# 1.

本次實驗選擇兩個關鍵超參數：Learning Rate (LR) 與 Hidden Layer Size (HS)，各自設置三種水準，分別為 [0.0005, 0.001, 0.01] 與 [64, 128, 256]，共產生 9 組組合。下表呈現 : 訓練 100 個 epoch，每種參數組合下 Train/Val/Test dataset中的 Loss 與 Acc。

一張含有 文字, 螢幕擷取畫面 的圖片

AI 產生的內容可能不正確。

# 2.

接續 #1.，我們將模型訓練與驗證的過程繪製成下列九張折線圖，並觀察不同參數組合下訓練集與驗證集上的 loss 與 accuracy 隨時間變化的趨勢，最後會進行超參數對模型效能影響之分析。

* 九種參數組合下的趨勢變化圖:
  + LR=0.0005、HS=64

一張含有 文字, 螢幕擷取畫面, 行, 繪圖 的圖片

AI 產生的內容可能不正確。

* + LR=0.0005、HS=128

一張含有 文字, 行, 字型, 筆跡 的圖片

AI 產生的內容可能不正確。

* + LR=0.0005、HS=256

一張含有 文字, 行, 字型, 繪圖 的圖片

AI 產生的內容可能不正確。

* + 一張含有 文字, 行, 字型, 繪圖 的圖片

    AI 產生的內容可能不正確。LR=0.001、HS=64
  + LR=0.001、HS=128

一張含有 文字, 行, 繪圖, 字型 的圖片

AI 產生的內容可能不正確。

* + LR=0.001、HS=256

一張含有 文字, 筆跡, 字型, 行 的圖片

AI 產生的內容可能不正確。

* + LR=0.001、HS=64

一張含有 文字, 行, 螢幕擷取畫面, 繪圖 的圖片

AI 產生的內容可能不正確。

* + LR=0.01、HS=128

一張含有 文字, 行, 繪圖, 字型 的圖片

AI 產生的內容可能不正確。

* + LR=0.01、HS=256

一張含有 文字, 行, 繪圖, 字型 的圖片

AI 產生的內容可能不正確。

* 學習率的影響
  1. 過小的學習率 (LR=0.0005) :

雖然訓練穩定，但收斂速度緩慢，且模型在訓練與驗證集上的準確率明顯偏低 (平均約 70%～74%)。由圖可見，Loss 雖緩步下降但最終停留在高原期，驗證集 Accuracy 曲線也波動劇烈 (如圖1~3)。

* 1. 適中的學習率 (LR=0.001) :

是本次實驗中表現最平衡的參數之一。其 Loss 曲線迅速下降後趨於穩定，驗證 Accuracy 穩定上升，最終可達 76%～79% 左右 (如圖4~6)，是泛化能力最佳的選項。

* 1. 過高的學習率 (LR=0.01) :

導致 Loss 曲線劇烈震盪甚至爆炸 (如圖7~9)，雖然模型能在訓練集上取得極高準確率 (86%↑)，但測試集準確率仍無法有效提升 (約 74%～77%)，明顯出現過擬合與學習不穩定現象。

* 隱藏層大小的影響
  1. 在相同學習率下，適度增加隱藏層大小有助於模型學習能力的提升 :

如，在 LR=0.001 下，隱藏層由 64 增加至 256 時，訓練 Loss 明顯下降，準確率提升約 6%，驗證 Accuracy 亦從 76.5% 提升至 79.0%，可見更大的模型容量有助於捕捉資料中的複雜模式。

* 1. 過大的隱藏層在高學習率下容易造成訓練不穩 :

如，LR=0.01 搭配 HS=256 時 (圖9)，Loss 雖迅速下降，但波動劇烈且驗證集準確率提升有限，顯示過度擴張模型並不一定有效。

* 整體最佳組合與建議
  1. 綜合上述曲線圖與 #1. 結果表格觀察，最佳組合為 LR=0.001 且 HS=256，能在訓練、驗證與測試集上取得平衡且穩定的表現。
  2. 若追求最高的測試準確率，則 LR=0.01 且 HS=256 組合可達 77.42%，但伴隨劇烈的 Loss 波動與不穩定性，須配合更嚴謹的正則化策略 (如 Dropout) 以避免過擬合。
* 本實驗結果顯示：學習率對模型收斂速度與訓練穩定性影響最劇，過小學習率導致學習緩慢、過大則易發散。隱藏層大小則影響模型容量與學習能力，但需搭配合適學習率才能發揮最佳效果。

# 3.

在訓練過程中，訓練集與測試集的 accuracy 不一致通常是由於過擬合造成的，模型在訓練集上學習到太多噪音與細節，而在未見過的測試集上表現不佳。此外，若資料分布存在差異 (distribution shift)、樣本不平衡或資料量不足，也會使得模型難以泛化，導致測試準確率較低。針對這些問題，可引入 dropout、early stopping 等正則化策略，以防止過度擬合並提高模型在新資料上的穩定性與表現。

# 4.

在表格型資料 (tabular data) 中選擇合適的特徵，能有效提升模型的準確度並降低過度擬合風險，也能減少運算成本。常見方法包含相關性分析，藉由計算特徵間相關係數以刪除重複或高度相關的欄位；遞迴式特徵消除 (Recursive Feature Elimination, RFE)，透過逐次移除影響較小的特徵來優化模型表現；以及基於樹模型的特徵重要性，如隨機森林或梯度提升演算法可衡量各特徵對預測的相對貢獻度。此外，互資訊 (Mutual Information) 等方法能捕捉資料間的非線性關係，使模型更具解釋力。整體而言，適當的特徵選擇有助於減少雜訊、提升效能並加速訓練。

* 參考資料 :
  + <https://blog.csdn.net/qq_40943760/article/details/134477728>
  + <https://ithelp.ithome.com.tw/articles/10264846>
  + <https://ithelp.ithome.com.tw/articles/10264846>

#5.

在處理表格型資料時，TabNet (由 Google Cloud 提出) 是一種深度學習模型的替代方案，能比傳統人工神經網路更有效率地應用於 tabular data。其核心理念結合了注意力機制 (Attention Mechanism) 與類似決策樹的稀疏選擇策略，讓模型在每個決策步驟中聚焦於最具判斷力的特徵，同時保持可解釋性。TabNet 透過多個「決策步驟」逐次萃取資料的重要訊息，並在運算過程中自動忽略噪音特徵，因而達到良好的泛化表現。相比一般的全連接式神經網路，TabNet 在不需要額外特徵工程的情況下，即可有效處理缺失值、非線性與類別特徵，並具備更強的可解釋度與穩定性。

* 參考文獻:
  + Arik, S. O., & Pfister, T. (2021). *TabNet: Attentive Interpretable Tabular Learning*. AAAI.
* 參考資料:
  + <https://blog.csdn.net/deephub/article/details/122535383>
  + <https://www.ithome.com.tw/news/139718>