\Psi_{s,\varepsilon}(y))_{2n}$ for p even $(=2n)$

and $(\psi_{s,\varepsilon}(y))_0 = 1$, by (25), we have

$$(\psi_{s,\epsilon}(y))_{2n} = \frac{1}{2\alpha(2n+1)}\left(\hat{f}_{2n+1,h}(y+\alpha) - \hat{f}_{2n+1,h}(y-\alpha) \quad -2\sum_{k=0}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}(\psi_{s,\epsilon}(y))_{2k}\right). \quad \cdots (27)$$

From now on, we want to estimate the error, starting with define

$$E_p = \| f_p - (\psi_{s,\varepsilon})_p \|_{W^{k,\infty}}$$

and we want for $p \le s$, $E_p \le E_p^* := \dfrac{2^{p/2}(1+\alpha)^{(p^2+p)/2}}{\alpha^{p/2}} \cdot \epsilon.$ (28)

where the coefficients of ε is designed for later estimation

If we choose h as in Lemma 3.1, then we directly have $\max\limits_{\substack{p\le s \\ p \text{ is odd}}} E_p \le \varepsilon$ (29)

which proves (28) for odd p.

Now we prove (28) also holds for even p using induction

First note that when $p=2$:

$$E_2 = \| f_2 - (\psi_{s,\varepsilon})_2 \|_{W^{k,\infty}} = \Big\| \frac{1}{6\alpha}\big((y+\alpha)^3 - (y-\alpha)^3 - 2\alpha^3\big)$$
$$- \frac{1}{6\alpha}\big(\hat{f}_{3,h}(y+\alpha) - \hat{f}_{3,h}(y-\alpha) - 2\alpha^3\big)\Big\|_{W^{k,\infty}}$$

$$= \Big\| \frac{1}{6\alpha}\big(|(y+\alpha)^3 - \hat{f}_{3,h}(y+\alpha)| + |(y-\alpha)^3 - \hat{f}_{3,h}(y-\alpha)|\big)\Big\|_{W^{k,\infty}} \le \frac{1}{6\alpha}\cdot(\varepsilon+\varepsilon)$$

$$= \frac{\varepsilon}{\alpha}\cdot\frac{1}{3} < \frac{\varepsilon}{\alpha}\cdot 2(1+\alpha)^3 = E_2^*$$, which proves the base step.

Let $n \in \mathbb{N}$ s.t. $2n+1 \le s$, and assume by the induction hypothesis that $E_{2k} \le E_{2k}^*$ for $k<n$

Similarly, we can also have the estimation: $E_{2n} \le \dfrac{1}{2\alpha(2n+1)}\left(E_{2n+1} + E_{2n+1} + 2\sum_{k=1}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}E_{2k}\right).$ (31)

Note that E_{2k}^* is increasing by definition and $E_{2k} \le E_{2k}^*$ for $k<n$

for $k=n-1$, we have $E_{2k} \le E_{2k}^* \le E_{2(n-1)}^*$

also by definition of $E_p^* = \dfrac{\overbrace{2^{p/2}}^{\ge 1}\overbrace{(1+\alpha)^{(p^2+p)/2}}^{\ge 1}}{\underbrace{\alpha^{p/2}}_{\le 1}}\varepsilon \ge \varepsilon$

Now we can estimate (31):

$$E_{2n} \le \frac{1}{2\alpha(2n+1)}\left(2\max_{\substack{p\le s\\ p \text{ is odd}}} E_p + 2\sum_{k=1}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}E_{2k}\right)$$

$$= \frac{1}{\alpha(2n+1)}\left(\max_{\substack{p\le s\\ p \text{ is odd}}} E_p + \sum_{k=1}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}E_{2k}\right)$$

$$\le \frac{1}{\alpha}\left(\varepsilon + E_{2k}\sum_{k=1}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}\right)$$

$$\underset{(j=n-k)}{\le} \frac{1}{\alpha}\left(E^{*}_{2(n-1)} + E^{*}_{2(n-1)}\,\alpha\sum_{j=1}^{n-1}\binom{2n+1}{2n-2j}\alpha^{2j}\right)$$

$2n+1$

$2n-2,\ 2n-4 \ \dots\ 2$

$$\le \frac{1}{\alpha}\left(E^{*}_{2(n-1)} + E^{*}_{2(n-1)}\,\alpha\sum_{j=0}^{2n+1}\binom{2n+1}{2n-2j}\alpha^{j}\right)$$

use all terms to bound

$$= \frac{1}{\alpha}\left(E^{*}_{2(n-1)} + E^{*}_{2(n-1)}\,\alpha(1+\alpha)^{2n+1}\right)$$

$$= \frac{E^{*}_{2(n-1)}}{\alpha}\left(1+\alpha(1+\alpha)^{2n+1}\right) \le \frac{2}{\alpha}E^{*}_{2(n-1)}(1+\alpha)^{2n+1}$$, which is (32)

Recall the definition (28), $E^{*}_{2(n-1)} = \dfrac{2^{n-1}(1+\alpha)^{(n-1)(2n-1)}}{\alpha^{n-1}}\,\varepsilon$

$$\Rightarrow E_{2n} \le \frac{2}{\alpha}\cdot\frac{2^{n-1}}{\alpha^{n-1}}(1+\alpha)^{(n-1)(2n-1)+(2n+1)}\,\varepsilon$$

$$= \left(\left(\frac{2}{\alpha}\right)^{n}\cdot(1+\alpha)^{2n^2-n+2}\right)\varepsilon \le \left(\frac{2}{\alpha}(1+\alpha)^{2n+1}\right)^{n}\varepsilon$$

$$= \frac{2^{n}(1+\alpha)^{2n^2+n}}{\alpha^{n}}\,\varepsilon = E^{*}_{2n}$$, which is (33).

Therefore, we have claimed (28) is also true of even p.

Next, by lemma A.2, we have the optimized $\alpha = \frac{1}{s}$

Lemma A.2. *Let $s \in 2\mathbb{N}-1$. It holds that*

$$\inf_{\alpha>0}\frac{2^{s/2}(1+\alpha)^{(s^2+s)/2}}{\alpha^{s/2}} \le \sqrt{e}(2es)^{s/2}, \tag{A.3}$$

where the infimum is reached at $\alpha = 1/s$.

Then, we have the conclusion that

$$\max_{p\le s}\left\|f_p - (\psi_{s,\epsilon})_p\right\|_{W^{k,\infty}} \le \sqrt{e}(2es)^{s/2}\epsilon. \tag{34}$$

Now, let's estimate the weight, we only need to compare the weight which is multiplied from weight in Lemma 3.1

The factor $= \dfrac{1}{2\alpha(2n+1)}\cdot\left(2+2\displaystyle\sum_{k=0}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}\right)$ from (27)

$$= \frac{1}{2\alpha(2n+1)}\left(2+2\cdot\alpha(1+\alpha)^{2n+1}\right) \leftarrow \text{have been estimated when estimating (31)}$$

$$\leq \frac{1}{2\alpha(2n+1)}\left(2(1+\alpha)^{2n+1}\right)$$

$$= \frac{(1+\alpha)^{2n+1}}{\alpha(2n+1)} < \frac{(1+\alpha)^{s}}{\alpha(2n+1)} < \frac{(1+\alpha)^{s}}{\alpha} \overset{(\alpha=\frac{1}{s})}{=} s\left(1+\frac{1}{s}\right)^{s} = O(s)$$

Multiply the weight bound given by Lemma 3.1, we have

$$O\left(\epsilon^{-s/2}s(2\sqrt[4]{2es}\sqrt{2(M+1)}(s+2))^{s(s+3)}\right) = O\left(\epsilon^{-s/2}(\sqrt{M}(s+2))^{3s(s+3)/2}\right) \tag{36}$$

Finally, from Lemma 3.1, for $i=0,1,\dots,\frac{s-1}{2}$ and $\beta \in \{-\alpha, 0, \alpha\}$, activation function is given by $\sigma((\frac{s}{2}-i)h(y+\beta))$ $\Rightarrow$ $\frac{s+1}{2}\cdot 3$ neurons needed. $\square$