

機器學習 Final Project

劉昱賢

Contents

1	目標與介紹	2
1.1	AI 的未來能力	2
1.2	所需的成分與資源	2
1.3	涉及的機器學習類型	3
2	toy model 設置	4
2.1	簡化問題	4
2.2	玩具模型簡化想法	4
2.3	資料設定	5
2.4	模型介紹	5
3	實作與結果	5
3.1	不加 noise	5
3.2	noise=0.3	8
3.3	noise=0.6	10
4	結果討論	13
4.1	噪音對模型表現的影響	13
4.2	訓練 Loss 與驗證 Loss 的差異	13
4.3	判斷錯誤的樣本特徵	13
5	結論	14

1 目標與介紹

1.1 AI 的未來能力

我認為 20 年後的 AI 可以做到 **跨物種的實時溝通與翻譯**

可以即時理解不同動物 (甚至昆蟲或植物) 的聲音、動作、化學訊號，並翻譯成可理解的人類語言。

同時也能將人類語言轉換為對應的訊號，實現雙向交流。具體應用場景：

1. 野生動物保護

即時理解動物需求與警示訊號，制定更有效的保護策略。

2. 促進生物圈的共存

理解不同物種間互動，提升生態系統的和諧與穩定。

3. 透過溝通及研究進一步發展仿生技術

從動植物行為與感知中獲取靈感，創新科技設計。

4. 透過動物視角對地球進行探索或生態保育

利用動物感知能力進行環境監測與保護，提供人類新視角。

5. 進行不同物種之間的協作任務

人類與其他物種協作完成任務，探索跨物種合作可能性。

重要性

因為它能加深對於生物的理解，並可能促進環境保護、動物福利以及跨物種的科學探索。

1.2 所需的成分與資源

1. 資料

(1) 聲音

原始波形、頻譜、聲音強弱、節奏、頻率變化等

用於推斷吠叫種類、情緒訊號。

(2) 行為影像

姿態、動作序列、表情細節等

用於推斷動作意圖、社會行為。

(3) 生理訊號

體表溫度、呼吸/心跳、荷爾蒙/費洛蒙等

用於推斷壓力、興奮、恐懼等內在狀態。

(4) 動物的真實意圖：依賴弱標籤、自監督資料等方式取得。

2. 模型

(1) 聲音：CNN、Transformer、時序模型等

(2) 影像：CNN、ViT、YOLO、姿態估測等

(3) 行為：RNN、LSTM、Transformer、3D CNN 等

(4) 生理訊號：1D CNN、RNN、LSTM、時序 Transformer 等

最後目標：多模態整合

將聲音、影像、姿態、生理訊號放入同一個模型中進行分析。

1.3 涉及的機器學習類型

我認為可以將三者組合使用，以下列了幾個例子

將資料來源設定為：動物研究記錄、身體各項數據 (心率、體溫等)、動物行為及聲音..... 等

監督式學習 (Supervised Learning)

用途：從已標記好的資料中學習，建立動物行為、聲音與情緒或意圖之間的對應關係，讓 AI 能在新情境下準確判斷或預測動物的狀態。

例子：判斷狗狗叫聲是「尋求關注」還是「警戒防禦」

研究人員蒐集大量狗狗在不同情境下的叫聲，並根據行為標記為兩種類別。接著透過這些標註資料進行分類訓練，學會區分「親近型叫聲」與「威脅型叫聲」，未來可用於寵物情緒辨識或互動回饋系統。

非監督式學習 (Unsupervised Learning)

用途：從沒有標記的動物聲音、行為或生理數據中，探索不同模式之間的潛在關聯。

例子：鯨魚海豚聲波與行為分類

透過分析大量鯨魚或海豚的聲波資料，AI 自動將聲音分群，發現某些特定的聲音型態常出現在捕食、社交或不同行為當中。

強化學習 (Reinforcement Learning)

用途：讓 AI 在與動物的互動中逐步優化溝通方式，從回饋中學習如何產生最有效的交流信號。

例子：與鸚鵡的互動學習。

透過不斷與鸚鵡進行聲音模仿遊戲，觀察鸚鵡是否回應或模仿。

根據鸚鵡的反應（如是否回覆相似聲音或接近聲源），調整輸出的頻率與節奏，最終學會以最自然的聲音節奏與鸚鵡進行交流。

2 toy model 設置

2.1 簡化問題

簡化問題提出

我將研究目標設定為「狗」，因為牠們是目前最容易取得相關資料與行為樣本的物種之一。

本研究的問題為：能否透過狗的聲音、行為影像或體表溫度變化（以紅外線熱像儀測得）判斷其當下意圖。

這個簡化問題在概念上如何代表你理想中的最終能力？

此問題的核心在於：AI 是否能準確理解動物當下的意圖。

若模型能在這個簡化的場景下達成穩定且可靠的判斷，可進一步擴展為更多意圖分類、不同物種間的對應分析，最終發展出能夠雙向溝通、即時互動的跨物種理解模型。

可測試性

此簡化問題具有明確且可量化的測試方式。

模型的輸入（聲音、行為影像、體表溫度變化）皆可被標準化與同步化，作為可重複蒐集的資料來源。

其次，狗的意圖可透過行為學專家、已知行為分類（如攻擊、放鬆、尋求互動、焦慮等）、以及生理指標（心率、壓力荷爾蒙）進行交叉檢視。

訓練後的模型可使用這些標註資料進行分類及評估。若模型能在不同的情境（不同狗、不同環境、不同情緒誘因）中維持穩定預測，則代表具有可驗證的能力。

2.2 玩具模型簡化想法

想法：

輸入聲波以及標籤，將聲波均分成許多點並丟進 *nerual network* 中，最後求出標籤為 0, 1，並驗證結果。

2.3 資料設定

爲了簡化問題，將資料設定爲 $\sin x$ 的波，週期爲 2π 。每個波的能量、週期均相同。

每個樣本的正弦波： $s_i(x) = \sin(x + \phi_i)$, $i = 1, 2, \dots, 5000$, $x \in [0, 2\pi)$, $\phi_i \sim \text{Uniform}(0, 2\pi)$

檢視最高點的位置，若最高點在 $(0, \pi)$ 之間，則將資料標示爲 1。

若最高點在 $(\pi, 2\pi)$ 之間，則將資料標示爲 0。

除此之外，會對不同情況進行討論，測試不加 *noise* 是否可行，接著再加上 *noise* 模擬不同狀況。共生成 5000 筆資料。

2.4 模型介紹

先將資料切分爲三個部分，*Training Set*, *Validation Set*, *Test Set*，分別占比例 80%, 10%, 10%。

每個聲波會先拆分成 1000 個點，每個點間隔均相同。接著將聲波的點放入 *Neural Network* 中進行訓練，最後得出分類狀況。

Input: $x \in \mathbb{R}^{1000}$, *Label*: $y \in \{0, 1\}$, *Output*: 0 or 1 (*Binary Classification*)

Model: $h_\theta(x) = \sigma(W_2 \cdot \text{ReLU}(W_1 x + b_1) + b_2)$, 訓練次數 3000 次

$x \in \mathbb{R}^{1000}$, $W_1 \in \mathbb{R}^{32 \times 1000}$, $b_1 \in \mathbb{R}^{32}$, $W_2 \in \mathbb{R}^{1 \times 32}$, $b_2 \in \mathbb{R}$, σ is *Sigmoid function*

Loss function: BCE loss

評估方法：*Confusion Martix* + *accuracy*

3 實作與結果

3.1 不加 noise

生成並繪製了 5000 個波形，並將所有結果繪製。紅線代表標籤 1，藍線代表標籤 0。

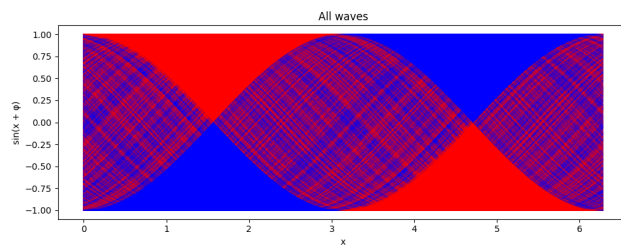


Figure 1: All wave with no noise

在生成完 5000 個波形後，我們使用神經網路對數據進行訓練，觀察在無噪音的理想情況下的模型表現。訓練 3000 次迭代後，結果如下：

- *Train loss* = 0.0007

- *valid loss* = 0.0024

對應的混淆矩陣如下：

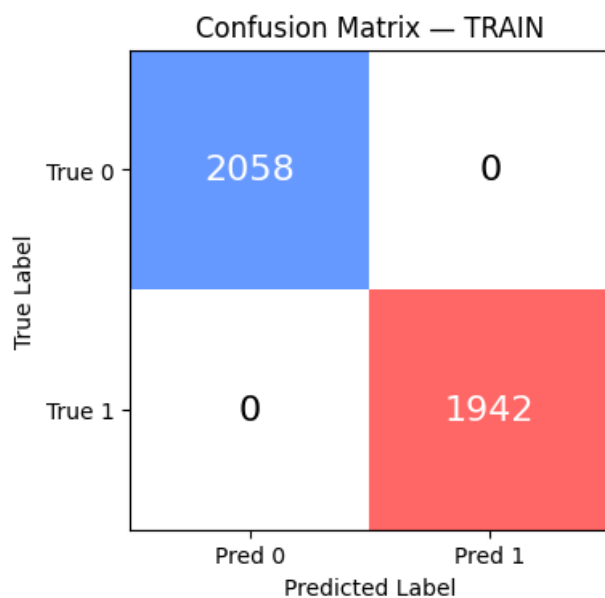


Figure 2: Confusion Matrix in train data with no noise

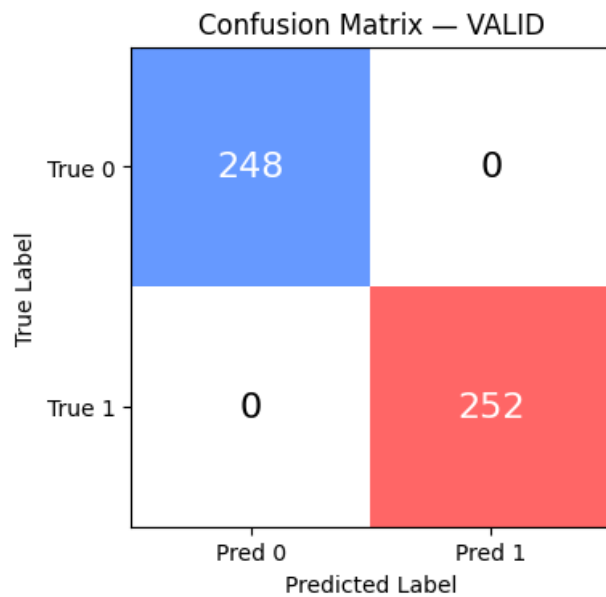


Figure 3: Confusion Matrix in valid data with no noise

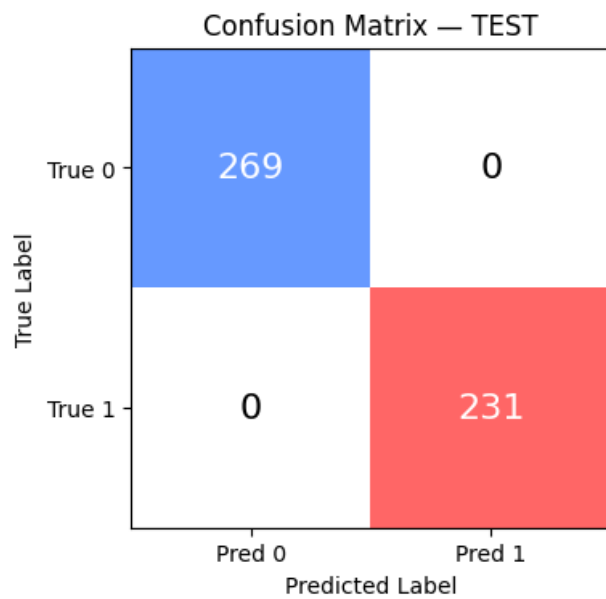


Figure 4: Confusion Matrix in test data with no noise

三個資料集的準確率如下：

$$\text{Train Accuracy} = 1.000, \quad \text{Valid Accuracy} = 1.000, \quad \text{Test Accuracy} = 1.000$$

訓練結果顯示，無噪音情況下模型能夠完美分類所有波形，三個資料集的準確率皆達 100%。

3.2 noise=0.3

我們生成並繪製了 5000 個波形，每個點上都加入了 *noise*。

紅線代表標籤為 1 的波，藍線代表標籤為 0 的波。

每個波的 1000 個點都加入了隨機噪音： $s_i(x_j) = s_i(x_j) + \epsilon_{ij}$, $\epsilon_{ij} \sim \mathcal{N}(0, 0.3^2)$, $i = 1, \dots, 5000$, $j = 1, \dots, 1000$

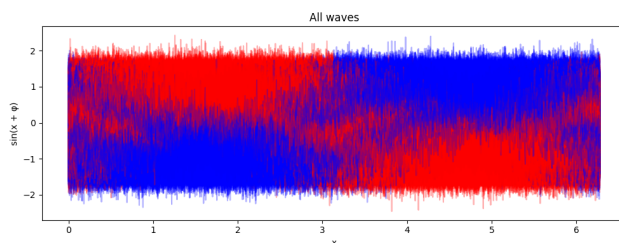


Figure 5: All wave with noise=0.3

單獨觀察第一個波型，如下

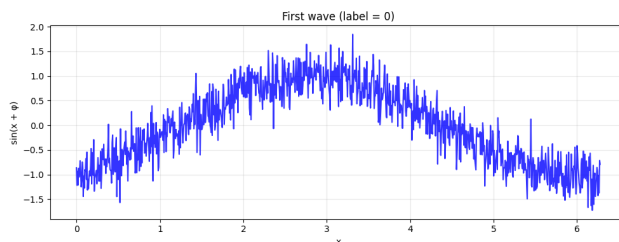


Figure 6: First wave with noise=0.3

在生成完 5000 個波形後，我們使用神經網路對數據進行訓練，觀察在加入噪音的情況下模型的表現。

訓練 3000 次迭代後，結果如下：

- Train loss = 0.0001

- Valid loss = 0.6151

對應的混淆矩陣如下：

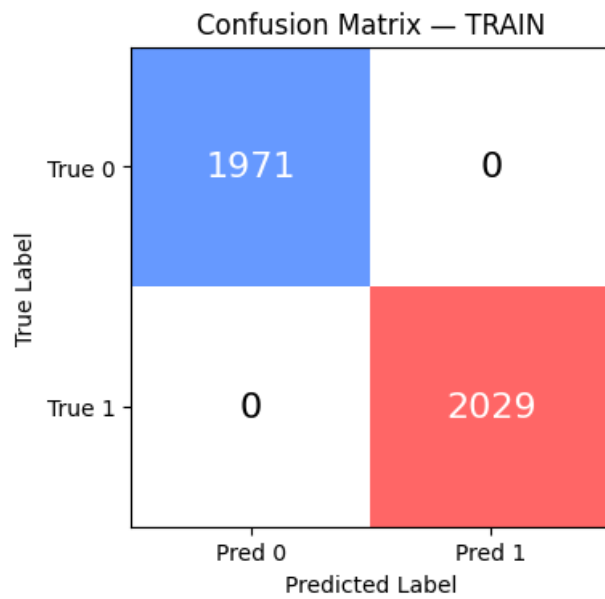


Figure 7: Confusion Matrix in train data with noise=0.3

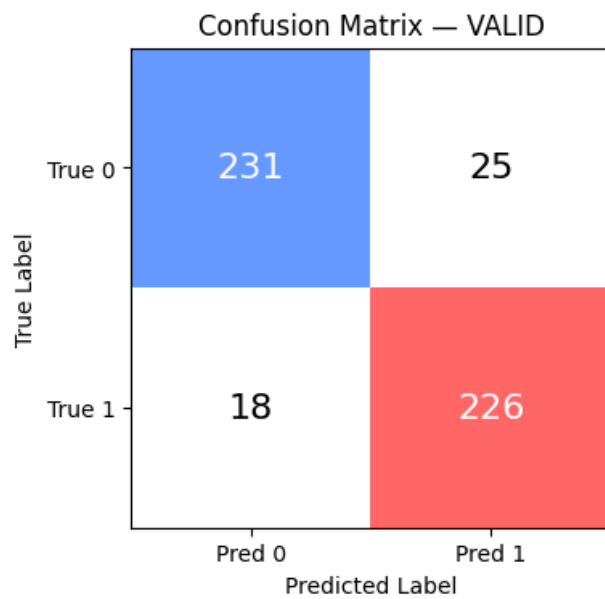


Figure 8: Confusion Matrix in valid data with noise=0.3

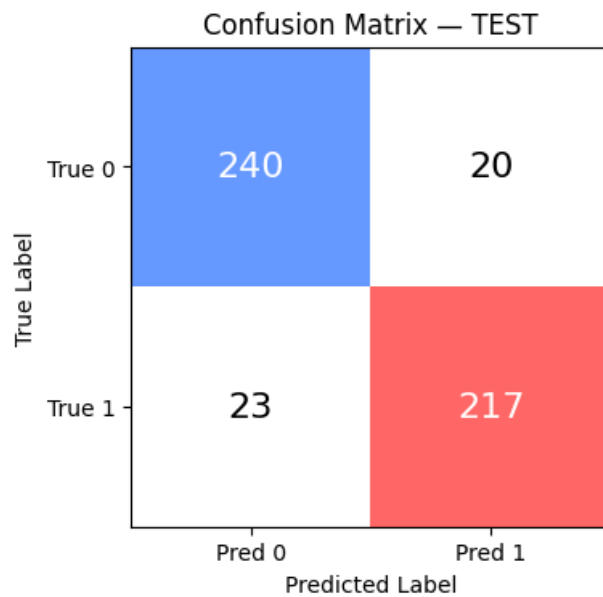


Figure 9: Confusion Matrix in test data with noise=0.3

三個資料集的準確率如下：

$$\text{Train Accuracy} = 1.000, \quad \text{Valid Accuracy} = 0.914, \quad \text{Test Accuracy} = 0.914$$

訓練結果顯示，在加入噪音後，模型在訓練集仍可達到完美分類，但在驗證集與測試集的準確率下降到約 91.4%，顯示噪音對模型造成了一定影響。

3.3 noise=0.6

由於在噪音 $\text{noise_std} = 0.3$ 時訓練結果仍然不錯，我們進一步測試 $\text{noise_std} = 0.6$ 的情況。

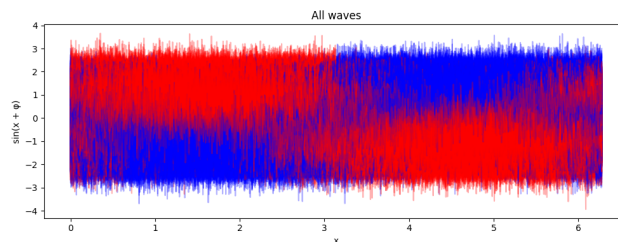


Figure 10: All wave with noise=0.6

單獨觀察第一個波型，如下

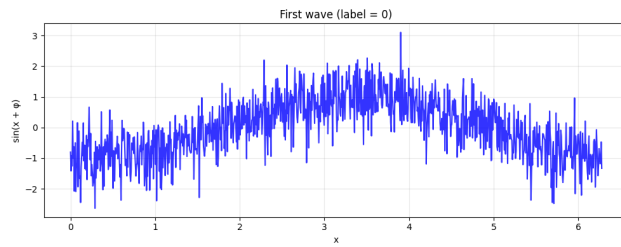


Figure 11: First wave with noise=0.6

在生成完 5000 個波形後，我們使用神經網路對數據進行訓練，觀察在加入較大噪音的情況下模型的表現。

訓練 3000 次迭代後，結果如下：

- *Train Loss* = 0.0001

- *Valid Loss* = 1.8199

對應的混淆矩陣如下：

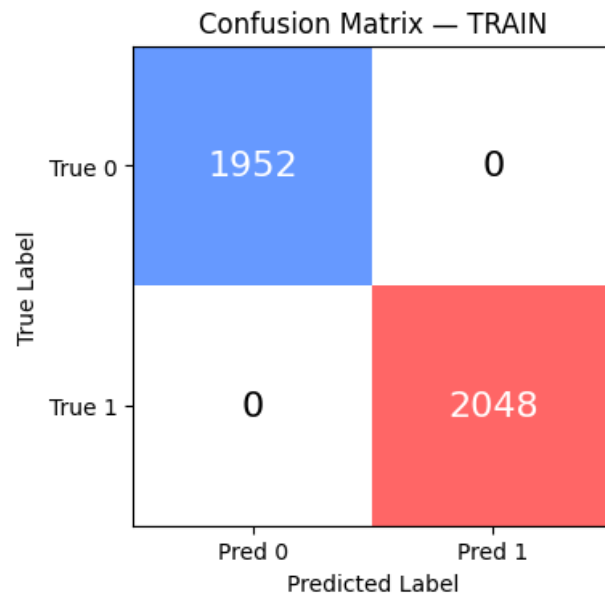


Figure 12: Confusion Matrix in train data with noise=0.6

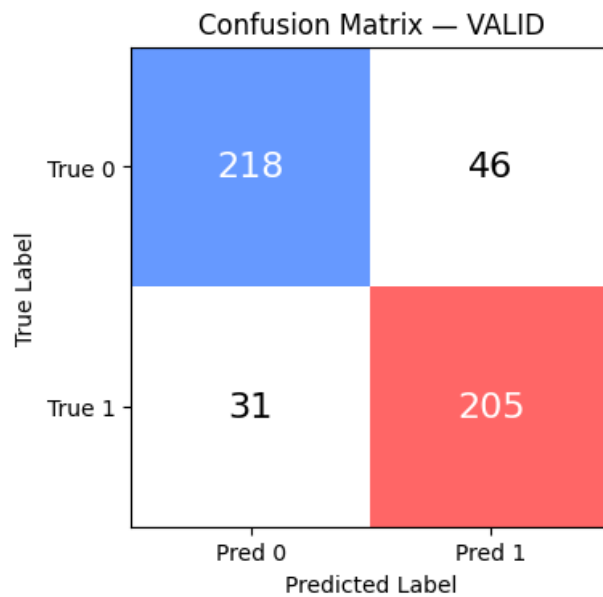


Figure 13: Confusion Matrix in valid data with noise=0.6

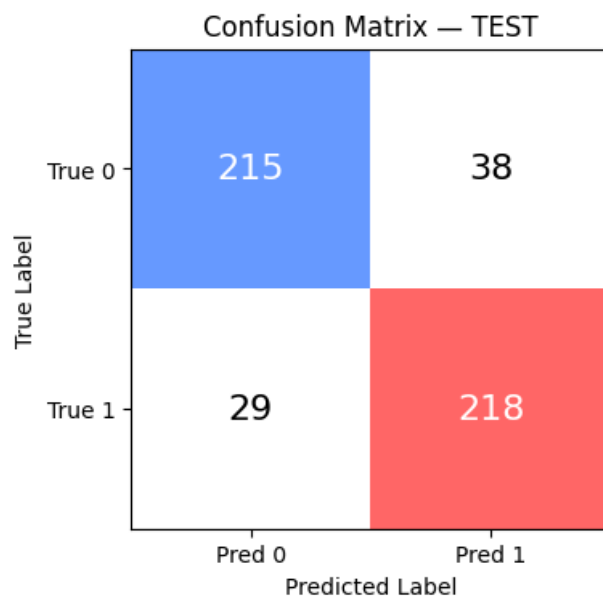


Figure 14: Confusion Matrix in test data with noise=0.6

三個資料集的準確率如下：

$$\text{Train Accuracy} = 1.000, \quad \text{Valid Accuracy} = 0.846, \quad \text{Test Accuracy} = 0.866$$

訓練結果顯示，在較大噪音的情況下，模型雖然在訓練集仍可達到完美分類，但在驗證集與測試集的準確率下降，顯示噪音對模型產生了較大影響，但結果依然還不錯。

4 結果討論

4.1 噪音對模型表現的影響

當 $\text{noise} = 0$ 時，模型完全學會了數據特徵，三個資料集的準確率皆為 100%。

當 $\text{noise} = 0.3$ 時，訓練集仍保持完美分類，但驗證集和測試集下降到約 91.4%，說明噪音增加了資料的變異性，使模型在未見過的資料上分類能力略微下降。

當 $\text{noise} = 0.6$ 時，驗證集與測試集下降到約 85—87%，顯示噪音過大會顯著影響模型能力，雖然訓練集仍可完美分類（可能存在過擬合現象）。

4.2 訓練 Loss 與驗證 Loss 的差異

當 $\text{noise} = 0$ 時，*Train Loss* 與 *Valid Loss* 均接近 0，顯示過擬合問題不存在。

當 $\text{noise} > 0$ 時，*Train Loss* 仍很低，但 *Valid Loss* 明顯上升，說明模型對訓練資料中的噪音過度擬合，而驗證資料時出現誤差。

4.3 判斷錯誤的樣本特徵

下圖顯示在 $\text{noise} = 0.3$ 與 $\text{noise} = 0.6$ 時被誤判的波形樣本。

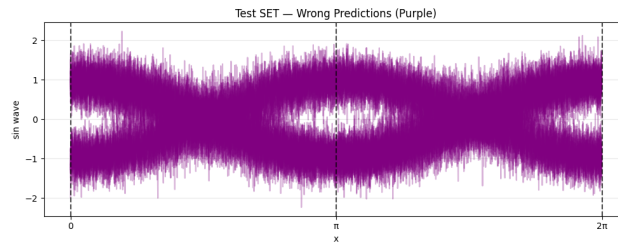


Figure 15: Wrongly predicted wave with noise=0.3

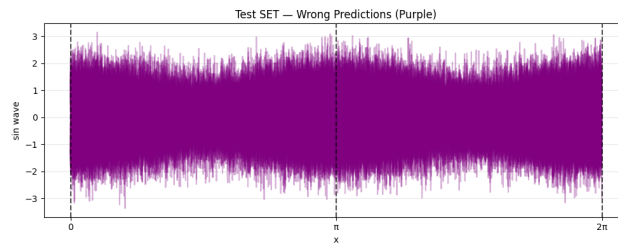


Figure 16: Wrongly predicted wave with noise=0.6

觀察發現，在有噪音的情況下，容易判斷錯誤的資料及其最大值會集中在 $0, \pi, 2\pi$ 附近，這些位置對分類來說較為模糊，容易受到噪音干擾。

5 結論

1. 模型對原始波形的學習效果非常好，能夠完美分類無噪音資料。
2. 隨著噪音增加，模型在訓練集上的表現仍然良好，但驗證集與測試集的準確率逐漸下降，顯示噪音對模型具有明顯影響。
3. 當噪音增加時，訓練 *Loss* 仍保持極低，但驗證 *Loss* 明顯上升，可能是模型對訓練資料的噪音過擬合所致。
4. 過大的噪音會降低驗證集與測試集的分類效果，但模型仍能保留一定辨識能力。
5. 若希望模型在更大噪音下穩定，可考慮增加資料量或使用更深層網路等方法。