

機器學習 HW3 - Written assignment

(一) Reading and Explaining Lemmas

Lemma 3.1: Let $k \in \mathbb{N}_0$ and $s \in \mathbb{N} \setminus 1$. Then it holds that for all $\varepsilon > 0$ there exists

a shallow tanh neural network $\underline{\Psi}_{s,\varepsilon} : [-1, 1] \rightarrow \mathbb{R}^{\frac{s+1}{2}}$

of width $\frac{s+1}{2}$ s.t. $\max_{\substack{p \leq s \\ p \text{ odd}}} \|f_p - (\underline{\Psi}_{s,\varepsilon})^{\otimes k}\|_{W^{k,\infty}} \leq \varepsilon$

Moreover, the weights of $\underline{\Psi}_{s,\varepsilon}$ scale as

$O(\varepsilon^{\frac{s}{2}} (\frac{1}{2(s+1)\sqrt{s\mu}})^{s(s+3)})$ for small ε and large s .

Lemma 3.1 說明了，用一個單隱藏層的神經網路 $\underline{\Psi}_\varepsilon$ (啟動函數用 tanh)

就能近似所有奇數單項式 (x^p , p 是奇數)。

函數 $\underline{\Psi}_{s,\varepsilon} : [-1, 1] \rightarrow \mathbb{R}^{\frac{s+1}{2}}$ 意思是在區間 -1 到 1 之間，若最大

次數為 s ，則有 $\frac{s+1}{2}$ 個單項式 ($x^1, x^3, x^5, \dots, x^s$)。

由於神經網路是單層且以 tanh 做為啟動函數，故可表示為

$$g(x) = \sum_{i=1}^N a_i \tanh(b_i x + c_i) + d$$

而 $\max_{\substack{p \leq s \\ p \text{ odd}}} \|f_p - (\underline{\Psi}_{s,\varepsilon})^{\otimes k}\|_{W^{k,\infty}} \leq \varepsilon$ 則是說在小於 s 的奇數單項式 f_p ，

與神經網路預測的誤差皆可控制在 ε 之內，而 k 則表示，

對於每個 f_p 的 m 階導數， $m \leq k$ ，都有 $g_p^{(m)}(x)$ 使兩者誤差小於 ε

最後一行的 0 則是在說. 神經網路裡的權重大小與

$$\varepsilon^{-\frac{s}{2}} (\sqrt{s+2}) \sqrt{\mu} \gamma^{(s+3)} \text{ 這個成 } \text{FCC}.$$

$\Rightarrow \varepsilon$ 取的越小. or 要求的奇單項式越多. 則權重越大

Lemma 3.2: Let $k \in \mathbb{N}_0$, $s \in \mathbb{N}-1$ and $\mu > 0$. For every $\varepsilon > 0$, there exists a shallow

tanh neural network $\psi_{s,\varepsilon}: [-\mu, \mu] \rightarrow \mathbb{R}^s$ of width $\frac{3(s+1)}{2}$ s.t.

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

Furthermore, the weights scale as $O(\varepsilon^{-\frac{s}{2}} (\sqrt{\mu(s+2)})^{3s \cdot \frac{s+3}{2}})$ for small ε and large s .

與 3.1 略有不同的是. Lemma 3.2 保證了要同時輸出 x, x', \dots, x^s

所須的隱藏層神經元數量為 $\frac{3(s+1)}{2}$, 其他和 Lemma 3.1 一樣.

(=) Unanswered Question

1. 資料總數和訓練輪數何者對訓練較有幫助?

2. 是否存在一函數可描述資料量、重複次數等參數

對訓練出的神經網路與目標的誤差的影響?