

Homework 2

學號：113034510 姓名：蘇謹

1. 選擇測試的 hyper parameter 為 hidden units 及 hidden layers。設定的數值分別為 hidden units = { 256, 512, 1024 }，hidden layers = { 5, 10, 50 }。接著實驗了九組數據，每組最後訓練 100 Epoch 結束之 Train、Val、Test 的 Loss、Accuracy 結果如下表所示：

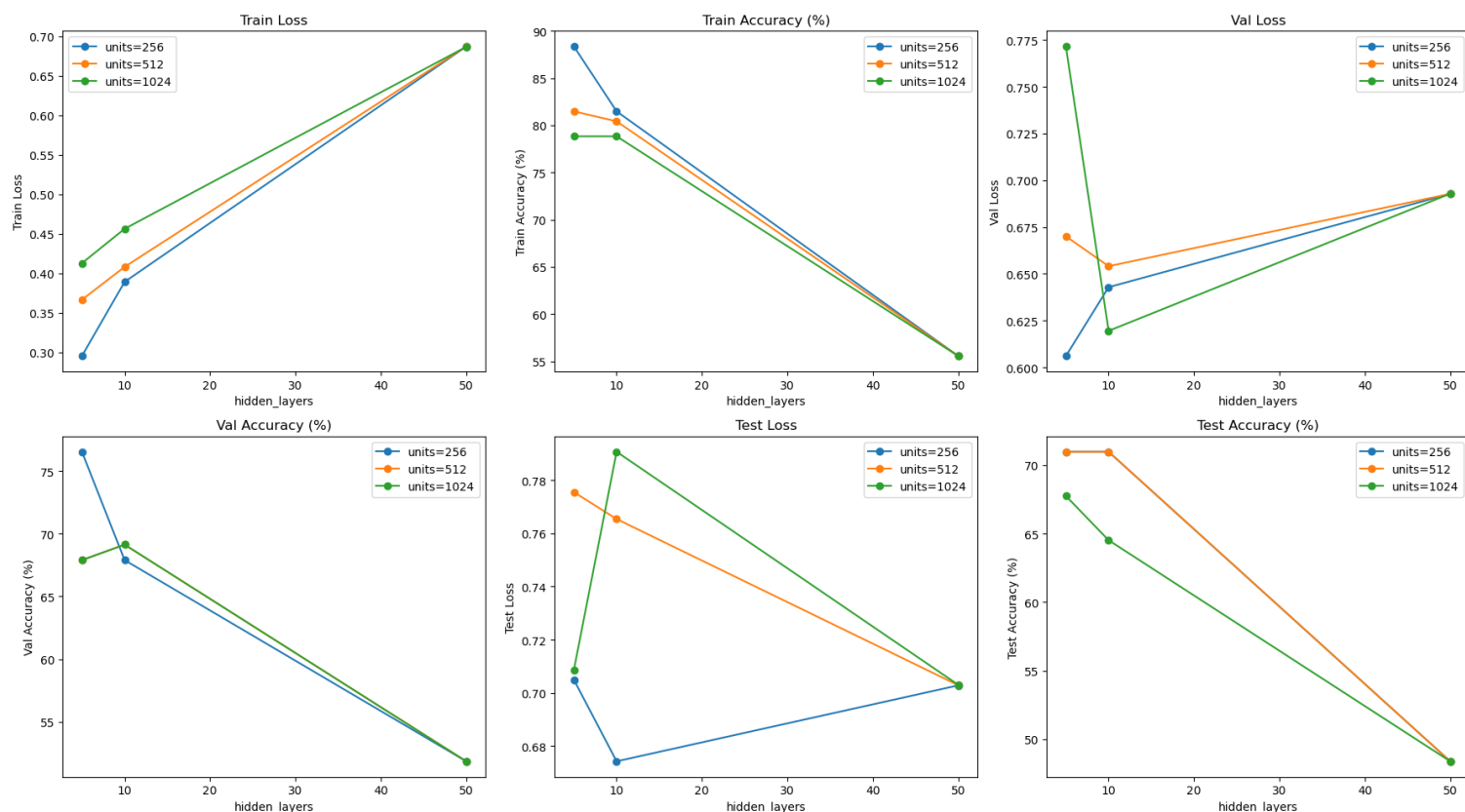
Hidden Units	Hidden Layers	Train Loss	Train Accuracy(%)	Val Loss	Val Accuracy(%)	Test Loss	Test Accuracy(%)
256	5	0.295	88.360	0.606	76.543	0.705	70.968
256	10	0.389	81.481	0.643	67.901	0.674	70.968
256	50	0.687	55.556	0.693	51.852	0.703	48.387
512	5	0.367	81.481	0.670	67.901	0.776	70.968
512	10	0.408	80.423	0.654	69.136	0.765	70.968
512	50	0.687	55.556	0.693	51.852	0.703	48.387
1024	5	0.413	78.836	0.771	67.901	0.709	67.742
1024	10	0.456	78.836	0.620	69.136	0.791	64.516
1024	50	0.687	55.556	0.693	51.852	0.703	48.387

2. 根據上表結果可以發現，當 hidden layers = 5 時，各項指標均表現較好；當 hidden layers = 10 時，模型表現普遍明顯下降；hidden layers = 50 時，模型在所有數據集上均表現最差。此外，可以發現當 hidden units 從 256 增加到 512 或 1024 時，模型的最佳表現並未有顯著提升。

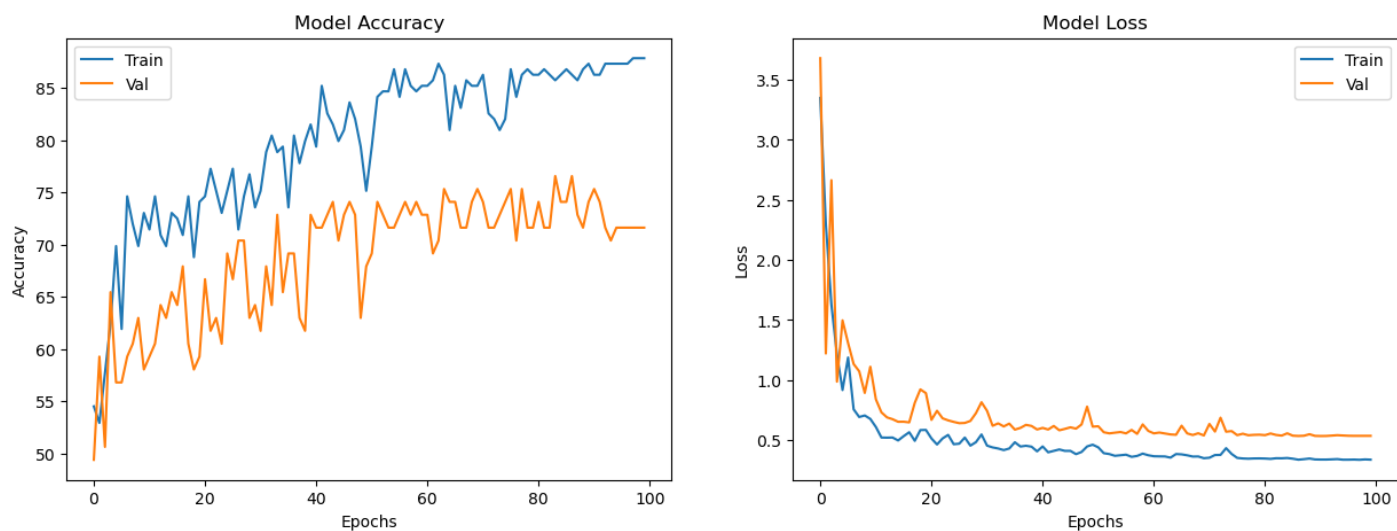
綜合結果來看，訓練架構的深度過高會影響模型效能，反而出現過擬合或訓練不穩定的現象，並且 hidden units 的改變可能沒有太大幫助。

另外參考下圖的各資料 Loss 可以發現受到 hidden layers 的影響劇烈，Train Loss 會隨 hidden layers 增加，並參考 Train Loss、Val

Loss、Test Los 的數據就可以知道，訓練過程中可能出現梯度消失或梯度爆炸問題，並且有嚴重的 Overfitting。



3. 主要原因應主要來自模型的 **Overfitting**，正如同下圖所示。模型可能在 Train data 上學到了過多細節，包括 noise 與 outlier。此外，與 Test data、Val data 分布不同也是有可能的原因，因此可能需要更大量的資料量，或是一些 **data Augmentation** 的手法



4. 特徵選擇對於模型的重要性在於可以減少 **Overfitting**、提升模型解釋性、減少計算成本並改善模型精度，主要有三種類型的方
法：

- A. **過濾法 (Filter Method)**：根據統計測試（如卡方檢驗、相關性分析）來選擇特徵。這些方法是獨立於模型的，並且計算快速。
- B. **包裝法 (Wrapper Method)**：使用特定的機器學習模型來評估特徵子集，這些方法通常能提供較好的預測性能，但計算量較大。
- C. **嵌入法 (Embedded Method)**：在模型訓練過程中進行特徵選擇，如 Lasso 回歸或決策樹方法，這些方法同時進行特徵選擇和模型訓練。

藉由以上的方法就可使用少量又快速的計算成本、並且提供具有解釋意義(統計方法等等，能用公式推導的算法)的特徵篩選方式，更重要的是篩選出的特徵除了有意義外，是可以在最後的模型訓練中降低多餘的計算成本，甚至篩選掉無意義、含有雜訊的不重要因子。例如主成分分析 (PCA) 就是很好的例子，對數據進行降維，保留主要特徵，消除冗餘特徵。

(Reference : GeeksforGeeks. (2025). Feature Selection Techniques in Machine Learning. <https://www.geeksforgeeks.org/feature-selection-techniques-in-machine-learning/>)

5. TabNet 是適合處理表格數據的深度學習模型。

TabNet 採用了 **sequential attention mechanism**，這代表著它會根據每一步的學習過程自動選擇最重要的特徵進行處理，這對於表格數據中的多變量特徵非常有效。其中有兩個很重要的設計：

- A. **決策機制**：根據當前的學習階段，動態選擇重要的 **sparse** 特徵。
- B. **共享神經網絡**：處理這些選中的特徵，並進行進一步的推理。

TabNet 的優勢在於它能夠減少過擬合，並且在模型訓練過程中實現特徵選擇，進而提高了模型的解釋性。相比於傳統的 **ANN** 模型，TabNet 在處理表格數據時通常能夠提供更好的性能，並且需要較少的計算資源。

(Reference : Arik, S. Ö., & Pfister, T. (2021). TabNet: Attentive Interpretable Tabular Learning. Proceedings of the 37th International Conference on Machine Learning.

<https://openreview.net/forum?id=BylRkAEKDH>)