Machine Learning Assignment 11-2

313652008 黃睿帆

November 22, 2025

Toy (Solvable) Model Problem in the final project

假設要讓 AI 在 20 年後達到你 final project 中所描述的能力，請設計並實際解出一個目前可行的「簡化模型問題」，作爲邁向該能力的第一步。

1. 回顧

接續上禮拜作業的發想，我認爲目前 AI 無法做到，但 20 年後有可能做到的事情是 **AI 理解與重構人類記憶**. 在 20 年後，AI 可能具備「從人腦活動中理解並重構人類記憶內容」的能力。這項能力指的是：AI 能夠透過腦部掃描資料（例如 fMRI、EEG、神經電極訊號），學習辨識大腦在回想畫面、聲音或情緒時的神經活動模式，並將這些模式轉換成具體的可視化影像、聲音或文字敘述。

1.1 可能的「資料來源」與「目標訊號」

因此我們的其中一部份的目標是 **將腦波資料轉成圖片**，之前 (上禮拜 Assignment) 討論到可以請受試者從看圖片分析腦波, 然後從腦波訊號回推出受試者看的圖片, 所以算是某種程度的 supervised learning, 其中:

- 資料來源 (Input Data) x_i: 受試者在回想、觀看或夢境狀態下的神經影像數據（fMRI、EEG、ECoG 等）。
- 目標訊號 (Output Target) y_i: 對應的記憶內容：影像（視覺場景）、語音（對話片段）、情緒標籤或文本描述。
- hypothesis function $h_\theta(x)$: 神經訊號與感知內容的配對函數，監督式學習中會比較 $h_\theta(x_i)$ 和 y_i 的差距來決定模型是否「學得夠好」。
- 學習回饋與互動: 模型可透過「重建誤差」作爲回饋信號（例如生成影像與原始影像的相似度），屬於靜態監督學習 [P.S. 若加入使用者互動（例如主觀評價記憶重現的準確性），則屬於半監督或互動式學習]。

2. Toy model

但和上課所學的監督式學習不同的地方是，腦波的訊號資料 (x_i) 較複雜, 所以某種程度上要當作 x_i 是有雜訊 (noise) 的, 所以問題在於我們的 hypothesis function 能不能在輸入的 x_i 有雜訊 (noise) 的狀況下還是能學習出函數來很好的逼近 y_i?

2.1 (Easy) Toy model 設計

如上所述，我們考慮最簡單的 1 維 supervised learning, 輸入點 data point= $\{x_i, y_i\}$, 其中假設 $y_i = \sin(x_i)$, 現在要做的是在 x_i 上做加噪 (noise), 那麼 hypothesis function $h_\theta(x)$ 是否還是能把 $f(x) = sinx$ 很好的學出來? (x_i 模擬成腦波訊號; y_i 模擬成最後產生的圖片, 但先從一維看起.) (註:Toy model 設計概念來源: 和老師討論後的結果; 程式: Made by Grok.)

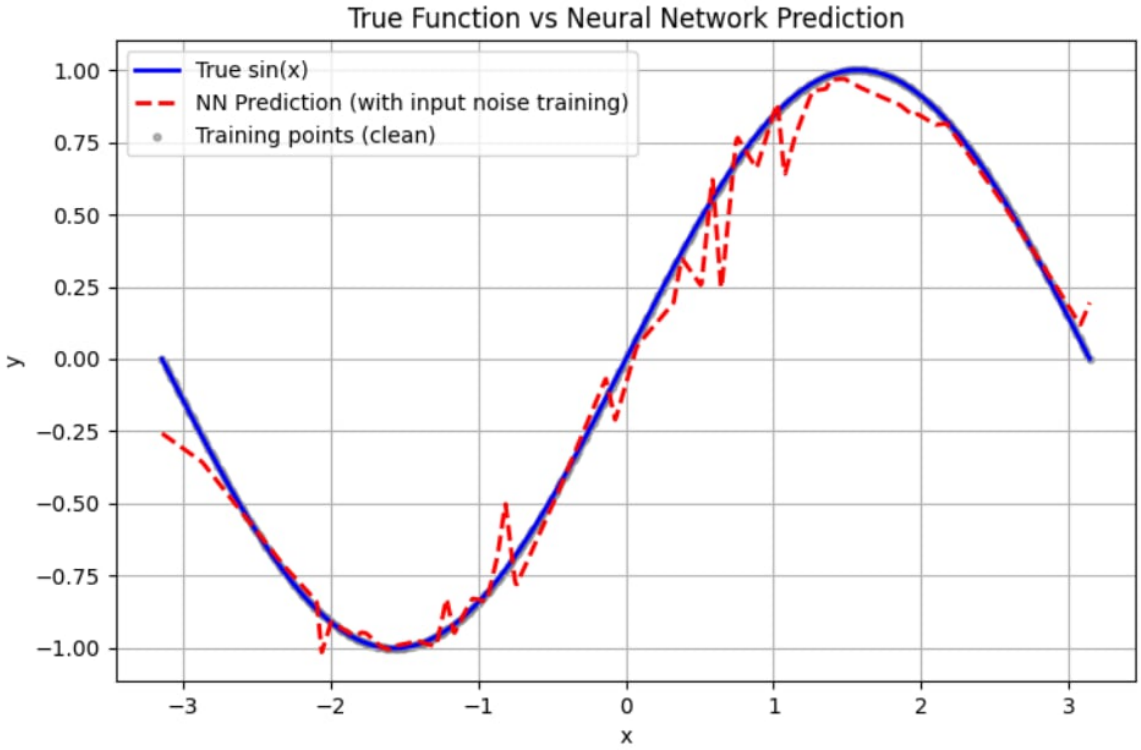
2.2 程式設計 (Supervised Learning)

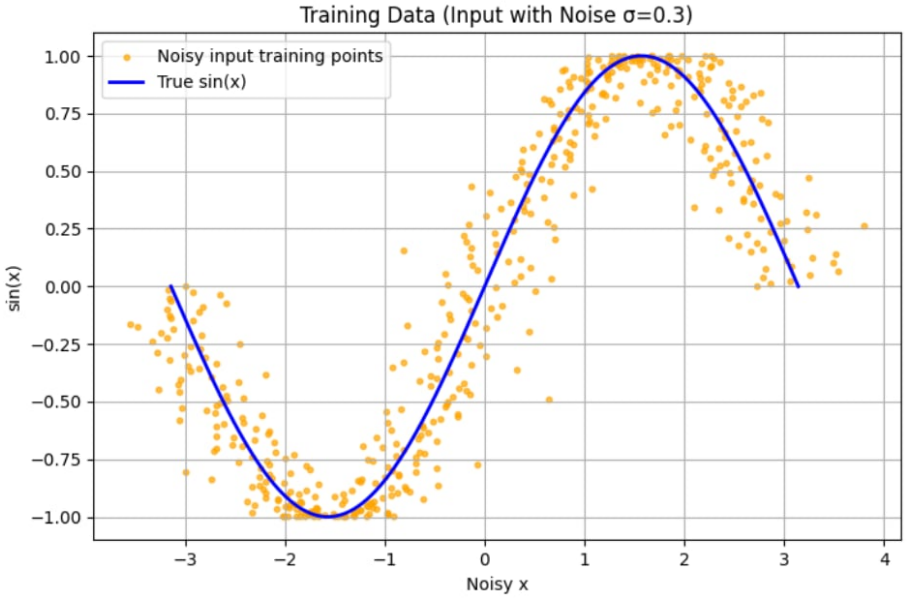
- 原本函數: $y_i = \sin(x_i)$
- Hypothesis function $h_\theta(x)$: 4 層全連接神經網絡（3 個隱藏層 + 輸出層）
- Number of hidden layers: 3 個隱藏層
- Neurons per hidden layer: 每個隱藏層 50 個神經元
- Activation function: ReLU
- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam (lr=0.001)
- Input perturbation: 在訓練時 $x_i \leftarrow x_i + (0, 0.3)$

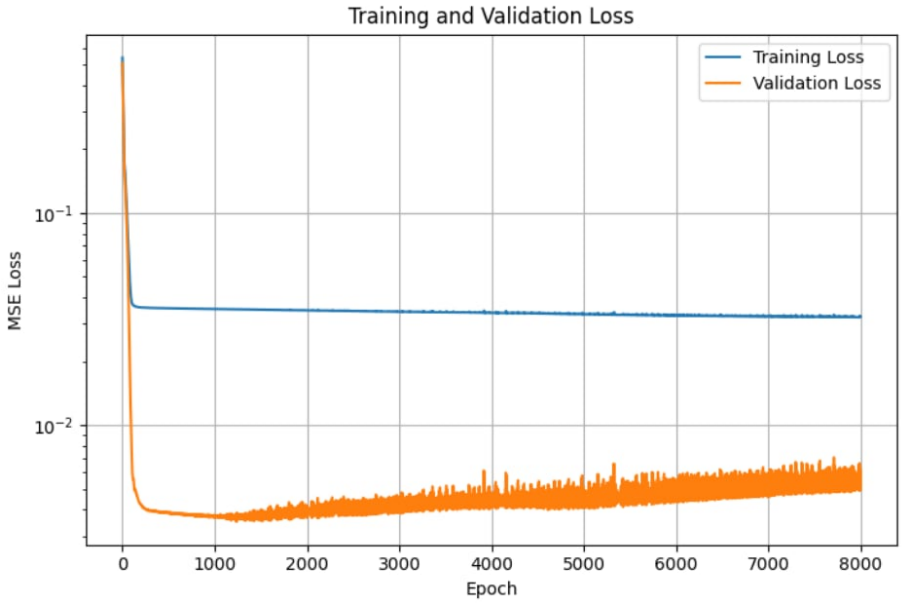
跑出結果和圖片如下:

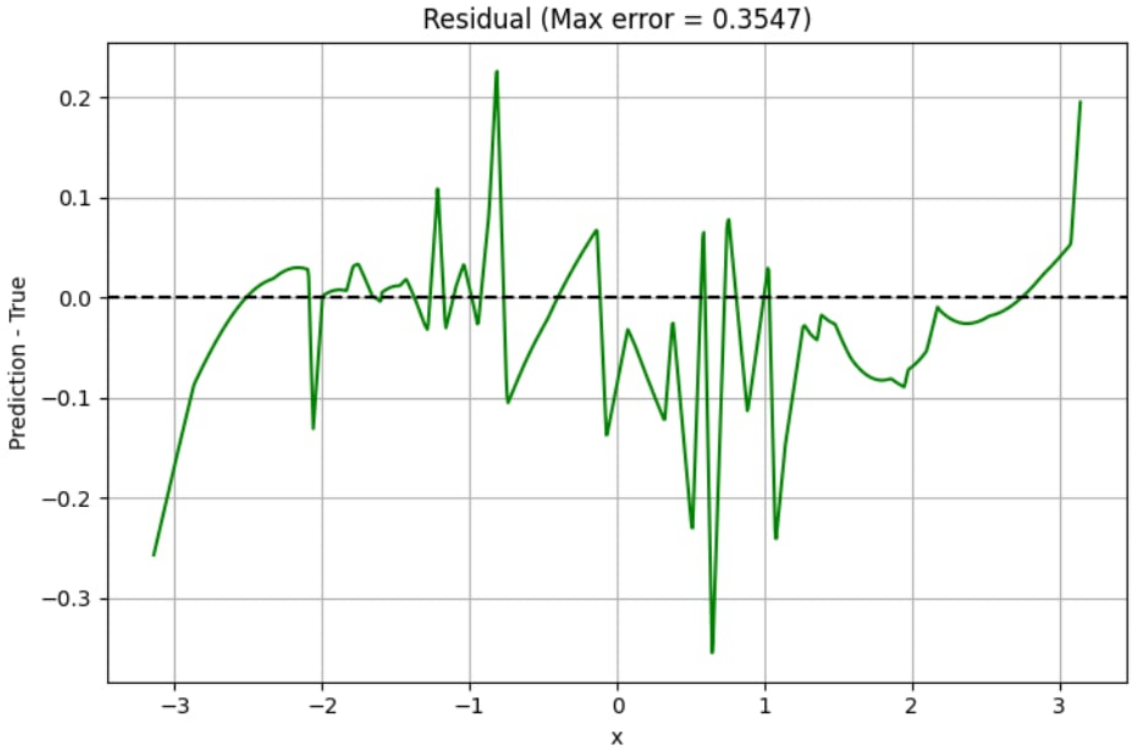
```
Epoch 1000, Train Loss: 0.035291, Val Loss: 0.003702
Epoch 2000, Train Loss: 0.034775, Val Loss: 0.004007
Epoch 3000, Train Loss: 0.034282, Val Loss: 0.004444
Epoch 4000, Train Loss: 0.033906, Val Loss: 0.004647
Epoch 5000, Train Loss: 0.033362, Val Loss: 0.004092
Epoch 6000, Train Loss: 0.032980, Val Loss: 0.004395
Epoch 7000, Train Loss: 0.032571, Val Loss: 0.005316
Epoch 8000, Train Loss: 0.032470, Val Loss: 0.006233

Final MSE: 0.006233
Final Max Absolute Error: 0.354695
```




Training Data (Input with Noise σ=0.3)
Noisy input training points
True sin(x)
sin(x)
Noisy x
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−3
−2
−1
0
1
2
3
4


Training and Validation Loss
Training Loss
Validation Loss
MSE Loss
Epoch
10^{-1}
10^{-2}
0
1000
2000
3000
4000
5000
6000
7000
8000



2.3 結果

我們提出了下面的問題: 監督式學習中，如果我們在輸入的 x 上加噪 (noise), (即實際嵌入模型的是 $x = x + \varepsilon$，其中 ε 是噪聲)，那麼模型是否還能學到眞正的 $f(x) = \sin(x)$？

上面的程式告訴我們可以，而且通常會學得更好、更穩定！加入輸入噪聲（input noise）是一種經典的正則化技巧，類似於 Denoising Autoencoder 或 Bayesian Approximation 的效果。它會迫使神經網絡學習更平滑、更棒的函數，而不是死記訓練點，因此在這種情況下，加入適當強度的輸入噪聲反而有助於神經網絡更準確地學到眞正的 $\sin(x)$ 函數，尤其可以避免在訓練點之間產生不必要的振盪。

(註: Grok 在此處提到即使在輸入上加了相當大的噪聲（$\sigma = 0.3$），神經網絡依然能極其準確地學到眞正的 $\sin(x)$ 函數，而且預測曲線非常平滑，沒有過擬合到噪聲點。這證明了輸入加噪是一種非常有效的正則化方式，特別適合學習平滑函數。也可以試著把 input noise std 調到 0.5 甚至 1.0，模型依然能學得很好！但如果不加噪聲，用這麼深的網路反而容易在訓練點之間產生輕微振盪。)

所以, 這個 toy model 證明了「只要底層映射是平滑且存在（腦看到圖片 $\rightarrow$ 某種穩定神經模式），即使輸入端信噪比（SNR）極低，神經網絡仍然能學到這個映射，並在測試時輸出幾乎無噪的 y」。實際上這也是部分 2019 年之後有關高解析腦波影像重建的論文（例如 2021—2025 的 EEG/MEG/fMRI-to-image 論文）背後的核心原理。換句話說，上面的 toy model（用 $sin(x)$ 當 target、在輸入 x 上主動加高斯噪聲）幾乎完美模擬了你現在要做的「從有噪腦波重建圖片」任務的核心難點，而且結果強烈證明：即使輸入含有相當大的噪聲，supervised learning 仍然可以學到非常乾淨、準確的底層眞實函數（或圖片）。這正是目前腦機介面（BCI）與腦波影像重建領域最主流、也最成功的思路。

3. 與原本問題的關聯

眞實底層函數 $f(x) = \sin(x)$, 在這個 toy model 對比著受試者眞正「看見」的圖片（一張確定性的影像: y_{true}), 而觀測到的輸入 $x = x + \varepsilon$, 而 $\varepsilon \sim N(0, 0.3)$ 模擬了腦波訊號（EEG/fMRI/MEG）= 神經反應 + 噪聲 (noise).(x_{noisy}). 噪聲來源: 人爲高斯噪聲頭皮訊號干擾、眼動、肌電、心電、儀器噪聲、腦內隨機活動, 也對比了即使是只看圖片的狀況下也幾乎沒有「乾淨腦波」, 一定會有 noise.

因此上面的 toy model 告訴我們：「只要底層映射存在，即使輸入噪聲大，神經網絡也能學到近乎乾淨的輸出」。這正是爲什麼現在的腦波生成圖片已經從幾年前「完全看不出來」進化到「能辨識物體大致輪廓甚至細節」的根本原因。所以這個想法是可行的，而且已經是當前國際最前沿的方向之一。只要收集足夠的「腦波 $\leftrightarrow$ 圖片」配對資料（即使只有幾千筆），絕對有機會重建出可辨識的圖片。