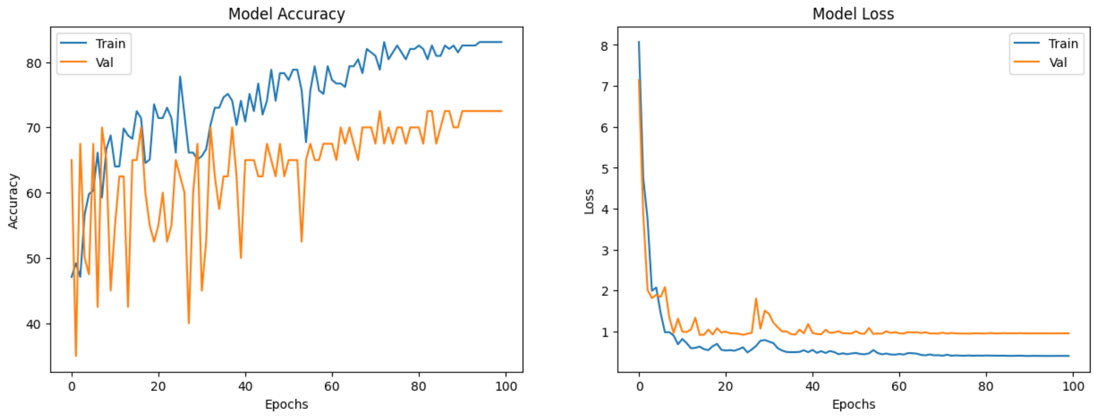
1. (20 pts) Select 2 hyper-parameters of the artificial neural network used in Lab 2 and set 3 different values for each. Perform experiments to compare the effects of varying these hyper-parameters on the loss and accuracy metrics across the training, validation, and test datasets. Present your findings with appropriate tables.

Hyperparameter	Test values
Learning Rate	0.0001, 0.001,0.01
Hidden Units	64, 128, 256

Learning Rate	Hidden Units	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
0.01	64	0.42	79.90%	0.6881	72.50%	0.62	64.52%
0.01	128	0.39	80.95%	0.6532	75.00%	0.58	74.19%
0.01	256	0.36	83.60%	0.6238	77.50%	0.56	77.42%
0.001	64	0.39	79.37%	0.6312	72.50%	0.60	67.74%
0.001	128	0.38	83.60%	0.8158	75.00%	0.63	74.19%
0.001	256	0.44	79.37%	0.7845	77.50%	0.69	77.42%
0.0001	64	0.42	82.01%	0.6811	77.50%	0.63	70.97%
0.0001	128	0.39	82.54%	0.6733	75.00%	0.63	67.74%
0.0001	256	0.45	78.84%	0.9493	72.50%	0.73	64.52%

- 較大的 hidden units（256）整體表現比較好，尤其在 learning rate = 0.01 和 0.001 時，256 都是 accuracy 最高的組合（77.42%）。
- Learning rate = 0.01 和 0.001 表現普遍比 0.0001 好，從表格中可見最低的 test accuracy 出現在 Learning Rate = 0.0001 且 Hidden Units = 256 的組合上（64.52%），代表學習太慢可能會影響學習效果。
- Test loss 與 accuracy 整體趨勢一致
  表中大部分 accuracy 較高的組合，其 Test Loss 也較低，表示沒有 overfitting 的嚴重情況。
- 當 learning rate 為 0.01 且 hidden units 為 256 時，模型測試準確率最高（77.42%），為整體表現最佳的組合。對於 learning rate 為 0.01 和 0.001 的情況，hidden units 較少時模型表現略差，可能因容量不足影響學習。而在 learning rate 為 0.0001 時，hidden units 增加反而讓模型表現下降，推測是學習率過小導致訓練不足或 overfitting，反而是小模型較容易收斂、效果更好。

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.)



根據實驗結果，模型效能隨著超參數的不同產生顯著變化。從表格可見，當隱藏單元數為 256 且學習率為 0.01 時，模型在訓練、驗證及測試集上的表現最佳（Validation Accuracy 為 77.50%、Test Accuracy 為 77.42%）。顯示增加隱藏層單元能提升表現能力。同時，適當的學習率亦為關鍵因素；相較於 0.0001，當學習率為 0.01 與 0.001 時模型更快收斂，準確率明顯提升，而過小的 learning rate 則可能導致收斂速度緩慢甚至停滯。由上圖可知，隨著 Epochs 上升時準確率穩定成長、Loss 明顯下降，顯示模型成功學習並趨於收斂。

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.)

造成訓練集與測試集準確率差異的原因有幾個，其中一個常見情況是過擬合。模型在訓練過程中容易記住資料中的雜訊或特定模式，導致無法良好泛化至測試資料。以結果來看，當隱藏層為 256 且學習率過小（如 0.0001）時，訓練準確率達 78.84%，但測試準確率僅為 64.52%，即為典型的過擬合現象。然而，若搭配適當的學習率（如 0.01），測試準確率可提升至 77.42%。此外，若訓練與測試資料分布不同，或缺乏如 dropout、early stopping 等正則化手法，也會加劇落差。最後，資料量不足或分布不均亦會影響泛化能力，因此需在模型複雜度與資料多樣性之間取得平衡。

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.)

在表格型資料中進行特徵選擇能有效提升模型效能、降低過擬合風險並加快訓練速度。常見方法包括：**過濾法**（Filter，如皮爾森相關係數、互資訊）、**包裝法**（Wrapper，如遞迴特徵消除 RFE）與**嵌入法**（Embedded，如 Lasso 或決策樹的重要性評估）。良好的特徵選擇有助保留具資訊價值的變數，移除冗餘與雜訊特徵，提升模型泛化能力與預測準確率。

參考文獻：

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.

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.)

雖然人工神經網絡（ANN）具備高度彈性，但在處理表格資料時，其效能與可解釋性往往有限。TabNet 是專為結構化資料設計的深度學習架構，透過逐步注意力機制（Sequential Attention）動態聚焦於最具資訊量的特徵，有效提升模型表現。其稀疏特徵遮罩（Sparse Feature Masking）只選取部分有意義的特徵參與訓練，降低過擬合風險並提升可解釋性。TabNet 還支援特徵重要性視覺化，便於理解模型行為，並具端對端特性，簡化資料前處理流程。實驗證明 TabNet 在多個公開表格資料集上皆表現優異，甚至超越傳統機器學習方法，特別適合應用於金融、醫療等需高準確率與解釋性的領域。

參考資料：

Arik, S. O., & Pfister, T. (2020). TabNet: Attentive Interpretable Tabular Learning. *arXiv preprint arXiv:1908.07442.*

Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). *Revisiting deep learning models for tabular data*. Advances in Neural Information Processing

Systems, 34, 18932–18943.