

Deep Learning and Industrial Applications

Homework 2

撰寫者 / 113034611 / 簡榮信

(一)、選擇人工神經網絡 (ANN) 在 Lab 2 中使用的兩個超參數，並為每個超參數設定三種不同的值。進行實驗，比較這些超參數變化對於訓練、驗證和測試數據集的損失 (loss) 和準確率 (accuracy) 指標的影響。使用適當的表格呈現您的發現。

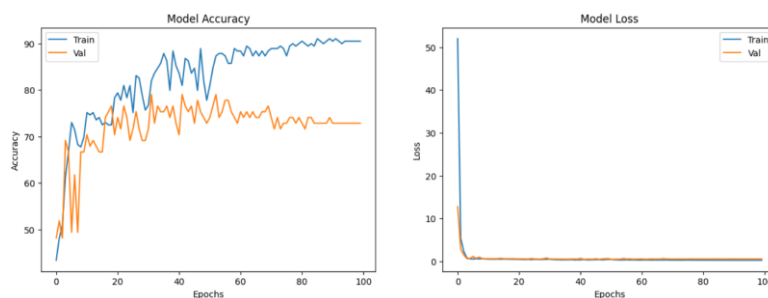
A：

此次作業我選擇 1.批次大小 (Batch Size) 和 2.學習率 (Learning Rate) 作為超參數，針對兩個參數設定三種不同的數值，以測試交互結果：

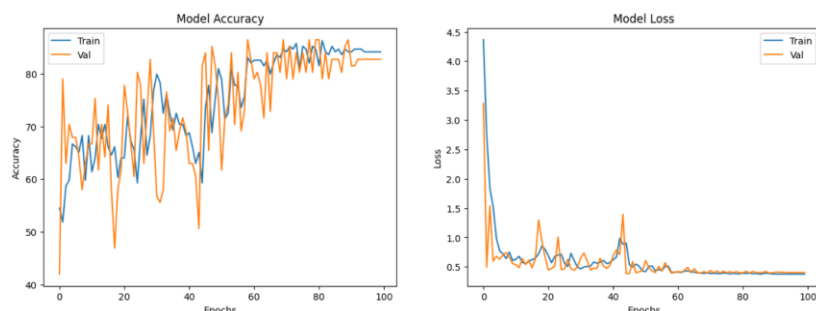
	Batch Size	Learning Rate
Level 1	32	0.01
Level 2	64	0.001
Level 3	128	0.0001

以下為各超參數不同水準下之交互參照結果，包括 Accuracy & Loss。

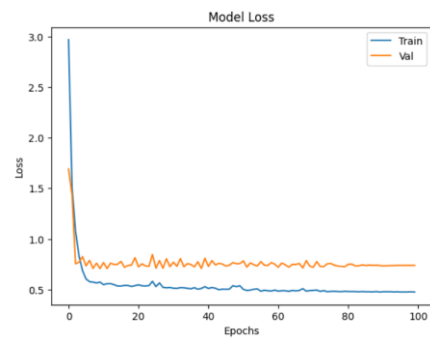
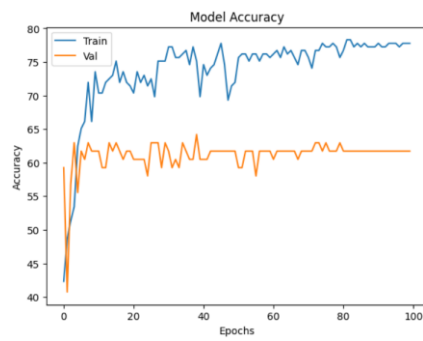
1. 【32, 0.01】



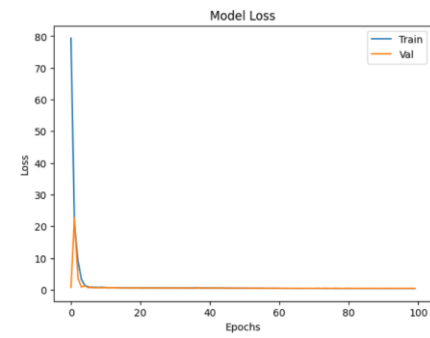
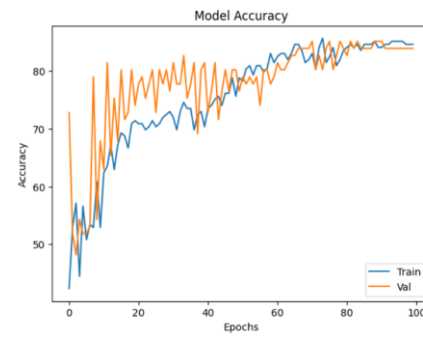
2. 【32, 0.001】



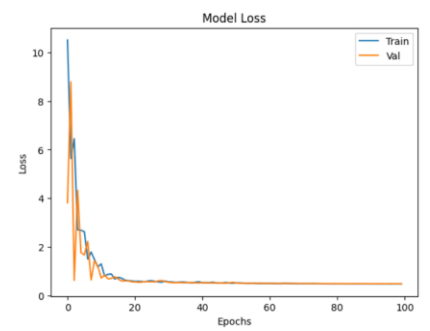
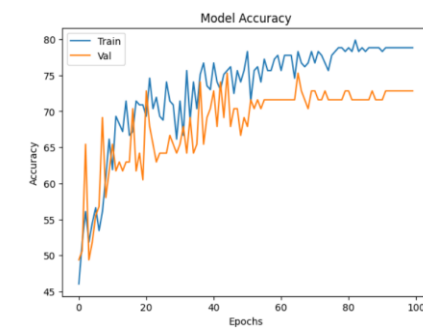
3. **【32, 0.0001】**



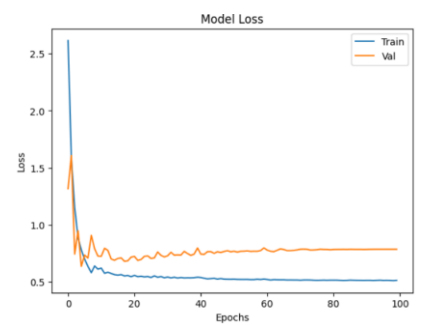
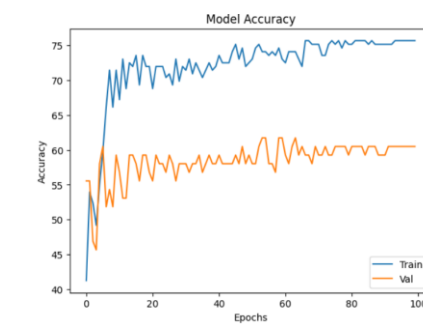
4. **【64, 0.01】**



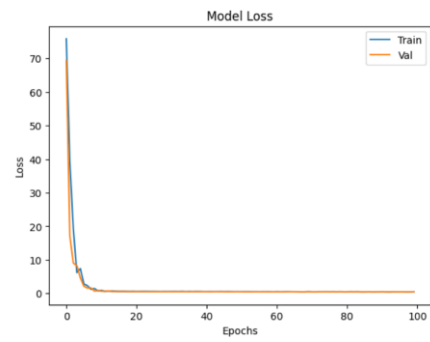
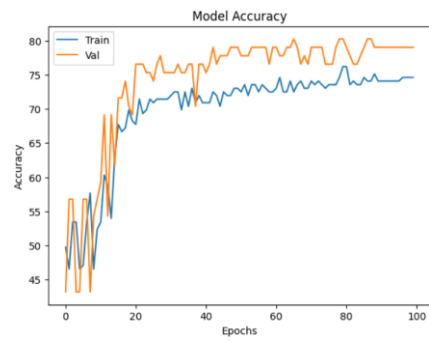
5. **【64, 0.001】**



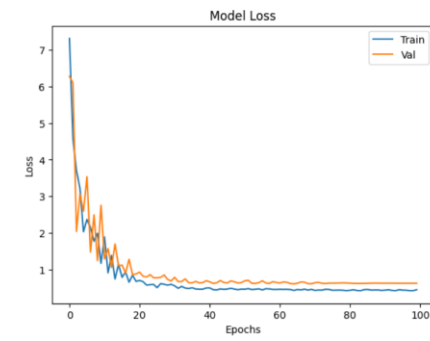
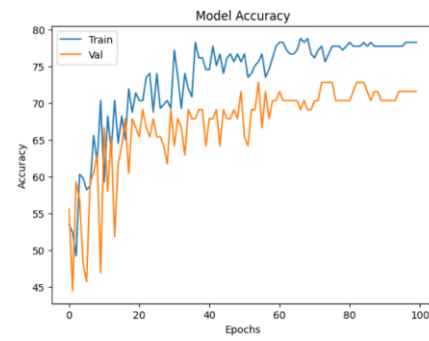
6. **【64, 0.0001】**



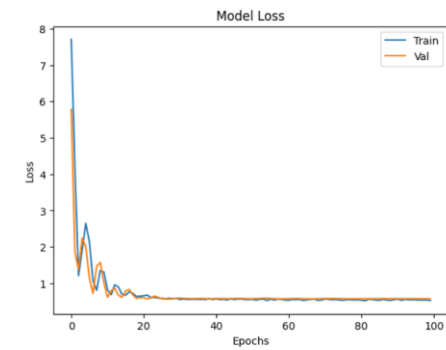
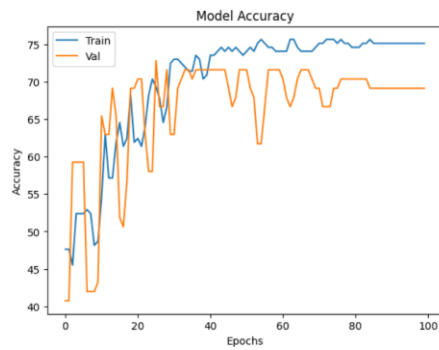
7. **【128, 0.01】**



8. **【128, 0.001】**



9. **【128, 0.0001】**



以上九張圖片之分析比較詳列於下頁表一。

表一、超參數分析比較

圖片編號	Batch Size	Learning Rate	收斂速度	訓練準確率	驗證準確率	過擬合	整體評價
1	32	0.01	快，但波動大	高 (~90%)	早期高 → 後期下降 (~72%)	明顯	不穩定，過擬合
2	32	0.001	中	中 (~85%)	波動但趨近訓練 (~83%)	無明顯	平衡良好
3	32	0.0001	慢	高 (~78%)	停留在低水準 (~62%)	有些過擬合	可能 underfitting
4	64	0.01	快	高 (~85%)	高 (~85%)	無明顯	整體最佳
5	64	0.001	中偏快	中上 (~80%)	穩定提升 (~73%)	輕微	訓練穩定
6	64	0.0001	慢	中 (~75%)	低 (~60%)	有	泛化差
7	128	0.01	快	中 (~75%)	高於訓練 (~80%)	需觀察	可能數據偏差
8	128	0.001	中	中上 (~78%)	中 (~72%)	小幅	表現穩定
9	128	0.0001	慢	中 (~75%)	中 (~70%)	小幅	進步緩慢

綜合以上九組訓練結果，模型表現最佳的超參數組合為 Batch Size = 64 與 Learning Rate = 0.01。此組合在訓練與驗證準確率間達到良好的平衡一致性，且收斂速度較快速，無明顯過擬合之情形。

● 此次實驗結果顯示：

- 較大的學習率 (0.01) 在本實驗中表現最佳：
 - 加快模型收斂速度。
 - 在合適的批次大小 (如 64) 下可維持良好的穩定性與泛化能力，無明顯過擬合問題。
- 較小的學習率 (0.0001) 雖能減緩參數更新幅度，但導致模型學習進度過慢，驗證準確率長期停滯，有跡象疑似 **underfitting**。
- 中等批次大小 (64) 在各種學習率下皆展現出不錯的平衡性，能夠兼顧準確率、loss 穩定性與泛化效果，為本次實驗的最佳選擇。
- 過小 (32) 或過大 (128) 的批次大小容易造成訓練不穩定或泛化能力下降，後續需依照實際模型與資料進一步調整變化。

綜合上述，建議未來模型訓練中可優先採用 Batch Size 為 64，Learning Rate 為 0.01 的組合作為起始設定，再視情況微調其他超參數或是參數以提升模型表現。

(二)、根據問題 1 的實驗結果進行分析。當超參數發生變化時，您觀察到哪些不同的影響？討論這些調整是否提升了模型的表現，並可以使用圖表來支持您的觀點。

A：

- 從第一題的實驗結果來看，(1) 學習率與 (2) 批次大小的選擇對於模型的收斂速度與泛化能力具有明顯影響：

■ **學習率**的影響：

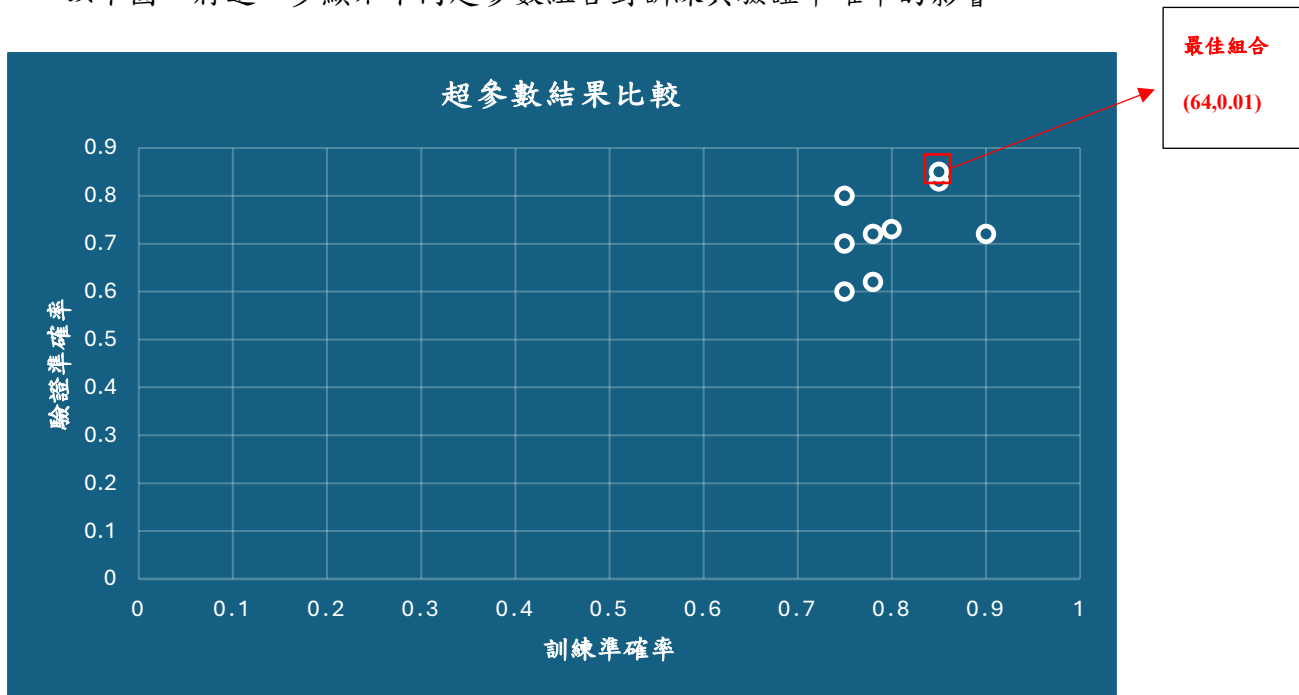
- **較高學習率 (0.01)** 在合適條件下 (如 Batch Size = 64)，不僅加快模型的收斂速度，且能有效提升驗證準確率，整體表現最佳。
- 然而在 Batch Size 較小的情況下 (如 32)，高學習率易造成訓練過快、準確率起伏大甚至過擬合，驗證表現反而下滑。
- 過低學習率 (0.0001) 則會導致模型學習進度慢，訓練與驗證準確率均會停留在低點，模型未能充分學習資料特徵。
- 總結：學習率需根據模型與批次大小調整，0.01 在本實驗中為表現最佳的設定，前提是需要搭配適當的 batch size。

■ **批次大小**的影響：

- **中等批次大小 (64)** 整體表現最穩定，不論搭配何種學習率，都較能兼顧收斂效率與後續驗證的泛化能力，訓練曲線與驗證趨勢一致性高，建議將此數值作為預設值。
- 批次過小 (32) 容易起伏劇烈，即使訓練準確率高，也可能導致驗證表現不佳，過擬合風險高。
- 批次過大 (128) 則會使模型更新步伐較緩慢，學習變得相對保守，驗證準確率上升有限，可能會導致部分組合欠擬合。
- 總結：中等 batch size 提供了收斂速度與準確率之間的良好平衡，64 為泛化能力最好的選擇。

- 若訓練損失持續下降但驗證損失開始上升，代表模型出現 **過擬合現象** (可能是學習率過高或 batch size 過小所致)。
- 若訓練與驗證損失皆長期維持在高位，代表模型尚未收斂，可能為學習率過低或模型學習能力不足。

以下圖一將進一步顯示不同超參數組合對訓練與驗證準確率的影響：



圖一、超參數準確率結果比較

(三)、在 Lab 2 中，您可能注意到訓練數據集與測試數據集的準確率存在差異。您認為造成這種現象的原因是什麼？討論可能導致**準確率差距**的潛在因素。

A：

● 根據題目所述，我認為的原因可能包括但不限於以下幾點：

1. **過擬合 (Overfitting) (最主要)**：模型在訓練數據上表現良好，但無法泛化至測試數據。通常發生在：
 - i. 模型太複雜。
 - ii. 訓練時間過長。
2. **測試集與訓練集分佈不同 (不同來源)**：如果測試集的數據分佈與訓練集不同，模型的泛化能力將受影響。(如：預測是否為貓咪，訓練集為自家貓咪照片，測試集卻為鄰居家貓咪照片)
3. **數據噪音 (Data Noise)**：如果訓練數據中存在錯誤的標註或是異常值，模型可能會學習到錯誤的特徵，進一步導致測試準確率的下降。
4. **過小的訓練數據集**：當訓練數據不足時，模型可能無法學習到一個穩定的狀態去判斷的模式，而僅記憶訓練數據，導致了準確率的差距。

● 解決方案包括但不限於以下幾種：

■ 使用正則化

- Dropout
- L2 Regularization

■ 增加訓練數據集

■ 調整超參數 (降低學習率) 等等。

(四)、討論在機器學習模型中選擇相關特徵 (feature selection) 的方法，特別是對於表格式 (tabular) 數據集的特徵選擇。強調**特徵選擇的重要性**及其對模型表現的影響。您可以參考外部資料來支持您的論點，請引用任何您參考的來源。

A：

- 存在表格式數據的機器學習模型中，**特徵選擇 (Feature Selection)** 是提升模型性能的關鍵步驟，其中最主要的方法包括以下幾種：
 - **基於統計的手法：**
 - 方差閾值 (Variance Threshold)：移除方差低的特徵。
 - 皮爾遜相關係數 (Pearson Correlation)：透過相關係數篩選與目標變數高度相關的特徵。
 - **基於模型的特徵選擇：**
 - **L1 正則化 (Lasso Regression)**：自動篩選重要特徵。
 - 決策樹/隨機森林 (Feature Importance in Tree-Based Models)。
 - **啟發式學習方法：**
 - 基因演算法 (Genetic Algorithm, GA)：尋找最佳特徵組合。
 - 深度學習中的注意力機制 (Self-Attention Mechanism, TabNet)。

特徵選擇能夠 **提高模型的解釋性、減少過擬合、加速訓練**，是表格式數據建模的重要步驟，參考自：【Shwartz-Ziv & Armon, 2022】。

(五)、雖然人工神經網絡 (ANN) 具有廣泛的適用性，但它未必是處理表格式數據的最佳選擇。請識別並描述一種更適合表格式數據的深度學習模型。解釋該模型專為表格式數據設計的原因，包括其關鍵特點和優勢。請確保引用您參考的外部來源。

A：

- 傳統 ANN 模型可以處理表格式數據，但可能不是最佳選擇，以下專門設計之深度學習模型可能可以更有效地處理表格式數據，詳列於下表二。

表二、Tabular 數據深度學習模型比較表

模型名稱	主要特點	適用場景	參考文獻
TabNet	<ul style="list-style-type: none">• 使用 注意力機制 動態選擇關鍵特徵• 解釋性強，適用於需要模型可解釋的應用	<ul style="list-style-type: none">• 中等規模數據 (如醫療、金融)• 對訓練時間要求不嚴苛的場景	Arik & Pfister, 2021
TabTransformer	<ul style="list-style-type: none">• 基於 Transformer 架構• 可學習 類別型特徵 關係• 自適應不同數據型別	<ul style="list-style-type: none">• 大規模表格式數據• 複雜類別特徵處理	Somepalli et al., 2021
DeepGBM	<ul style="list-style-type: none">• 結合 GBDT 與 DNN 優勢• 同時具備 解釋性與表達力	<ul style="list-style-type: none">• 金融風險預測、保險模型等應用• 需高精度預測任務	Ke et al., 2019

- 選擇建議
 - 小規模數據：使用 XGBoost、LightGBM 等傳統學習模型。
 - 中等規模數據：可考慮 TabNet。
 - 大規模數據：使用 TabTransformer 或 DeepGBM 等方法可能更具優勢。

参考文献

- Schwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90.
- Arik, S. Ö., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 8, pp. 6679-6687).
- Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., & Goldstein, T. (2021). Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*.
- Zhu, Y., Brettin, T., Xia, F., Partin, A., Shukla, M., Yoo, H., ... & Stevens, R. L. (2021). Converting tabular data into images for deep learning with convolutional neural networks. *Scientific reports*, 11(1), 11325.
- Ke, G., Xu, Z., Zhang, J., Bian, J., & Liu, T. Y. (2019, July). DeepGBM: A deep learning framework distilled by GBDT for online prediction tasks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 384-394).

資料參數意涵

- ✓ **age**：病患年齡 (以歲為單位)
- ✓ **sex**：性別 (男性/女性)
- ✓ **cp**：胸痛類型 (四種：低、中、高、嚴重)
- ✓ **trestbps**：靜息血壓
- ✓ **chol**：血清膽固醇 (單位：mg/dl)
- ✓ **fbs**：空腹血糖是否 > 120 mg/dl (1 = 是，0 = 否)
- ✓ **restecg**：靜息心電圖結果 (數值 0、1、2)
- ✓ **thalach**：最大心率
- ✓ **exang**：運動誘發心絞痛 (1 = 是，0 = 否)
- ✓ **oldpeak**：運動相對於靜息時的 ST 段下降幅度
- ✓ **slope**：運動 ST 段的坡度
- ✓ **ca**：透視檢查時染色的主要血管數量 (0-3)
- ✓ **thal**：3 = 正常，6 = 固定缺陷，7 = 可逆缺陷
- ✓ **target**：是否患病 (1 = 是，0 = 否)