

113034553_黃熾云_HW2

- (20 pts) Select 2 hyper-parameters of the artificial neural network used in Lab 2 and set 3 different values for each. Perform experiments to compare the effects of varying these hyper-parameters on the loss and accuracy metrics across the training, validation, and test datasets. Present your findings with appropriate tables.

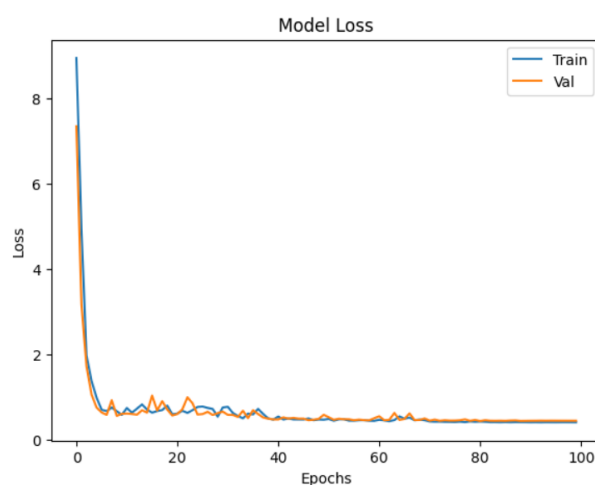
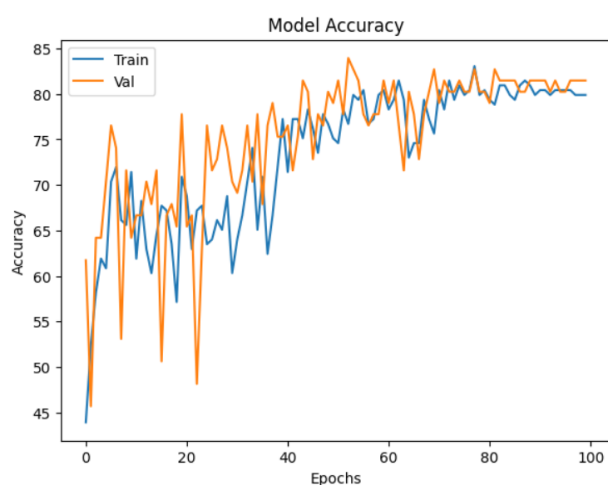
調整內容	參數設定一（小）	參數設定二（中）	參數設定三（大）
Batch size	8	16	64
Hidden layer nodes	256	256 * 2	256 * 4
Eta_min	0.0000	0.0001	0.01

參數調整的實驗順序如下表：

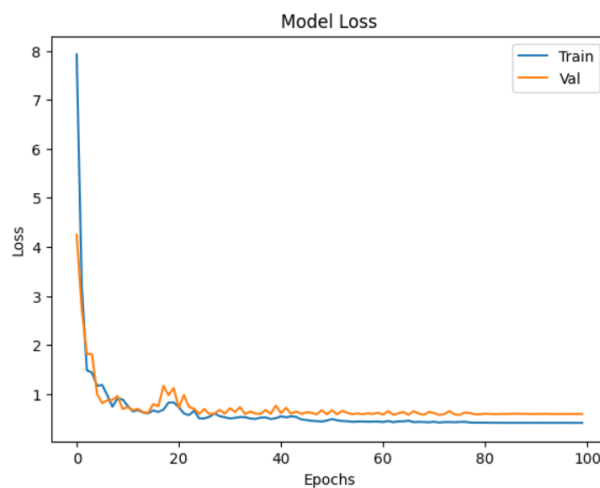
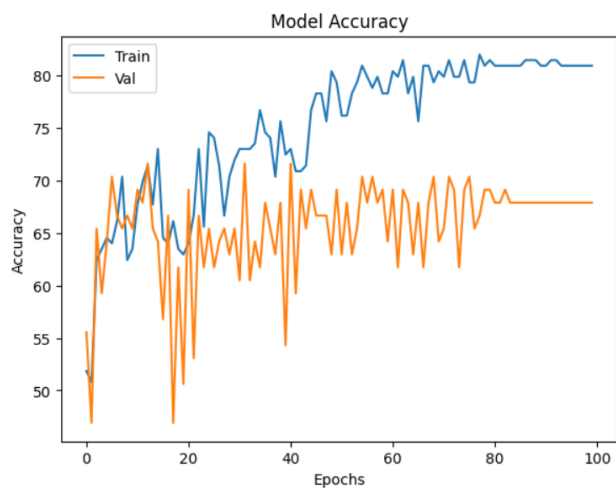
實驗	調整內容	調整參數	測試集準確度	小結
1	Batch size	32→64	64.5%→61.3%	覺得 Batch size 變小比較好，於是嘗試實驗 2
2	Batch size	32→16	64.5%→77.4%	得在這個資料中實驗 1 的猜測正確，繼續調小
3	Batch size	32→8	64.5%→74.2%	調整回 Batch size =16
4	模型結構	nn.Linear(13, 256*2), nn.ReLU(), nn.Linear(256*2, 256), nn.ReLU(), nn.Linear(256, 2)	77.4%→83.9%	Node 數量增加會提高準確度
5	模型結構	nn.Linear(13, 256*4), nn.ReLU(), nn.Linear(256*4, 256), nn.ReLU(), nn.Linear(256, 2)	74.2%→77.4%	調整回實驗資料五
6	lr_scheduler	lr_scheduler = CosineAnnealingLR(optimizer, T_max=epochs/2, eta_min=0.01)	77.4%→83.5%	* T_max 決定學習率從初始值降低到最小值 (eta_min) 所需的 epoch 數。當訓練達到 T_max epoch 時，學習率

				會降低到 <code>eta_min</code> ，然後重新開始新的週期（如果有 <code>restart</code> ）。
7	<code>lr_scheduler</code>	<code>lr_scheduler = CosineAnnealingLR(optimizer, T_max=epochs/2, eta_min=0.0001)</code>	77.4%→83.9%	* <code>eta_min</code> 是學習率的最小值，當學習率下降到這個值後，不會再降低。

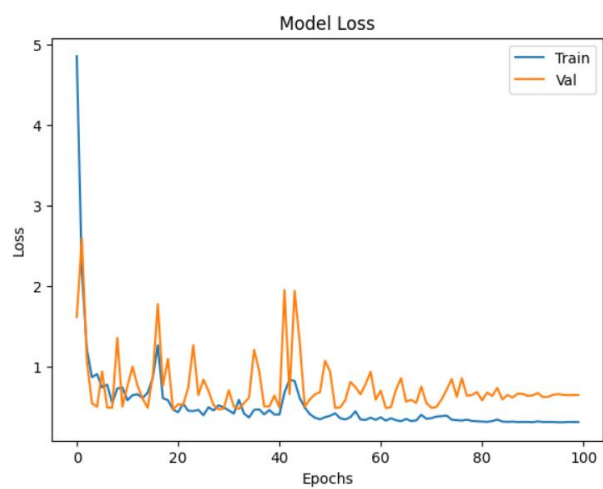
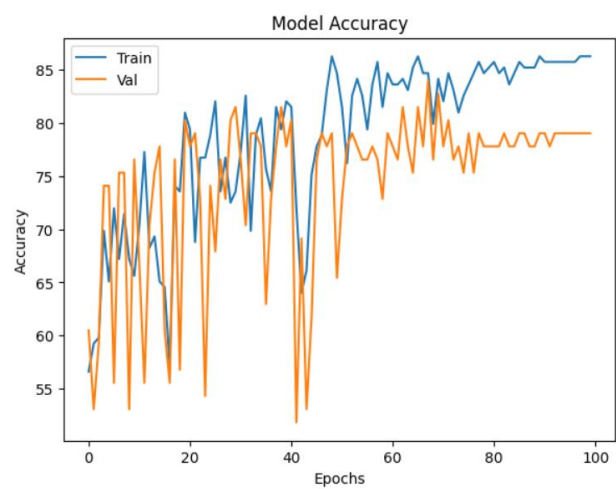
未調整之初使設定



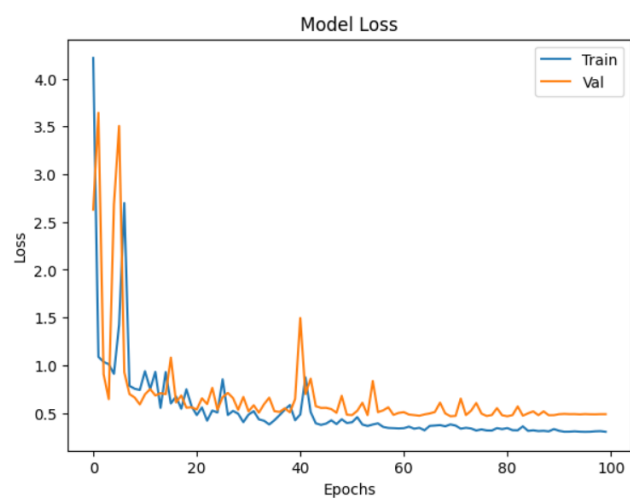
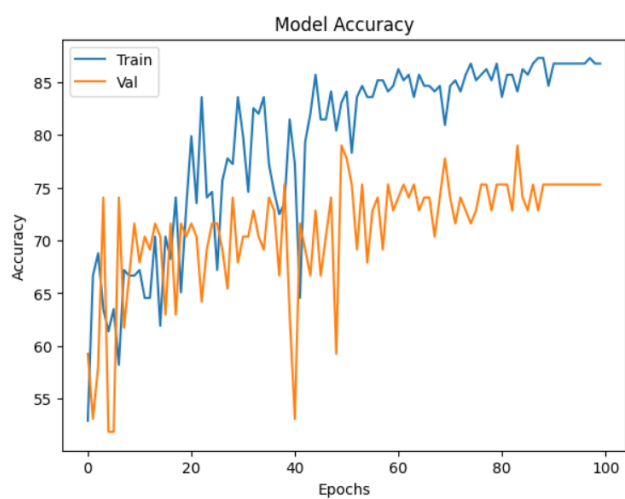
實驗資料一



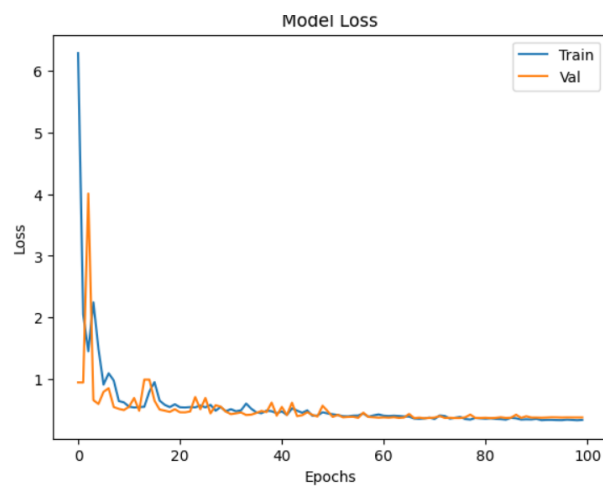
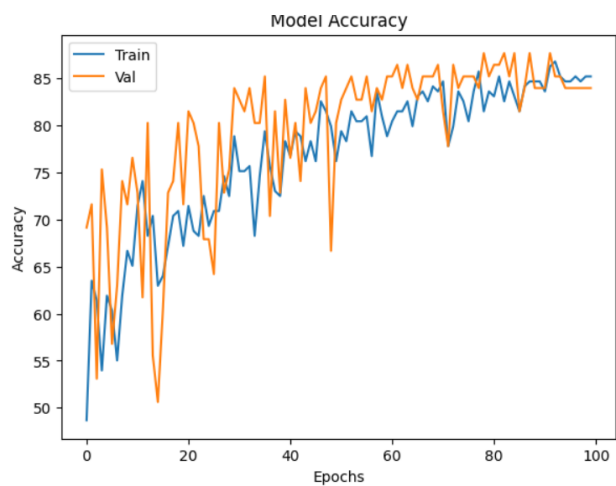
實驗資料二



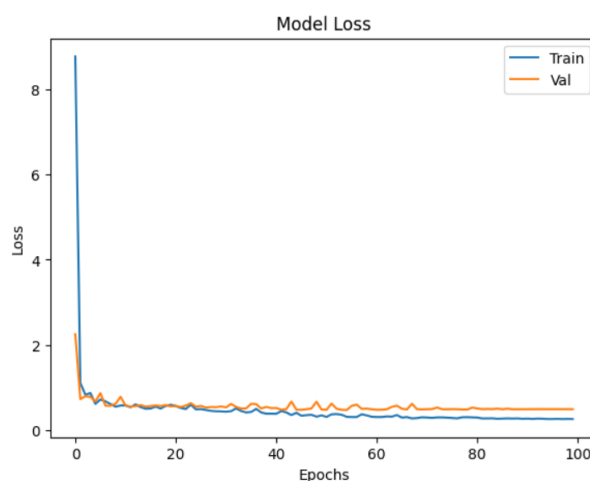
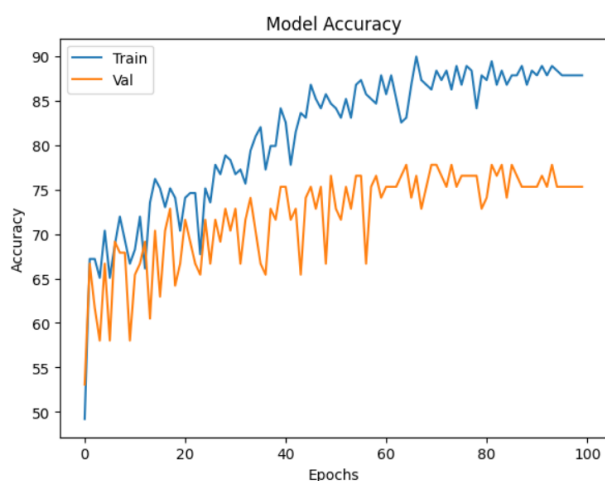
實驗資料三



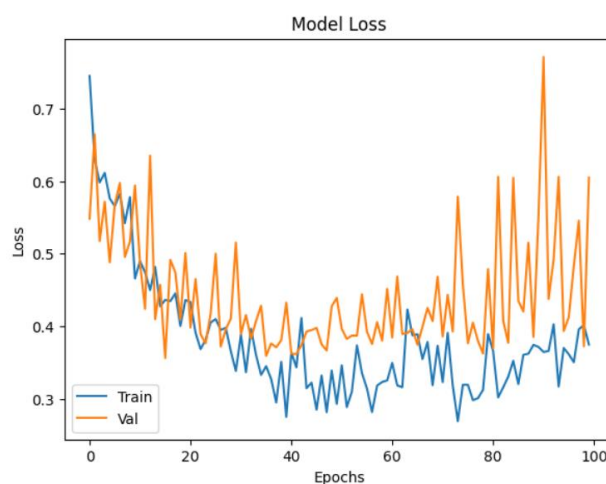
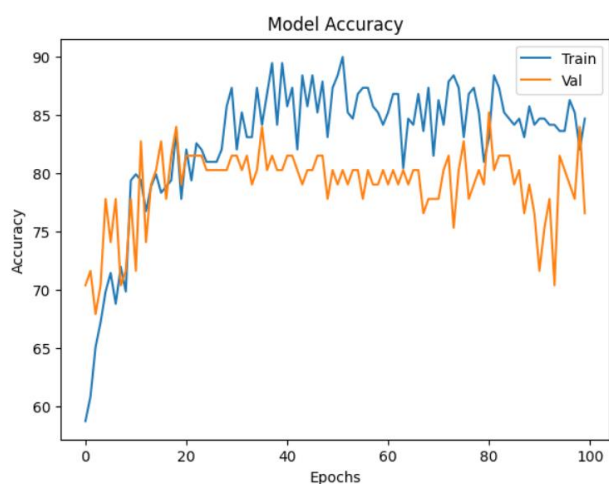
實驗資料四



實驗資料五



實驗資料六



觀察與分析

在訓練準確度或驗證準確度出現大幅度波動（鋸齒狀），可能是由於以下幾種原因：

- ✓ Batch Size 太小：每個 batch 的樣本分佈變化較大，導致梯度更新不穩定，造成準確度大幅波動。
- ✓ Learning Rate 過大：梯度更新幅度太大，導致模型無法穩定收斂，準確度在高低間來回震盪。
- ✓ 訓練集與驗證集的數據分佈不同：如果訓練數據與驗證數據的分佈不一致（如某些類別在訓練集中很多，但在驗證集中很少），則模型可能在不同數據集上的表現大幅波動。

2. (20 pts) Based on your experiments in Question 1, analyze the outcomes. What differences do you observe with the changes in hyper-parameters? Discuss whether these adjustments contributed to improvements in model performance, you can use plots to support your points. (Approximately 100 words.)

根據第一題的實驗結果顯示：

- ✓ Batch Size 的影響
 - 減小 Batch Size (從 32 → 16) 後，測試集準確率提升 (64.5% → 77.4%)。
 - 進一步減小至 8，準確率下降 (74.2%)，說明太小的 Batch Size 可能導致不穩定的學習。
 - 這顯示出適當的 Batch Size (16) 有助於提高泛化能力，但太大或太小都可能影響效果。
- ✓ 模型結構 (神經元數量)
 - 增加神經元數量 (256 → 512)，測試準確率從 77.4% 提升至 83.9%。
 - 但當神經元進一步增加 (1024)，準確率反而回落至 77.4%。
 - 顯示模型可能變得過於複雜，出現過擬合。
- ✓ 學習率調度 (lr_scheduler)
 - 使用 CosineAnnealingLR，準確率維持在 83.5%。
 - 學習率調整有助於穩定訓練，但未帶來額外的明顯提升。
- ✓ 過擬合的可能性：當模型神經元數量過多時，測試準確率下降，說明模型學習了過多訓練資料的細節，導致泛化能力下降。
- ✓ 欠擬合的可能性：當 Batch Size 過大時 (如 64)，測試準確率下降，可能是因為模型學習不足，無法有效捕捉資料特徵。

根據第一題圖表，有訓練準確度 (藍色曲線) 遠高於驗證準確度 (橘色曲線)，這通常表示過擬合 (Overfitting)，也就是模型在訓練資料上表現良好，但無法泛化到未見過的資料。

可能的原因包括：

- ✓ 模型過於複雜 (參數太多，如神經元數量過多)：模型學習了訓練資料中的細節和噪音，而非真正的模式，導致在驗證資料上表現不佳。
- ✓ 訓練次數過多 (Epoch 太多)
 - 訓練時間過長，導致模型記住了訓練資料，而不是學習一般性的特徵。
- ✓ 正則化不足 (Lack of Regularization)
 - 沒有使用 Dropout、L2 正則化 (Weight Decay)，導致模型過度擬合訓練資料。
- ✓ Batch Size 選擇不當

- Batch Size 過小時，模型更新波動較大，可能影響泛化能力。
- Batch Size 過大時，可能導致模型缺乏足夠的多樣性來學習泛化特徵。

3. (20 pts) In Lab 2, you may have noticed a discrepancy in accuracy between the training and test datasets. What do you think causes this occurrence? Discuss potential reasons for the gap in accuracy. (Approximately 100 words.)

- 過擬合 (Overfitting): 模型在訓練時學得太仔細，甚至記住了訓練資料中的雜訊，導致在新資料上表現不佳。
- 資料不均衡 (Data Imbalance): 如果某一類資料比其他類多，模型可能會偏向預測較多的那一類，影響測試結果的準確性。例如在 cp 欄位，值為 0 的個數遠大於其他值的個數。
- 正則化不足 (Lack of Regularization): 如果沒有使用 Dropout、權重衰減 (Weight Decay) 等技術，模型可能會變得過於複雜，影響泛化能力。
- 資料分佈不同 (Data Distribution Mismatch): 訓練集和測試集的資料來源或特徵分佈不同，可能會讓模型在測試集上表現不佳。

4. (20 pts) Discuss methodologies for selecting relevant features in a tabular dataset for machine learning models. Highlight the importance of feature selection and how it can impact model performance. You are encouraged to consult external resources to support your arguments. Please cite any sources you refer to. (Approximately 100 words, excluding reference.)

在表格資料中，若兩個欄位之間具有高度相關的特徵，可能會對模型帶來相似的資訊，進而影響學習的效果。因此特徵選擇方法有助於避免共線性問題 (Multicollinearity)，若兩個特徵的相關係數趨近於 1，容易造成回歸模型的係數變得不穩定，模型難以穩定收斂，因而影響最終模型訓練表現；也可以協助提升模型效能，移除冗餘或不相關的特徵有助於降低模型的計算複雜度，加快訓練速度，並改善預測準確率。並且減少過擬合 (Overfitting)，過多的特徵可能導致模型過於擬合訓練數據，使其無法在測試數據上泛化。適當的特徵選擇能夠減少噪音，提高模型的泛化能力；若欄位過多在後續的解讀分析上也容易造成紊亂，模型難以解釋，也不知道哪個特徵是模型的關鍵影響特徵，而特徵選擇手段可以協助提升模型可解釋性 (Interpretability)。常見的方法包括：

- 篩選法 (Filter Methods): 使用統計方法 (如相關係數、交互資訊) 來篩選重要特徵。例如 forward stepwise selection method.
- 包裹法 (Wrapper Methods): 透過演算法 (如遞歸特徵消除 RFE) 來測試不同特徵組合。
- 嵌入法 (Embedded Methods): 在模型訓練過程中進行特徵選擇，例如 Lasso 回歸 (L1

正則化)。

- 降維技術 (Dimensionality Reduction)：例如主成分分析 (PCA) 能移除冗餘特徵。

參考資料：[Feature selection](#)

5. (20 pts) While artificial neural networks (ANNs) are versatile, they may not always be the most efficient choice for handling tabular data. Identify and describe an alternative deep learning model that is better suited for tabular datasets. Explain the rationale behind its design specifically for tabular data, including its key features and advantages. Ensure you to reference any external sources you consult. (Approximately 150 words, excluding reference.)

根據 TabNet: Attentive Interpretable Tabular Learning 文獻回顧，TabNet 是專門為表格資料設計的深度學習模型。TabNet 採用決策感知注意力機制 (decision-aware attention)，使模型能夠根據資料特徵動態選擇重要資訊。其主要優勢包括：

- 可解釋的特徵選擇：透過稀疏注意力 (sparse attention) 機制，針對每個決策步驟選擇關鍵特徵。欄位之間的關係在 ANN 中難以捕捉，表格資料的特徵通常具有不同類型，例如數值型 (如年齡、價格) 與類別型 (如職業、地區)。這些特徵的相依關係在 ANN 的全連接層 (Fully Connected Layers) 中難以有效學習。ANN 沒有固定的空間結構，單純的矩陣相乘難以學習非線性特徵關係。
- 更好的泛化能力：透過特徵稀疏性來降低過擬合的風險。
- 提升離散特徵的學習能力：表格資料中的離散特徵 (如類別編碼)，在 ANN 中可能無法有效學習特徵重要性，而透過注意力機制，模型可以根據決策步驟動態調整不同特徵的影響力。
- 適應表格結構：深度學習的梯度下降對表格資料不夠高效。ANN 的權重更新方式基於梯度下降 (Gradient Descent)，但對於表格資料，這種方式可能不夠靈活。稀疏注意力允許模型動態學習特徵重要性，類似 XGBoost 的特徵分裂策略，使學習更具針對性。
- 克服 ANN 固有特徵學習缺陷：ANN 在沒有額外機制 (如注意力機制 Attention) 的情況下，對所有特徵一視同仁地學習權重，無法像 XGBoost 那樣自動學習「哪些特徵更重要」。

結論

- ✓ ANN 無法像決策樹 或 TabNet 一樣，自動學習特徵之間的關係，這使得它在表格資料上表現較差。

- ✓ 梯度下降法不適合離散或高維類別資料，而 XGBoost 使用決策樹結構來適應這類資料。
- ✓ TabNet 改進了 ANN 的弱點，透過「決策感知注意力機制」來動態選擇重要特徵，使其在表格資料上比標準 ANN 更有效。

[TabNet: Attentive Interpretable Tabular Learning](#)