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

$$\max_{\substack{p \le s \\ p \text{ odd}}} \| f_p - (\Psi_{s,\varepsilon})_{\frac{p+1}{2}} \|_{W^{k,\infty}} \le \varepsilon.$$

$$W^{k,p}(\Omega) = \{ f \in L^p(\Omega) : D^\alpha f \in L^p(\Omega) \text{ for all } \alpha \in \mathbb{N}_0^d \text{ with } |\alpha| \le k \}$$

這個 sobolev space 搜集了函數屬於 $L^p(\Omega)$ space 的函數，並它們的偏導函數也屬於 $L^p(\Omega)$ space，偏導數的次數總共要小於等於 k

e.g. $f(x,y) = x^2+y^2$, $|\alpha| \le 2$. $\frac{\partial f}{\partial x \partial y} = 0$, $|\alpha| = 1+1 = 2$

對於 s 是任意奇數，任意 $M > 0$，存在一個只有一層以 tanh 為 activation function 的類神經網路，寬度為 $\frac{s+1}{2}$

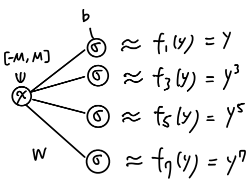
可以使得同時所有奇次多項式 $\{y^1, \dots, y^p\}$ ($p \le s$, p 是奇數) 與該類神經網路的 $W_{k,\infty}$ norm 誤差 $< \varepsilon$

根據定義 $\|f\|_{W^{k,\infty}(\Omega)} = \max_{0 \le m \le k} |f|_{W^{m,\infty}(\Omega)} = \max_{0 \le m \le k} \left(\max_{|\alpha|=m} \|D^\alpha f\|_{L^\infty(\Omega)} \right) = \max_{0 \le m \le k} \left(\max_{|\alpha|=m} \inf\{ C \ge 0 : |D^\alpha f| \le C \text{ for almost every } x \} \right)$

$\max_{\substack{p \le s \\ p \text{ odd}}} \| f_p - (\Psi_{s,\varepsilon})_{\frac{p+1}{2}} \|_{W^{k,\infty}} \le \varepsilon$ 的意思是把那些 $y^1, \dots, y^p$ 的多項式函數和類神經網路構造出來的函數，函數相減計算

最小上界，因為總共可以微分 $|\alpha| = m$ 次，把所有 $|\alpha| = 0, \dots, m$ 中上界最大的取出，再從 $y_1, \dots, y_p$ 這些剛剛計算得出的結果取

最大值，就得到 $W_{k,\infty}(\Omega)$ norm 了。



那作者是透過將 tanh 函數平移和係數相乘加總，

來去逼近 $x, x^3, x^5, \dots$ 奇次方的多項式行為。

在上課的時候有提到說，要如何設計一個 neural network 去呈現各種不同特性的函數，就現在來看一些複雜的系統可以用 neural network 來訓練得到不錯的結果，但是如果真的要將一些函數和其導函數透過 neural network 來進行函數合成搭配佩適當的係數就可以完成的話，那麼要使用什麼樣的函數來當作 neuron 呢？現在常常看到這一些 activation function 它們肯定是具備了某些條件而且效果不錯才被拿來使用，不然為什麼不依照傅立葉展開用 $\sin x$, $\cos x$ 來逼近不同的函數呢？

Lemma 2.

Let $k \in \mathbb{N}_0$, $s \in 2\mathbb{N}-1$ and $M > 0$. For every $\varepsilon > 0$, there exist a shallow tanh neuron network $\psi_{s,\varepsilon} : [-M, M] \longrightarrow \mathbb{R}^s$ of width $\frac{3(s+1)}{2}$ such that

$$\max_{p \le s} \| f_p - (\psi_{s,\varepsilon})_p \|_{W^{k,\infty}} \le \varepsilon.$$

Furthermore, the weight scale as $O\left(\varepsilon^{-s/2} \left(\sqrt{M}(s+2) \right)^{3s(s+3)/2} \right)$ for small ε and large s.

作者把偶次方寫成奇次方與較低偶次方的公式，如下 ($\alpha > 0$)：

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

透過遞迴關係式可以一直遞迴到剩下奇次方項的多項式，再根據 Lemma 3.1

就能讓偶次方與奇次方項的多項式與類神經網路所構成的函數誤差 $< \varepsilon$。

小例子 ($s = 3$)

Lemma 3.1：有一個寬度 $(3+1)/2 = 2$ 的淺層 tanh 網路 $\Psi_{3,\varepsilon}$，同時把 x, x^3
近似到 ε (含導數到第 k 階)

Lemma: 3.2：根據 Lemma 3.1，擴充成一個寬度 $3(3+1)/2 = 6$ 的淺層網路 $\psi_{3,\varepsilon}$，
可以同時把 x, x^2, x^3 都近似到 ε。這裡 x^2 是用上面的遞迴公式，從 x^3 與較低次組回來的。