Written Assignment 3

Problem 1

這兩個關鍵的引理，核心想法是：一個淺層的 tanh 神經網路可以用來逼近單項式 x^p，而且在 Sobolev 範數 $W^{k,\infty}$ 下的誤差可以被控制到任意小，連導數也能同時被逼近。

Lemma 3.1 – 奇數次冪

1.定理描述

考慮 $\tanh(x)$ 的展開式：

$$\tanh(x) = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - \dots$$

可以發現所有項都是奇數次冪，若對 $\tanh$ 做縮放和平移，再進行線性組合，就有辦法抽取出特定的奇數次項。

透過有限差分算子：

$$\delta_h^p[f](x) = \sum_{i=0}^{p}(-1)^i \binom{p}{i} f\Big(x + \Big(\tfrac{p}{2} - i\Big)h\Big),$$

我們有近似關係：

$$\delta_h^p[\tanh](x) \approx h^p \tanh^{(p)}(x).$$

在 $x = 0$ 時做歸一化，可以定義

$$\hat{f}_{p,h}(x) := \frac{\delta_{hx}^p}{\tanh^{(p)}(0)\, h^p}.$$

隨著 $h \to 0$，這個近似會收斂到 $f_p(x) = x^p$，而且在 $W^{k,\infty}$ 範數下也成立。因此，對所有奇數 $p \le s$，可以構造出一個寬度為 $\frac{s+1}{2}$ 的淺層網路 $\Psi_{s,\varepsilon}$，保證

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

換句話說，奇數次冪函數以及其導數都能被網路近似，誤差小於 ε。誤差估計大約是 $O(h^2)$，而權重的大小則需隨 $\varepsilon^{-s/2}$ 成長。

---

2. 誤差控制

對 $m \le k$ 的導數，有

$$\big\|\hat{f}_{p,h}^{(m)} - (x^p)^{(m)}\big\|_{L^\infty([-M,M])} \;\le\; C_{p,k,M}\, h^2.$$

因此在 Sobolev 範數下

$$\|\hat{f}_{p,h} - f_p\|_{W^{k,\infty}} \le C_{p,k,M}\, h^2.$$

選 $h = \sqrt{\varepsilon / C_{s,k,M}}$，即可保證所有奇數 $p \le s$ 的誤差 $\le \varepsilon$。

---

3. 寬度與權重

- **寬度**：由於可以共用同一組隱藏神經元（對稱性），總寬度為

$$\tfrac{s+1}{2}.$$

- **權重大小**：輸出層係數量級為

$$O(\varepsilon^{-s/2}\, \Phi(s, M)),$$

Lemma 3.2 – 偶數次冪,多項式

1. 定理描述

Lemma 3.1 已經處理了所有奇數次冪，現在要補上偶數次冪。由於 $\tanh$ 是奇函數，展開式中沒有偶數項，因此需要使用代數恆等式：

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

這個公式表示：偶數次冪可以由「兩個奇數次冪的差」再扣掉一些低階偶數冪來得到。因此，如果奇數次冪已經能夠近似，就可以用遞迴的方式構造出偶數次冪。

令 $\psi_{s,\varepsilon}$ 為這個構造出的網路，則可以同時逼近所有 $f_p(x) = x^p$（$p \le s$），並保證

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

---

2. 誤差控制

定義誤差

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

- 當 p 為奇數時，誤差界由 Lemma 3.1 直接給出：
  $E_p \le \varepsilon$。

- 當 $p = 2n$ 為偶數時，利用上式中的遞迴公式，誤差來自兩部分：

  1. 由奇數次冪的近似誤差傳遞而來；
  2. 由低階偶數次冪誤差累積而來。

透過數學歸納法可得：對所有 $p \le s$，都有

$$E_p \le C_{s,k,M}\,\varepsilon,$$

其中常數 $C_{s,k,M}$ 只依賴於最大次數 s、導數階數 k 與區間大小 M。因此選擇合適的 α（例如 $\alpha = 1/s$），即可確保所有冪次的誤差 $\le \varepsilon$。

---

3. 寬度與權重

- **寬度**：因為每個偶數次冪需要額外的奇數次冪組合，總寬度比 Lemma 3.1 大，為

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

- **權重大小**：偶數次冪的構造中，每一步遞迴都會引入額外的係數，大小大約隨 h^{-p} 成長。和 Lemma 3.1 相同，選取 $h \sim \varepsilon^{1/2}$，可得權重階為

$$O(\varepsilon^{-s/2}\,\Phi(s,M)),$$

參考文獻的證明:lemma3.1(18-25) lemma3.2(27-37)，並有使用gpt修飾並整理過

Problem 2

雖證明是針對 tanh 的，那如果換成 ReLU 或 sigmoid，是否還能有類似的結果？