深度學習作業二

1. 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.

   Setting hyper-parameter : learning rate = [0.01, 0.001, 0.0001]

   Setting hyper-parameter : batch size = [32, 64, 128]

learning rate	batch size	Train loss	Train accuracy	Val loss	Val accuracy	Test loss	Test accuracy
0.01	32	0.22629	91.5343	0.70506	77.7777	0.41851	74.1935
0.01	64	0.25145	89.9470	0.64547	76.5432	0.43500	77.4193
0.01	128	0.26944	91.5343	0.58110	76.5432	0.45543	80.6451
0.001	32	0.30017	91.5343	0.56248	77.7777	0.48328	77.4193
0.001	64	0.37429	86.2433	0.63101	74.0740	0.56579	70.6977
0.001	128	0.45701	78.3068	0.60006	71.6049	0.63636	67.7419
0.0001	32	0.48181	76.1904	0.61834	70.3703	0.63100	67.7419
0.0001	64	0.51741	73.5449	0.58365	76.5432	0.65299	67.7419
0.0001	128	0.55006	69.8412	0.60926	74.0740	0.68595	61.2903

2. 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.

   - Learning rate 對表現的影響

     較高的 Learning rate (0.01)，Train accuracy 穩定維持高水準使得收斂快速，Val & Test Accuracy 亦穩定維持在 75%~80%。同時 Train & Test loss 也維持在較低的表現。

     隨著 Learning rate 降低，Train & Val & Test Accuracy 明顯降低，且 Train & Test loss 明顯上升，具有 underfitting 的現象。

     由此可見，learning rate 太低(0.0001)會導致模型學習過慢甚至訓練不足，較高的 learning rate(0.01)較佳。

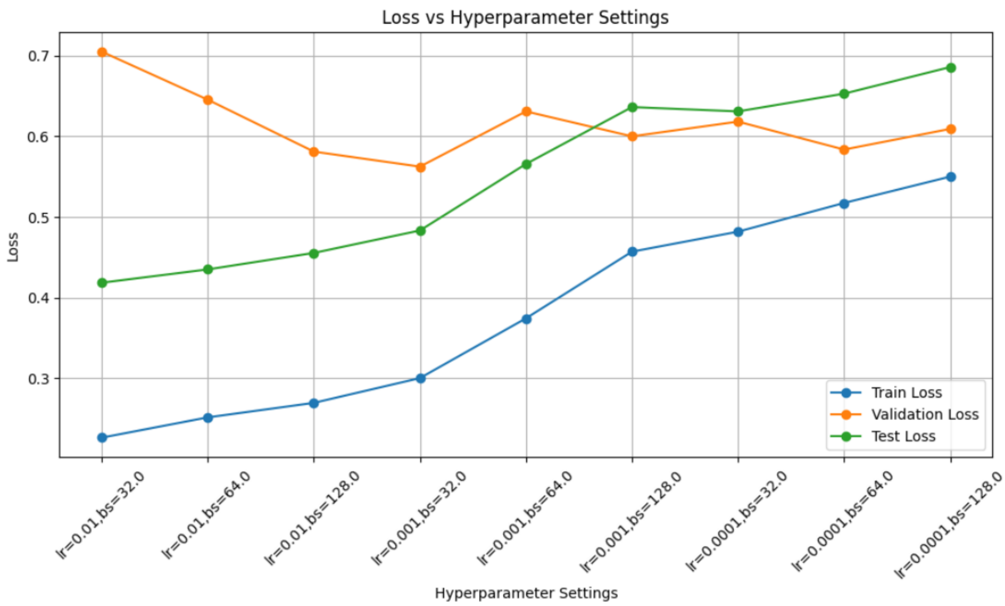
- Batch size 對表現的影響

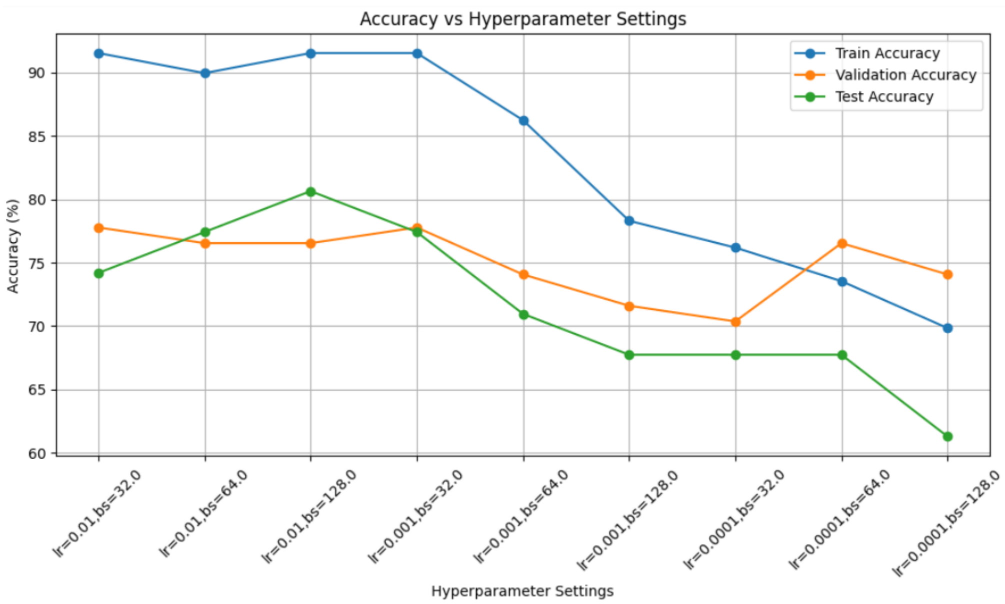
  learning rate = 0.01 下，batch size = 128 時 Train & Test Accuracy 些微較高，batch size 較小時準確率的表現則相較較差。

  learning rate = 0.001/0.0001 下，batch size 越高則 Train & Val & Test loss 有明顯的上升，accuracy 亦有明顯的下降。

  由此可見，當 learning rate 大時，batch size 對表現的影響不明顯，但隨著 learning rate 下將，過大的 batch size 容易導致收斂效果較差。

- 結論: 當 learning rate=0.01、batch size=128 為最佳化表現的組合。





3. 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.

- Overfitting
  當 learning rate=0.01 時，Train & Test accuracy 尤其大，顯示模型在訓練集學習過多細節或噪聲，導致在未見過的測試及上表現明顯較差。
- Dataset 不平衡
  train set 資料量(270)明顯大於 test set 資料量(31)，理想情況下訓練集資料量:測試及資料量應為 4:1，可能造成 sample bias，模型在訓練時過度擬合到 train set 的特徵分布，導致泛化到 test set 時 accuracy 降低。
- 資料量不足
  training set 資料量有限，模型會傾向記憶訓練資料而非學到模式。

4. 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.

- Filter Methods
  常用技術包含: 相關係數、卡方檢定、互信息等。優點為計算速度快，適合初步快速過濾不重要特徵，缺點為無法考慮特徵間的交互作用。
- Wrapped Methods
  常用技術包含: RFE、Forward Selection、Backward Elimination。優點為考慮特徵間的交互作用，缺點為計算成本高，尤其在高維資料上，訓練時間會大幅增加。
- Embedded Methods
  常用技術包含: 正則化、Decision Tree、Random Forest、XGBoost 等數模型重要性分析。與模型緊密結合，選出的特徵通常對該模型效果最佳。

Feature selection 的重要性在於，利用刪減無關或多餘的特徵，可以提升模型泛化能力、減少模型複雜度，並提高計算效能。

參考資料: Guyon, I., & Elisseeff, A. (2003). "An introduction to variable and feature selection." Journal of Machine Learning Research, 3(Mar), 1157-1182.
參考資料: Chandrashekar, G., & Sahin, F. (2014). "A survey on feature selection methods." Computers & Electrical Engineering, 40(1), 16-28.

5. 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.

   表格數據的特性為結構化的，使用 ANN 處理時，ANN 基於 dense layer，每一層會無條件將所有特徵結合，容易導致不必要的特徵干擾，也很難像樹模型一樣「自動挑選重要特徵」。因此針對表格型資料，Tabnet 處理更適合。

   TabNet 結合了決策樹與深度學習的優點。使用序列注意力機制，逐步篩選重要特徵，並允許模型在不同決策階段關注不同的特徵子集，每一步都選取與決策最相關的少量特徵。

   TabNet 關鍵特典包含: 特徵選擇能力強、具備可解釋性、類似樹模型，無需手動 one-hot encoding 類別型特徵，可間容樹執行與類別型資料。

   參考資料: Arik, S. Ö., & Pfister, T. (2019). TabNet: Attentive Interpretable Tabular Learning. arXiv preprint arXiv:1908.07442.