National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name:李宜庭 Student ID: 112030751

Due on 2025.03.27

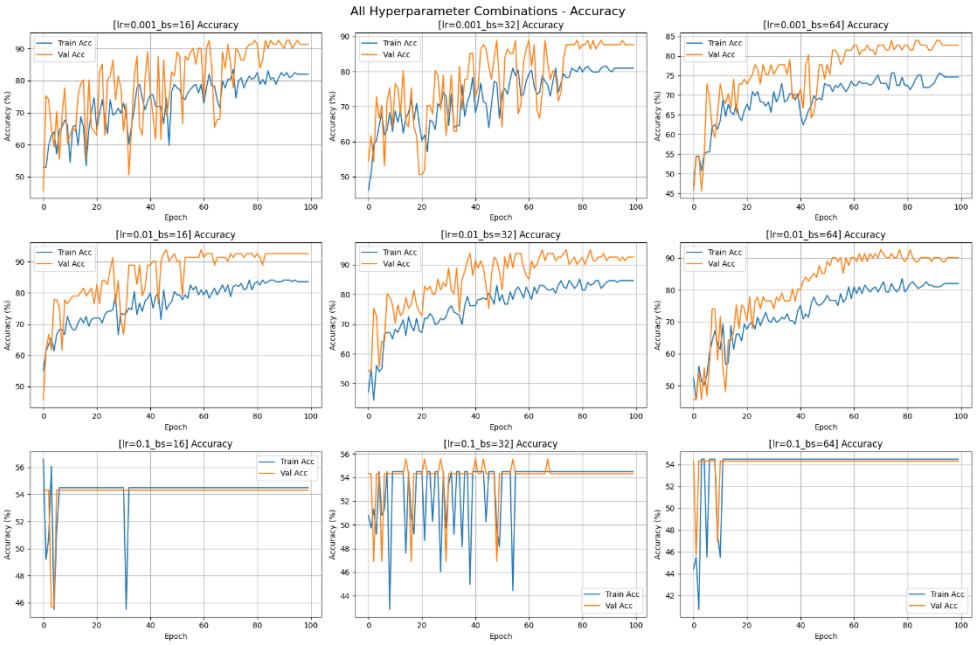
1. (20 pts) Select 2 hyper-parameters of the artificial neural network used in Lab 2 and set 3 different values for each. Perform experiments to compare the effects of varying these hyper-parameters on the loss and accuracy metrics across the training, validation, and test datasets. Present your findings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

實驗參數組合：Learning Rate(0.1, 0.01, 0.001), Batch Size(64, 32, 16)

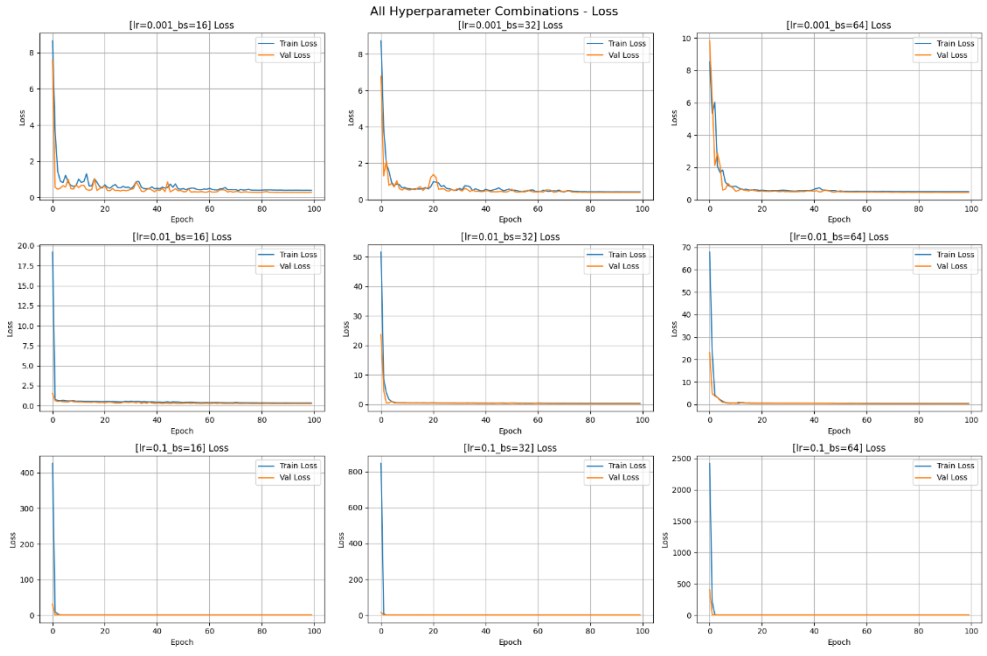
	Learning Rate	Batch Size	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss	Test Accuracy	Test Loss
0	0.001	16	83.60	91.36	0.3725	0.2712	83.87	0.4507
1	0.001	32	79.89	87.65	0.4317	0.3916	70.97	0.5476
2	0.001	64	78.31	83.95	0.4697	0.4884	70.97	0.5842
3	0.010	16	83.60	92.59	0.3396	0.2227	77.42	0.4424
4	0.010	32	80.95	93.83	0.3862	0.2924	70.97	0.4672
5	0.010	64	83.60	86.42	0.3421	0.2980	80.65	0.5088
6	0.100	16	54.50	54.32	0.6885	0.7044	48.39	0.6999
7	0.100	32	82.54	87.65	0.4502	0.3163	74.19	0.4671
8	0.100	64	54.50	53.09	0.6893	0.6913	48.39	0.7000

當 Learning Rate 為 0.01 或 0.001 時，模型在Training Set與Validation Set上表現穩定；但當 lr 提高至 0.1，模型表現不穩定、難以收斂 ; Test Set 部分，lr = 0.001, bs = 16 達最高 Test Accuracy（83.87%）。

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.)



圖一：不同 Hyperparameter 組合之 Accuracy Curve



圖二：不同 Hyperparameter 組合之 Loss Curve

由圖一與圖二，當 lr 為 0.001 或 0.01 時，Train 與 Val ACC Curve 表現穩定，模型能有效收斂；相對地，lr = 0.1 時，模型之ACC Curve震盪較劇烈、Loss Curve在初期極高、不穩定。整體而言，最佳組合為「lr = 0.001、bs = 16」，Train/Val/Test 三者差異小，且Test Accuracy最高、Test Loss僅次於 「lr=0.01, bs=16」。

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.)

在以上實驗中，Training 與 Test Set Accuracy 的差異可能來自於 overfitting。以 lr=0.01, bs=16 為例，其 Validation Accuracy 雖然很高，但 Test Accuracy 明顯下滑，顯示模型可能對 Validation Set overfit。此外，資料量有限、缺乏Regularization等機制等也可能是影響因素。

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.)

針對 Tabular Data，常見的 Feature Selection 方法包含 filter-based approaches、wrapper methods 與 embedded methods。Filter-based 方法與模型無關，透過統計量（如相關係數）篩選特徵，速度快但未考慮與模型的關聯性；Wrapper 方法則依據模型表現評估特徵組合，準確度高但計算成本較大；Embedded 方法則在模型訓練過程中同時進行特徵選擇，具備良好效率。透過移除不相關或多餘的欄位，能使模型更簡潔，有助於降低 overfitting問題，並提升模型可解釋性與泛化能力。

References：*Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.*

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.)

TabNet是一種專為Tabular Data Set打造的深度學習架構，透過**Sequential Attention機制**，在每個decision step中選擇與當前樣本最相關的特徵，模擬tree-based 模型的特徵選擇邏輯。其主要特色包含**sparse feature selection**、**instance-wise masking**以及**built-in interpretability**，使模型能針對樣本聚焦於最具資訊代表性的特徵子集，因此特別適用於中小型Tabular Data Set，且相較於傳統 ANN 參數更少、模型泛化能力更強。

References：*Arik, S. Ö., & Pfister, T. (2019). TabNet: Attentive Interpretable Tabular Learning*