HW3 的 Lemma 大解析

1. 我們的目標:在區間$[-M,M]$上，用一層隱藏層的 tanh 網路近似單項式f_p(y)=y_p，而且用 Sobolev 的誤差來量化。關鍵工具是中心差分算子套到 tanh 上，做出一個像y_p的函數，然後證明它真的很像!!

2. Sobolev 範數的誤差：

Define: if $f \epsilon W^{k,\infty}(D)$, then

$$\|f\|_{W^{k,\infty}} = \max_{0\le m\le k}\max_{|\alpha|=m}\|\partial^\alpha f\|_{L^\infty(D)}$$

Moreover, $W^{k,p}(D)=\{f\epsilon L^\alpha(D): D^\alpha f\epsilon L^p(D), \forall \alpha \epsilon N_0^d \ with\ |\alpha|\le k\}$

為什麼我們要使用 Sobolev 誤差呢？如果只在意義下近似函數，有可能函數值很準，但導數很差。Sobolev 誤差強迫函數和導數一起是好的。所以只近似函數值還不夠，還要保證導數也準

3. Lemma3.1:

1.1 對所有奇數 p ≤ s，有一個淺層 tanh 網路，St. $\|y-((\psi_{s,\varepsilon})_p)\|_{W^{k,\infty}} \le \varepsilon$，而且同時處理所有奇數次到 s。首先我們先來定義中心差分，給定步長 h>0、整數 p≥1，對函數 f 定義：

$$(\Delta_h^{(p)}f)\,(x) := \sum_{i=0}^{p}(-1)^i\binom{p}{i}f\left(x+\left(\frac{p}{2}-i\right)h\right)$$

(1) $x+\left(\frac{p}{2}-i\right)h$ ：以 x 為中心，左右對稱的取樣點。

(2) $(-1)^i\binom{p}{i}$ ：來自二項式結構，讓我們有消去的效果。

1.2 Define 位移算子E^αf(x) :=f(x + αh).then,

$$(\Delta_h^{(p)}) = \sum_{i=0}^{p}(-1)^i\binom{p}{i}E^{\left(\frac{p}{2}-i\right)}=E^{\frac{p}{2}}(1-E^{-1})^p$$

1.3 Using E = e^{hD},where D = $\frac{d}{dx}$:

$$\Delta_h^{(p)}=(e^{\frac{h}{2}D}-e^{-\frac{h}{2}D})=(2sinh(\frac{h}{2}D))^p$$

Moreover ,$sinh()$只有幾數次冪，這會直接造成只出現 p , p + 2 , p+4 , ……階的 derivative。

1.4 Suppose z = $\frac{h}{2}D$, for $sinhz$ = z + $\frac{z^3}{3!}$ + O (z^5),then we have the following:

$$(2sinhz)^p = (2z)^p(z+\frac{z^2}{3!}+O(z^4))^p =$$

把 z 換回$\frac{h}{2}D$並且作用在 f:

$$(\Delta_h^{(p)}f)\,(x) = h^pf^p(x)\frac{p}{24}h^{p+2}p^{p+2}(x)O(h^{p+4})$$

$$\frac{(\Delta_h^{(p)}f)\,(x)}{h^p} = f^p(x)\frac{p}{24}h^{p+2}p^{p+2}(x)O(h^{p+4})$$

所以我們知道中心 p 階導數是二階精度，主要誤差係數是$\frac{p}{24}$ 。

1.5 我們現在要把上面算好的中心差分套用在 tanh 中，做出一個長得很像y^p的函數：

Define 定義近似器：

$$\boxed{\hat{f}_{p,h} = \frac{[\Delta_h^{(p)}(tanh)](hy)}{\tanh^{(p)}(0)h^p}}$$

(1) 上面分子是把 tanh 放進中心差分，在把 x 換成 hy(就只是在做尺度調整而已)

(2) 分母則是正規化，把主要係數調成 1

1.6 對 y 在 0 做 Taylor Expansion tanh(ay),where a is a coefficient.

Suppose that $a_i = (\frac{p}{2} - \mathrm{i})\mathrm{h}$.

$$\tanh(a_i \mathrm{y}) = \sum_{l=0}^{\infty} \frac{tanh^{(l)}(0)}{l!}(a_i \mathrm{y})^l$$

and then put it into 中心差分的權重，we get:

$$\boxed{S_l := \sum_{i=0}^{p}(-1)^i \binom{p}{i} \left(\frac{p}{2} - \mathrm{i}\right)}$$

S_l:決定了哪些階會被消除掉。

$$S_l = \begin{cases} 0\,, l < p \\ p\,, l = p, l = p+2 \\ 0, l = p+1 \end{cases}$$

(1) $\Delta_h^{(p)}$是 p 次差分，他殺掉了所有低於 p 次的 polynomial($0\,, l < p$)

(2)又因為中心(對稱性的 reason)，連$l = p + 1$都被對稱性消掉

(3)所以我們第一個留下的是$l = p$(而且係數剛好是 p!) ，下一個是$l = p + 2$

1.7 By 1.6 we can get:

$$[\Delta_h^{(p)}(tanh)](hy) = \tanh^{(p)}(0)h^p y^p + \mathrm{C}\ h^{p+2}y^{p+2} + \ldots\ldots$$

Where, C is the coefficient combination by $\tanh^{(p+2)}(0)$ and S_{p+2}

$$\hat{f}_{p,h} = y^p + \mathrm{O}\,(h^2)\cdot\ y^p + \cdots\cdots$$

1.8 升到 Sobolev：對於每個導數階 m≤k ，差距也就只是 O(h^2)，這是因為導數換到和，常數，tanh 的導數上依然都是 bonded。Therefore,$\exists C = C(k, M, p)$ is a coefficient St.

$$\left\| y^p - \hat{f}_{p,h} \right\|_{W^{k,\infty}([-M,M])} \leq \mathrm{C}h^2.$$

Let h = $\sqrt{\frac{\varepsilon}{c}}$,then

$$\boxed{\left\| y^p - \hat{f}_{p,h} \right\|_{W^{k,\infty}([-M,M])} \leq \varepsilon}$$

1.9 為什麼可以一次處理所有的奇數 p≤s 而且寬度是$\frac{s+1}{2}$?

(1)對最大的奇數 s，它用到的取樣點集合（那些（$\frac{s}{2} - i$）h）包含了所有較小奇數需要的點。

(2)tanh is odd function：tanh(−z)=−tanh(z)。所以不需要為正負各建一個神經元，一個就能兩邊用（權重帶負號就可以實現了）。

(3) 於是總共只要$\frac{s+1}{2}$個不同的正側取樣點（含 0 的那個），就能生成所有奇數單項式的輸出（每個 p 只是不同的輸出層線性組合）。

到此我們說明完 Lemma3.1。

4. Lemma3.2:

2.1 (1)Lemma 3.1 已經能處理所有奇數次冪y^p（p奇數），誤差是 $O(h^2)$，但是偶數次單項式 y^{2n}不是奇函數，沒辦法直接用 tanh （奇函數）生成。

(2)解法：用一個「代數恆等式」把 y^{2n}寫成奇數冪和低階偶數冪的組合。

2.2　從二項式展開差：

從二項式展開的差：$(y+\eta)^{2n+1}-(y-\eta)^{2n+1}$.

對兩邊做展開：

$$(1): (y+\eta)^{2n+1}=\sum_{m=0}^{2n+1}\binom{2n+1}{m}y^{2n+1-m}\eta^{m}.$$

$$(2): (y-\eta)^{2n+1}=\sum_{m=0}^{2n+1}\binom{2n+1}{m}y^{2n+1-m}(-\eta)^{m}.$$

Therefore, we have: $(y+\eta)^{2n+1}-(y-\eta)^{2n+1}=\sum_{m=0}^{2n+1}\binom{2n+1}{m}y^{2n+1-m}(\eta^m-(-\eta)^m)$.

2.3　只留下奇數次m，因為：

(1)如果m偶數，$\eta^m-(-\eta)^m=0$。

(2)如果m奇數，$\eta^m-(-\eta)^m=2\eta^m$。

$$Then, we\ have\ \ (y+\eta)^{2n+1}-(y-\eta)^{2n+1}=2\sum_{\substack{m=1\\ m\ \text{odd}}}^{2n+1}\binom{2n+1}{m}y^{2n+1-m}\eta^{m}.$$

2.4:　現在把y^{2n}拿出來，在上式中，取$m=1$的項：

$$2\binom{2n+1}{1}y^{2n}\eta=2(2n+1)y^{2n}\eta.$$

把它移到等號左邊，其餘$m\geq 3$（奇數）的項留在右邊：

$$(y+\eta)^{2n+1}-(y-\eta)^{2n+1}=2(2n+1)y^{2n}\eta+2\sum_{\substack{m=3\\ m\ \text{odd}}}^{2n+1}\binom{2n+1}{m}y^{2n+1-m}\eta^{m}.$$

然後我們兩邊同除$2\eta(2n+1)$可以得到：

$$\boxed{y^{2n} = \frac{(y+\eta)^{2n+1} - (y-\eta)^{2n+1}}{2\eta(2n+1)} - \sum_{\substack{m=3 \\ m\ \mathrm{odd}}}^{2n+1} \binom{2n+1}{m} \frac{y^{2n+1-m}\eta^{m-1}}{2n+1}.}$$

2.5 改索引，變成低階偶數，令 $m = 2(n-k)+1$，其中 $k = 0,1,\dots,n-1$。

(1) 當 $m = 3 \Rightarrow k = n-1$。

(2) 當 $m = 2n+1 \Rightarrow k = 0$。

代回：$y^{2n+1-m} = y^{2k}, \eta^{m-1} = \eta^{2(n-k)}$.

經過一番整理我們得到：

$$\boxed{y^{2n} = \frac{(y+\eta)^{2n+1} - (y-\eta)^{2n+1}}{2\eta(2n+1)} - 2\sum_{k=0}^{n-1} \binom{2n+1}{2k} \eta^{2(n-k)+1} y^{2k}.}$$

2.6 為什麼這公式有用？

(1)第一項：只含$(2n+1)$次冪（奇數），但輸入換成$y+\eta$、$y-\eta$。可以用 Lemma 3.1（奇數冪近似）去處理，只需要在神經網路的輸入加偏置。

(2)第二項：是更低階的偶數冪（$y^0, y^2, \dots, y^{2n-2}$）。可以用歸納（已經處理好低階偶數）來完成。

所以我們現在知道了：用奇數冪的近似 + 歸納，就能構造出所有偶數冪!!!!!!

2.7 誤差控制

(1)奇數部分（第一項）：誤差已經是$O(h^2)$。

(2)偶數部分（第二項）：由歸納假設，誤差也都是$O(h^2)$。

(3)選一個合適的η（paper 選$\eta = \frac{1}{s}$）來避免係數太大。

最後我們可以保證：$\boxed{\max_{p\le s} \| y^p - (\psi_{s,\varepsilon})_p \|_{W^{k,\infty}([-M,M])} \le \varepsilon.}$

5.網路寬度的解釋

為了用第一項，要同時近似 $(y-\eta)^{2n+1}, y^{2n+1}, (y+\eta)^{2n+1}$。所以需要三組「奇數冪近似」的神經元。每組需要 $(s+1)/2$ 個隱藏元（Lemma 3.1 的結論）。總寬度：

$$\boxed{\frac{3(s+1)}{2}}$$