

一、 超參數選擇： Hidden size, Learning rate (固定 Batch_size=32, epochs=100)

Hidden size	Learning Rate	Train Loss	Train ACC	Val Loss	Val ACC	Test ACC
64	1e-2	0.3177	86.7725%	0.5104	81.48%	77.419%
64	1e-3	0.3965	82.5379%	0.626	71.60%	58.065%
64	1e-4	0.5252	71.9577%	0.6178	71.60%	64.516%
128	1e-2	0.384	87.3016%	0.3491	80.2469%	70.97%
128	1e-3	0.4322	80.4233%	0.4074	82.72%	80.65%
128	1e-4	0.5026	75.6614%	0.6717	61.73%	64.516%
256	1e-2	0.3033	87.8307%	0.3732	86.42%	74.193%
256	1e-3	0.4119	81.4815%	0.4561	83.95%	80.645%
256	1e-4	0.5081	73.5450%	0.5735	76.54%	58.065%

二、

根據實驗結果，較大的 Hidden size (如 256) 整體表現最佳，特別是搭配學習率 1e-3 時，在驗證與測試集上皆達到約 80% 的準確率，顯示模型有良好的學習能力與泛化性。相較之下，學習率太大 (1e-2) 雖訓練快速但易過擬合，太小 (1e-4) 則學習不足。整體而言，Hidden size = 256 且 Learning rate = 1e-3

是效果最穩定且準確的組合。結果亦可用曲線圖顯示不同組合的 loss / accuracy 變化趨勢。

三、

我有發現在 Lab 2 中，訓練集與測試集準確率之間的差異可能源自於模型過度記住訓練資料（過擬合），導致泛化能力下降。此外，資料量有限、資料分布不均或模型複雜度過高，也都可能造成模型在測試集表現較差。若模型在訓練集上準確率高，但在測試集中比較低，代表它沒有學到通用的規則。

為了能夠改善此情況，我認為可以加入 Dropout、使用 Early Stopping 或減少模型參數等方式，提升泛化能力。

四、

特徵選擇是提升模型效能與泛化能力的關鍵步驟。常見方法包括：根據統計指標（如皮爾森相關係數）篩選特徵、使用樹模型（如 Random Forest）的特徵重要性排序，以及基於模型表現的遞迴特徵消除（RFE）。良好的特徵選擇能降低維度、減少過擬合、提升訓練效率與預測準確度。此外，它也有助於模型解釋性，讓我們更了解哪些變數真正有助於預測結果。

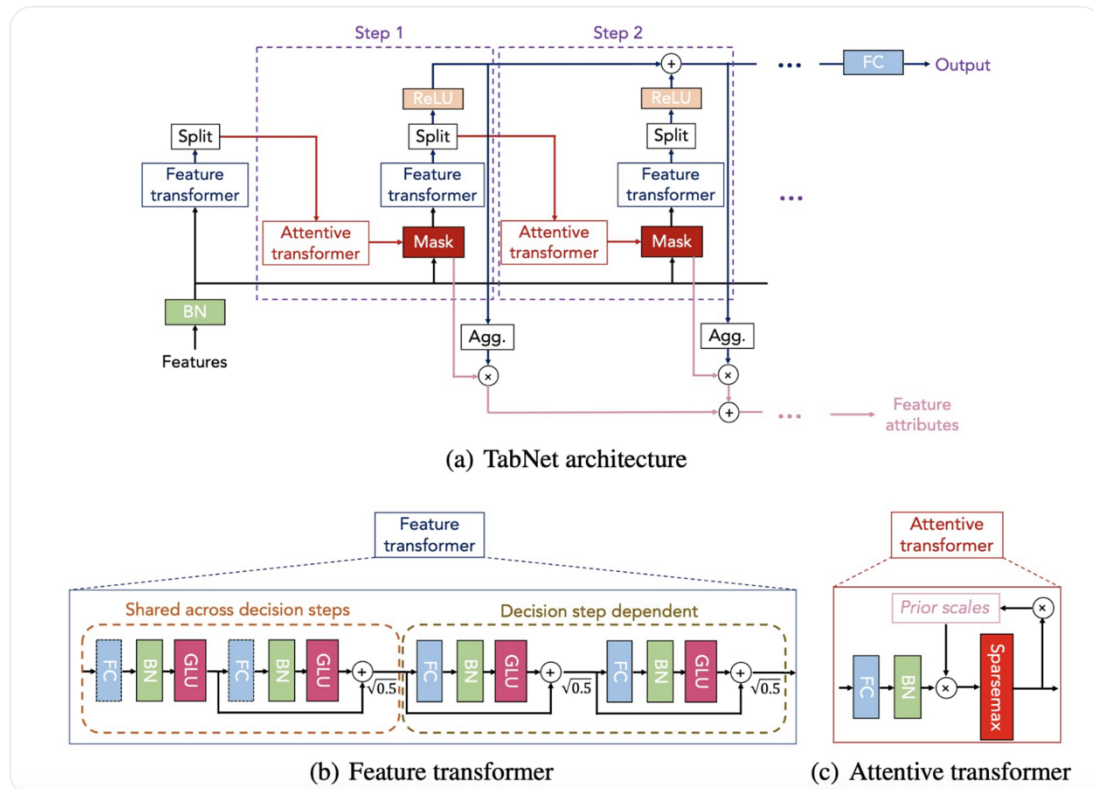
Reference: <https://machinelearningmastery.com/an-introduction-to-feature-selection/>

https://scikit-learn.org/stable/modules/feature_selection.html

五、

在處理表格型資料（**Tabular Data**）時，傳統的人工神經網路（**ANN**）雖然具備強大的表達能力，但經常因無法有效捕捉特徵間的結構與意義，而無法達到最佳效果。為了解決這個問題，**Google Cloud AI** 提出了 **TabNet**，這是一種專門為表格資料設計的深度學習模型。

TabNet 的核心創新在於其序列式注意力機制（**Sequential Attention Mechanism**），這讓模型能夠在每一個決策步驟中動態地選擇「看哪些特徵」，就像人類做判斷時會依據不同情境關注不同資訊一樣。這種做法不僅讓模型能學到稀疏、可解釋的特徵表示，也有助於避免資訊冗餘與過擬合。此外，**TabNet** 還融合了類似樹模型的局部決策能力與深度學習的端對端優勢。在多個基準資料集上，**TabNet** 不僅表現優異，還能顯著降低對特徵工程的依賴，並提供可視化的特徵貢獻解釋，特別適用於金融、醫療等需要模型透明度的應用場景。



Reference: <https://arxiv.org/abs/1908.07442>

<https://cloud.google.com/blog/products/ai-machine-learning/improved-tabnet-on-vertex-ai>