Machine Learning Assignment 11-2

313652008 黃睿帆

November 30, 2025

1. AI 二十年後的未來能力：理解與重構「人類記憶」

我認爲在未來二十年間，AI 可能發展出一項突破性的能力 「從人腦活動中理解並重構人類的記憶內容」。所謂的重構，指的是 AI 能從腦部掃描資料（例如 fMRI、EEG、神經電極訊號）中，解讀人在回想畫面、聲音、語言或情緒時的大腦活動模式，並進一步將這些模式轉換成具體影像、聲音，甚至完整的文字敘述。

我的靈感其實源自多年前的一部日劇《神探伽利略》第二季。劇中提到一個高維度矩陣研究中心，能將人腦中的影像「映像化」來協助破案。雖然最後揭露這只是湯川教授爲了逼迫眞兇露出破綻所編造的故事，但它在當時給我留下深刻印象。畢竟在 1990 年代，電腦剛剛能做出蒙太奇效果，距離解析大腦影像遙不可及；而三十多年後的今天，AI 已展現出驚人的推理與生成能力（前幾天聽說 google 的 gemini 3.0 已經能模仿人的字跡並解數學題目的過程）。這讓我開始思考——或許在未來，這樣的「記憶映像化」不再只是虛構。換句話說：我認爲在未來 **AI 能把「腦活動模式」映射回「記憶的內容」本身。**

這種從腦訊號理解並重構記憶的技術，將帶來多面向的應用，例如：

- **醫療復健：**協助失智症、腦外傷患者找回破碎的記憶，或幫助 PTSD 患者安全地「重整」創傷場景，提升治療效果。
- **司法鑑識：**在嚴格倫理審查與個人同意下，還原目擊者或被害人的事件記憶，提高證詞可靠度，協助釐清眞相。
- **人機介面：**讓人類以「想像」直接與 AI 溝通，例如只需在腦中描繪畫面，就能控制電腦、生成影像或創作虛擬世界。
- **文化與歷史保存：**未來或許能「記錄並分享」個人的主觀記憶，形成全新的文化檔案與生命史敘述方式。

這項能力的重要性在於：它能突破語言與表達的限制，使我們更接近理解「人類思考與記憶如何運作」。然而，它也將重新定義個人隱私與意識邊界，帶來前所未有的倫理挑戰。因此，在追求技術突破的同時，更需要謹慎、負責與成熟的制度來引導其應用。

2. 所需的成分與資源（Ingredients）

如果要讓 20 年後的 AI 實現上面描述的能力，也就是 AI 能做到：從人腦活動訊號（fMRI、EEG、ECoG、深部電極等）中解析出一個人「正在回想的內容」，並以明確方式「重建」出來。具體來說，我認爲所需要用到的「成分（ingredients）」有：(P.S. 參考資料來源包含 Chat-GPT 與 Wikipedia)

2.1 資料 (Data)

首先，資料包含了多模態的「腦活動－內容」對照資料，因爲我們要讓 AI 學習「腦活動 $\rightarrow$ 記憶內容」之間的映射，因此也需要大量的 paired datasets：例如腦部訊號資料（fMRI / EEG / ECoG / MEG），包括高時間解析（例如 EEG、ECoG），高空間解析（例如 fMRI），也可能需要記錄：EEG + fMRI 的同步資料，這些資料的主要作用在於 **提供模型學習人腦不同區域在記憶、視覺化、語言回想、情緒喚起時的活動模式。**

再者，資料也包含其對應的外部內容（Ground Truth Memory Content），依想達成的任務不同，可以是：影像 (讓受試者觀看或回想某張照片)、聲音 (聽某段音樂或在腦中回想旋律)、語言 (回想某段文

字、情節)、情緒 (回想特定事件並標記情緒與強度)、事件記憶 (episodic memory) (訪談作爲標註),這些資料主要作用在於提供**監督訊號 (supervision),使模型能學習「某個記憶下的大腦活動應該對應什麼外部內容」。**

還有記憶相關的輔助資料,包含大規模的大腦結構資料 (connectome, wiring diagram),描述了人腦不同區域間的連結,還有從 connectomics 得到的神經路徑,主要的作用在於**幫助模型理解「記憶如何在結構上被存取」。**以及語意資料庫 (semantic datasets),主要是爲了讓 AI 重建語言或敘事型記憶,此部分可能額外需要:大型語言模型,結構化人類語意概念資料 (WordNet、ConceptNet),才能達成 **將抽象記憶 (如「一場吵架」) 編碼成語意描述。**

資訊也包含時態資料 (Temporal Data),由於記憶是動態,因此需要時間序列的腦活動訊號,事件回想時的腦區活化變化,紀錄記憶強度、熟悉度、情緒強度的變化等,達到**讓 AI 捕捉「記憶回想的時間結構」,而不是只看靜態瞬間**的作用。

備註: 多模態學習是一種深度學習方法。它能整合和處理多種類型的資料,例如文字、音訊、圖像或影片。這些不同類型的資料叫做模態。這種整合能夠更全面地理解複雜資料,(在我們的例子中,就是讀取到的腦波資料 (即腦活動內容)),從而提高模型在視覺問答、跨模態檢索、文字到圖像生成、美學排名和圖像字幕等任務中的效能。(資料來源: 維基百科)

2.2 工具 (Tools)

AI 要處理訊號並理解人腦記憶必須使用多種工具組合。包含以下最核心的幾項工具:

1. 高維訊號處理與數學工具–PDE 或是動態系統 (Dynamic Systems): 用於處理模型化腦部電訊號演化以及模仿神經活動的 propagation,幫助理解記憶如何被從腦區 A「喚醒」至腦區 B。**主要用途爲建立腦活動的時間動態模型。**
2. 高維訊號處理與數學工具–多變量統計與矩陣分解 (PCA / ICA / CCA): 用於在 fMRI 裡分離不同認知模式,來找出腦訊號的共同群組 (functional networks) **主要用途爲第一層表示 (representation) 提取。**
3. 深度學習工具–Graph Neural Networks (GNN): 大腦某種程度上算是一個圖 (graph) 結構 (neurons → nodes, synapses → edges)。**主要用途在於對「大腦結構連結」建模,將不同腦區活動整合成「整體記憶表徵」**
4. 深度學習工具–Transformer / Multimodal Encoder: 需要一個模型能同時吃入 fMRI voxel 時序、EEG 時序、語意或視覺資料、記憶強度等 metadata. **主要用途是將多模態腦訊號轉成統一的 latent space.**
5. 深度學習工具–生成模型 (Diffusion model / VAE / GAN / LLM): 負責將腦訊號轉成:視覺影像 (Diffusion)、聲音 (Audio diffusion / codec models)、敘述 (LLM)、多模態記憶 (multimodal reconstruction) 等資料,**主要用途爲最終步驟「重建記憶內容」。**
6. 解釋工具 (Explainable AI)–Causal Inference (因果模型): 讓模型不只知道「會出現記憶 A」,而是知道:爲何這段記憶被喚起、哪些腦區「造成」記憶重建,**主要用途爲建立可解釋、可追蹤的記憶重構。**

2.3 硬體與環境 (Hardware / Environment)

要讓 AI 要處理腦波訊號並理解人腦記憶,所使用到的設備包括:

- 感測設備: 要取得高品質,無雜訊的腦波訊號,可能需要更高解析度的 fMRI (例如 <0.5 mm voxel) (註: 現代 fMRI 還太粗,20 年後可能能以近神經元解析度成圖像等)
- 感測設備: 非侵入式高速 EEG (高密度 > 2000 channel) 目前已逐漸往高密度方向發展,可以捕捉 millisecond-level 訊號
- 感測設備: 新型腦機介面感測器。例如:納米級光學感測,含石墨烯的非侵入式腦電偵測,超低雜訊磁場量測 (MEG-like but wearable) **感測設備主要作用在於提供精準、穩定的時間與空間訊號給模型。**
- 運算硬體: 包含大規模 GPU / TPU 集群,量子運算,記憶壓縮與神經符號模型的混合架構等等

- 標記環境（Annotation Environment）: 未來可能需要受試者描述記憶的交互介面、心理學家協助標記情緒、語意、事件結構、眞人回想的行爲紀錄（語音、眼動、反應時間）

2.4 學習架構（Learning Setup）

要讓 AI 要處理大量腦波訊號並理解人腦記憶，學習架構需要結合以下的各種不同學習範式:

1. Self-supervised Learning（自監督）: 因爲腦訊號資料很難完全標註，所以用對比學習（contrastive learning）並使用 fMRI/EEG 互相預測，從腦區 A 預測腦區 B（cross-brain prediction）**主要作用在標註有限時，仍能學到腦訊號結構。**

2. Supervised Learning（監督式學習）: 也是必要的，用於 fMRI 和圖像、EEG 和聲音、fMRI 和敘述等轉換。**主要作用是讓模型學到「記憶類型、內容、語意」的明確 mapping。**

3. Multitask Learning（多任務）: 因爲人類記憶涉及視覺、聽覺、語言、情緒、時序事件等。**主要作用在於讓系統學到不同模態的共通 representation.**

4. RLHF（或人類回饋）: 可能用於受試者評價重建記憶的準確度，根據回饋來讓 AI 改良 reconstruction. **主要作用是去微調生成模型，使結果更接近主觀記憶。**

5. Meta-learning: 因爲不同人腦的神經排列不同，此技術能讓模型快速適應新個體（few-shot brain alignment），快速適應個人差異。

2.5 小結

總結，要讓 AI 學會處理大量腦波訊號並理解人腦記憶，是一個很複雜的課題，目前大致會用到的整體架構如下:

- 硬體決定你能收集到什麼「腦訊號形態」。
- 資料決定 AI 能學習哪些 mapping（腦 → 記憶內容）。
- 工具 (深度學習 + 數學) 負責建構高效的「腦表徵模型」與「內容生成模型」。
- 學習架構決定整體如何組合，包括表示、重建、個體化調適與人類回饋。

註: **特徵學習（feature learning）或表徵學習（representation learning）**是學習一個特徵的技術的集合：將原始資料轉換成爲能夠被機器學習來有效開發的一種形式。它避免了手動擷取特徵的麻煩，允許電腦學習使用特徵的同時，也學習如何擷取特徵：學習如何學習。(資料來源: 維基百科)

上述過程形成一個清楚的 pipeline：感測器 → 腦訊號 → 多模態 encoder → latent representation → 生成模型 → 記憶重建 → 人類評價 → 更新模型。

3. 涉及的機器學習類型

如上所述，讓 AI 去處理腦波訊號並理解人腦記憶的技術，所使用的主要機器學習方法可能包含：監督式學習 (Supervised Learning)、生成式模型 (Generative Models)（如 Diffusion Models、Variational Autoencoders、GANs）、多模態學習 (Multimodal Learning)、表徵學習 (Representation Learning) 等，其中:

- 監督式學習：要讓 AI 學會「神經活動模式」和「視覺或語意內容」的對應關係，必須以已知配對資料訓練，像是 (1) 輸入資料 (人在觀看或想像特定影像時的大腦神經訊號)；(2) 標籤資料：對應的實際影像內容。透過監督學習，AI 能逐漸掌握不同腦區活動與記憶內容之間的對應特徵。
- 生成式模型：因爲人腦的記憶重構不是單純分類，而是「從腦部訊號生成視覺或語音內容」。所以生成模型（如 diffusion model 或 VAE）或許能將潛在的腦部特徵向量轉換成影像或聲音，重建人類曾見或想像過的場景。
- 多模態學習：記憶往往包含多種感官成分（影像、聲音、情緒）。所以 AI 必須學會在不同感官模態間對齊與整合資訊，建立統一的「記憶表徵空間」。

- 表徵學習：同時，爲了讓 AI 能理解「記憶」這種高維、抽象概念，它需透過表徵學習在神經信號與語意之間建立橋樑，學習穩定的潛在特徵（latent representations）。

而在「AI 重構人類記憶」的能力中，最核心的技術主要屬於監督式學習（Supervised Learning），並輔以非監督式學習（Unsupervised Learning）的方法。因爲要使用**監督式學習**讓模型學會「腦部活動訊號 $\rightarrow$ 外部可觀察內容」的對應關係，例如：fMRI/EEG/神經電極訊號（輸入項），受試者看到的影像、聽到的聲音、文字描述（是標記的 target），因爲資料是成對的（大腦活動 vs. 其對應的外部刺激），所以非常適合以監督式學習訓練。模型透過大量配對資料學習如何從神經訊號預測影像、聲音或內容。**資料來源是腦部神經訊號（fMRI/EEG/神經電極）；目標訊號是外界刺激或受試者描述之內容（影像、語音、文字）。**

而非監督式學習作爲輔助角色，是因爲記憶內容高度複雜且低維度標記稀缺，因此需要非監督式方法來壓縮、抽取大腦活動的潛在表示（latent representation），從大量未標記的神經資料中找出結構，並和生成模型（如 VAE、Diffusion models）結合重建畫面與聲音，也就是說，非監督學習能讓系統建立「大腦訊號空間」與「記憶表徵空間」的共同結構。因此在「AI 重構人類記憶」的能力中，**監督式學習用來建立腦活動與外界刺激的對應，並使用非監督式或生成式模型搭配，來捕捉腦訊號的隱含結構並生成重建內容。**

4. 第一步的「可實作模型問題」(Toy Model/ Solvable Model Problem)

我們的研究目標之一，是**將腦波或神經訊號資料轉換爲影像**。在先前（上週的 Assignment 11）中，我們討論過一種基本策略：請受試者觀看圖像、記憶圖像，或在想像／夢境狀態下產生內在影像，並同步記錄神經訊號。接著利用這些資料訓練一個模型，使其能從腦波訊號反推出受試者所見或所想的內容。這本質上是一種監督式學習（supervised learning），其中：

- **輸入資料** x_i：受試者在觀察、回想或想像場景時所產生的神經影像數據，例如 fMRI、EEG 或 ECoG。
- **目標訊號** y_i：與神經活動對應的外在或主觀內容，如視覺圖像、語音片段、情緒標籤或文字描述。
- **Hypothesis function** $h_\theta(x)$：一個將神經訊號映射到感知內容的函數。在監督式學習中，我們透過比較 $h_\theta(x_i)$ 與 y_i 的差距來判斷模型是否成功學習兩者的對應關係。
- **學習回饋方式**：最常見的是透過「重建誤差」作爲回饋，例如重建影像與目標影像的相似度。這屬於靜態監督式學習；若加入主觀評分或使用者互動，則可延伸爲半監督學習或互動式學習。

4.1 Toy Model：簡化問題

然而，與課堂上的傳統監督式學習不同，眞實的神經訊號 x_i 高度複雜且具有大量雜訊。因此，我們特別關注一個核心問題：(註:Toy model 設計，概念來源: 和老師討論後的結果; 程式: Made by Grok)

> **在輸入包含雜訊的情況下，神經網絡模型是否仍能學到正確的映射關係？**

爲了先在極簡化的環境中驗證這個問題，我們先設計一個一維的 Toy Model：假設資料點爲 x_i, y_i，其中 $y_i = \sin(x_i)$，並在訓練時對 x_i 加入隨機雜訊 $x_i \leftarrow x_i + \mathcal{N}(0, 0.3)$。這裡 x_i 模擬腦波訊號，y_i 模擬最終的重建影像（但先從一維函數開始）。

4.2 Toy Model 的模型架構

Toy example 採用標準的監督式神經網絡設定：

- **目標函數**： $y_i = \sin(x_i)$
- **Hypothesis function**： 4 層全連接神經網絡（3 個隱藏層 + 1 輸出層）
- **隱藏層**： 3 層，每層 50 個神經元
- **Activation function**： ReLU
- **Loss function**： MSE（Mean Squared Error）

- **Optimizer：** Adam（learning rate = 0.001）
- **Input noise：**訓練時加入 $x_i \leftarrow x_i + \mathcal{N}(0, 0.3)$

這樣的簡化模型可以檢驗：即使輸入具有雜訊，模型是否仍能學出一個足夠平滑且接近眞實函數 $f(x) = \sin(x)$ 的映射。這將爲後續更高維度的腦波 $\rightarrow$ 影像重建模型奠定直覺與實驗基礎。

4.2.1 問題設計（Problem Formulation）

如上所述，爲了讓 AI 達成「從人腦活動中理解並重構人類的記憶內容」的能力，我們需要從腦部訊號（如 EEG）中重建人類正在回想的影像內容。然而眞正的腦訊號極度高維且含有大量雜訊，因此第一步需要設計一個可控、可求解、可驗證的簡化模型問題（Toy Model）。其中 Toy Model 的概念對應:

- 腦波訊號（EEG）$\rightarrow$ 加入雜訊的 x 值
- 人類記憶的眞實內容（圖片）$\rightarrow$ 乾淨且可解析的目標函數 $y = \sin x$
- 影像重建模型 $\rightarrow$ 神經網路 $h_\theta(\cdot)$

因此，我們的簡化版本是：給模型觀察到的 noisy input: $x_i + \varepsilon$，是否仍能學習並恢復 underlying function $y_i = \sin(x_i)$？這對應到未來記憶重建的核心挑戰：**Input noisy 在 EEG 含大量非訊息噪聲，而 True signal smooth but unknown（記憶內容）的情況下，模型是否能在高雜訊下仍重建有意義的資訊？**

換句話說，我們明確定義輸入（input）：$\tilde{x}_i = x_i + \varepsilon_i, \text{where } \varepsilon_i \sim \mathcal{N}(0, 0.3)$，和輸出 (target) $y_i = \sin(x_i)$.，任務目標是訓練模型 $h_\theta(\tilde{x}_i)$ 使其能逼近眞實的 $\sin x$，即: $h_\theta(\tilde{x}) \approx \sin x$ ，這測試了：模型是否具備「抽取訊號、忽略噪訊」能力（等價於去雜訊能力）以及模型是否能從模糊輸入中重建潛在語義（類似未來從腦波推影像）。

4.2.2 模型與方法（Model & Method）

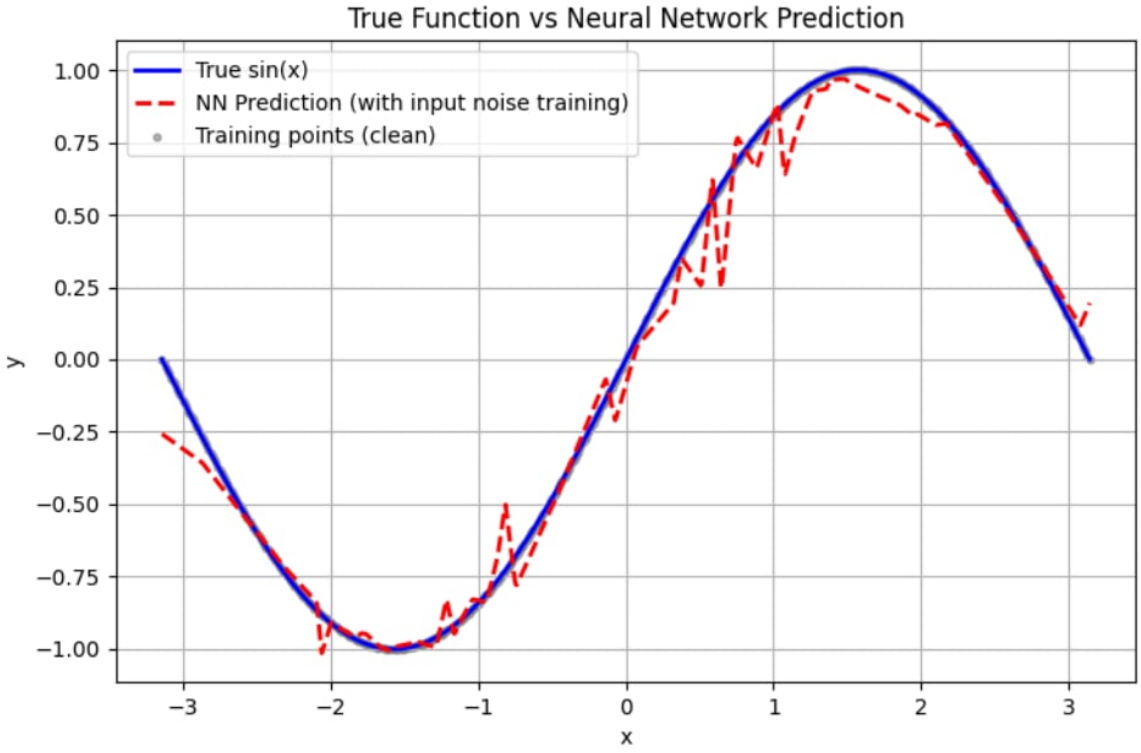
我們在 Supervised Learning 的框架下，來做簡化後的模型。模型架構 (如上所述)，包括 3 個隱藏層，每層 50 個神經元，Activation：ReLU(ReLU 架構簡單，訓練穩定，能觀察噪聲與非線性回歸的基本行爲)，Loss：Mean Squared Error，Optimizer：Adam (lr = 0.001)。這足以測試**從 noisy input 推回原訊號**，這與未來從 noisy EEG decoding 出 mental image 的任務本質上是類似地。這個架構簡單，重點在於能把大的問題化爲一個可操作的數學學習任務，並用這學期機器學習的課上所學就能親手驗證，但仍能清楚反映未來大問題的核心難題：**在存在雜訊的前提下，模型是否能恢復潛在的有意義訊號？**

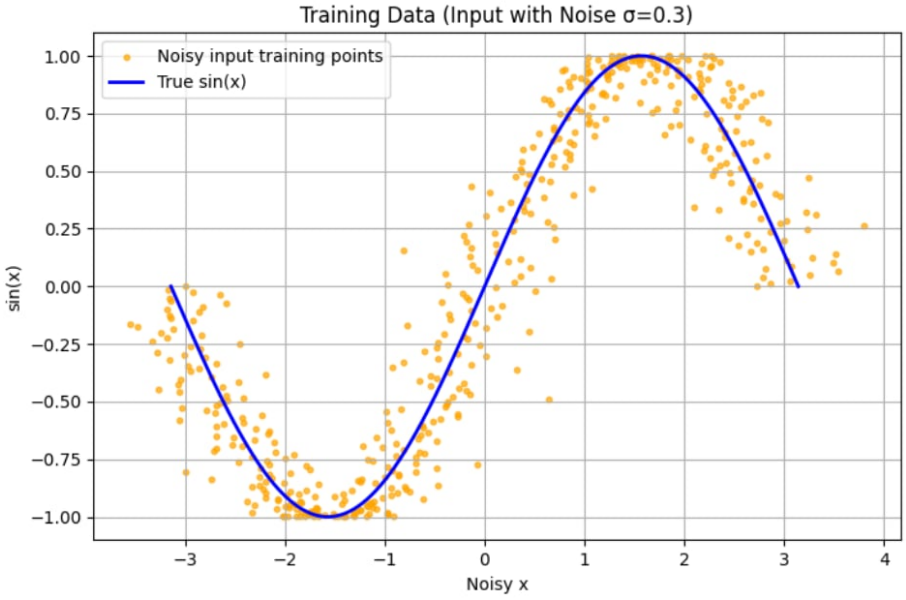
4.2.3 實作與結果（Implementation & Results）

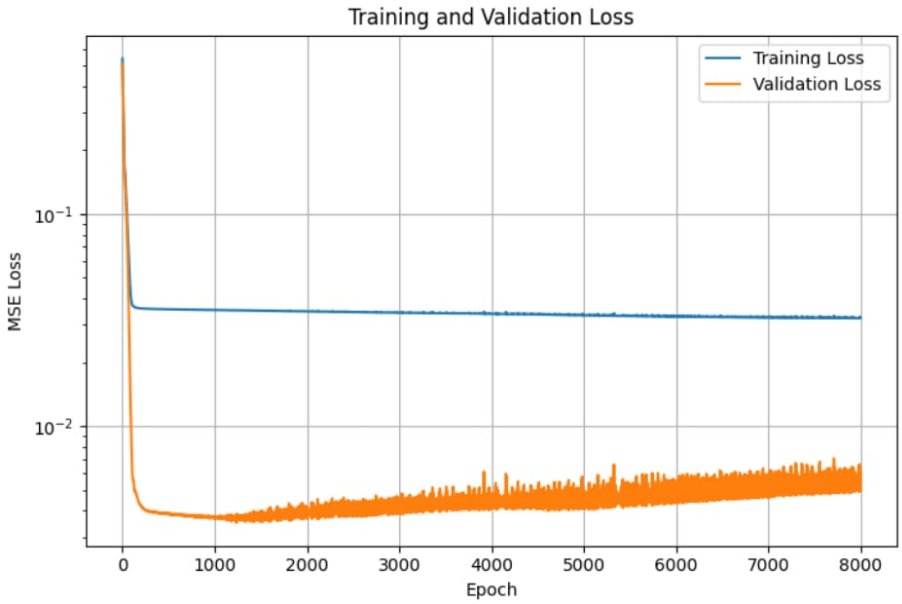
根據上面所述的方法跑程式 (輸入加噪，能否繼續學好 $\sin x$ 函數?)，在 Assignment 11 的結果 (跑 Python 程式)，跑出結果和圖片如下:

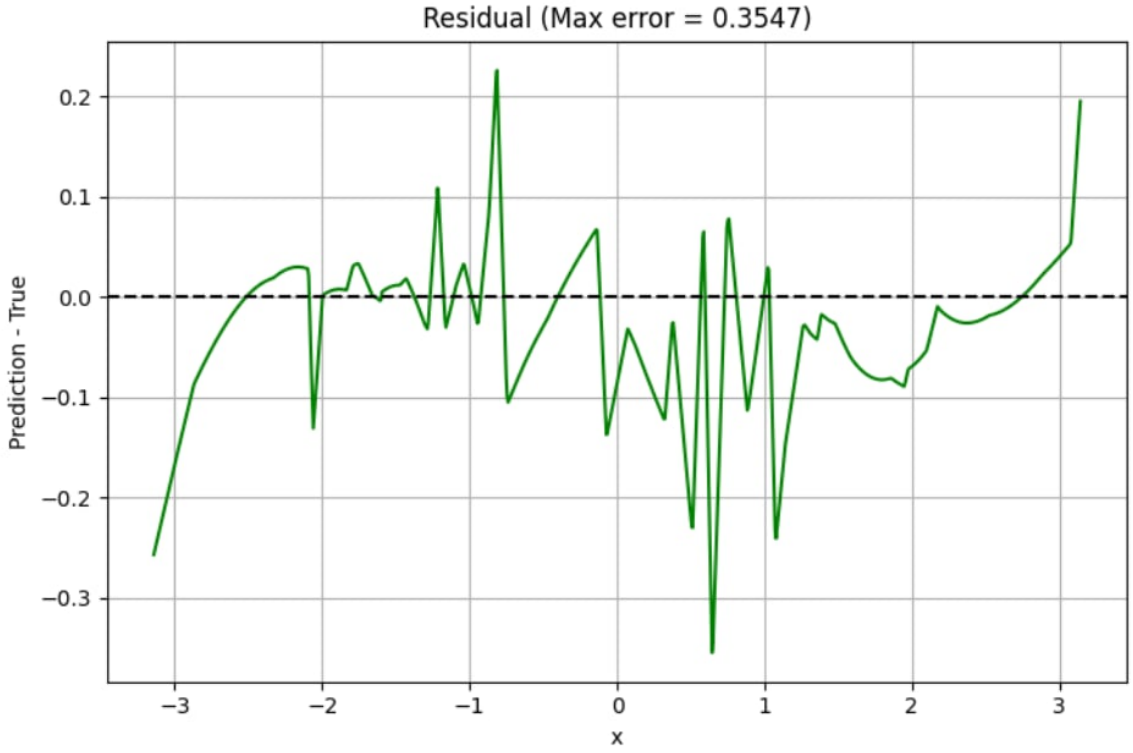
```
Epoch 1000, Train Loss: 0.035291, Val Loss: 0.003702
Epoch 2000, Train Loss: 0.034775, Val Loss: 0.004007
Epoch 3000, Train Loss: 0.034282, Val Loss: 0.004444
Epoch 4000, Train Loss: 0.033906, Val Loss: 0.004647
Epoch 5000, Train Loss: 0.033362, Val Loss: 0.004092
Epoch 6000, Train Loss: 0.032980, 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
```


True Function vs Neural Network Prediction
True sin(x)
NN Prediction (with input noise training)
Training points (clean)
y
x


Training Data (Input with Noise σ=0.3)
Noisy input training points
True sin(x)
sin(x)
Noisy x


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


Residual (Max error = 0.3547)
Prediction - True
0.2
0.1
0.0
−0.1
−0.2
−0.3
−3
−2
−1
0
1
2
3
x

用 Python 完成的 Supervised Learning，結果表明：

```
Epoch 8000
Train Loss    0.0325
Val Loss      0.0062
```

而最終 MSE= 0.006233；RMSE 0.079，而最大誤差（Max Abs Error）：0.354695，結果顯示模型確實學到了接近 sin 的 mapping，(RMSE 0.079)，表示平均誤差小，在 $[-1, 1]$ 範圍屬良好表現。所以，雜訊輸入並未阻止模型重建 underlying function，這驗證了：**「即使輸入受噪聲污染，神經網路仍可恢復隱含的目標函數。」**，但最大誤差偏大 (0.35) ，代表模型在部分區域（通常是陡峭或曲率變化大的區段）仍有不穩定性。而 Train Loss > Val Loss，代表模型正在學「去雜訊」，訓練資料：noisy(有加噪音); 驗證資料：clean 這樣的結果合理。整體而言，Toy Model 成功展示：在噪聲污染的輸入下，模型依然能 reconstruct 出潛在的眞實訊號。

4.2.4 討論（Discussion）

從這個 Toy Model 我們學到: 噪聲輸入並非致命問題，因爲模型能在有加噪的輸入下，學習統計上的平均結構，逼近眞實函數。這類似於未來能從 noisy EEG 重建影像。意即模型自然學到了「去雜訊」能力，因爲訓練輸入 noisy，但 supervision 是乾淨的 $\sin(x)$。這意味著模型 implicitly 做了：$\tilde{x} = x + \varepsilon \Rightarrow h_\theta(\tilde{x}) \approx \sin(x)$. 這模擬未來記憶重建中眞正困難的部分：recover signal from noisy neural data

而最大誤差提醒了更大的挑戰：不穩定性，若用在腦波 $\rightarrow$ 圖片重建中，局部高誤差會造成影像重建模糊、變形。這代表未來更大系統必須：使用 denoising 模組，使用更平滑的 function class（如 tanh、SIREN）或增加資料量、multi-trial averaging，上面的 Toy model 驗證了：「這樣的簡化是可行的」，因爲 Toy Model 已經展示出關鍵能力：**「模型能從 noisy input 中 recover underlying structure。」**

4.3 結論 (Conclusion)

這個簡化模型問題成功展示了：即使輸入資料含有高程度噪聲，神經網路仍能透過監督式學習恢復目標訊號。此結果提供了強烈證據，支持未來 AI 有能力從 noisy neural activity（如 EEG）中解析並重建較高層次的記憶內容。它不僅是可行的第一步，也揭示了後續研究最核心的課題：如何在巨大噪聲與高維訊號下，更穩定地重建人類主觀經驗的內容。