National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name:方品心　　　　　Student ID:312704005

Due on 2024.03.21

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.

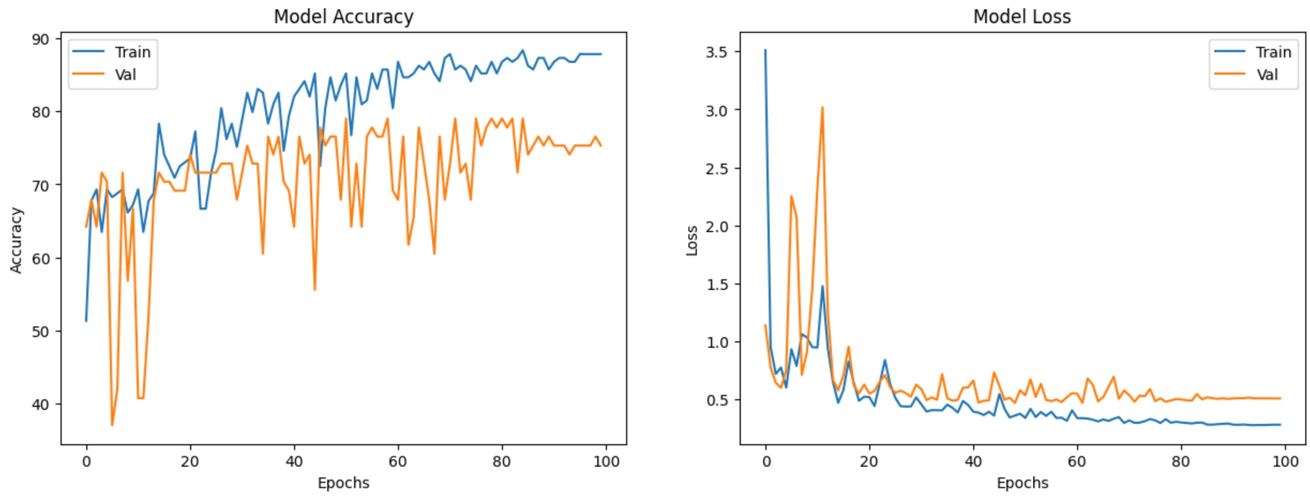
	Train loss	Train acc	Best Val loss	Best Val acc	Test acc
batch_size = 8	0.2830	87.83%	0.4706	79.01%	77.42%
batch_size = 16	0.3180	85.19%	0.3980	82.72%	70.97%
batch_size = 32 (原始)	0.4228	84.66%	0.4250	87.65%	74.19%
batch_size = 128	0.4124	85.71%	0.6081	72.84%	61.29%

	Train loss	Train acc	Best Val loss	Best Val acc	Test acc	Run time
epoch = 10	0.5776	68.25%	0.6062	71.60%	58.06%	0.78 secs
epoch = 20	0.5896	69.84%	0.4178	83.95%	64.52%	0.92 secs
epoch = 50	0.4445	79.37%	0.4593	80.25%	70.97%	1.32 secs
epoch = 100 (原始)	0.4228	84.66%	0.4250	87.65%	74.19%	1.86 secs

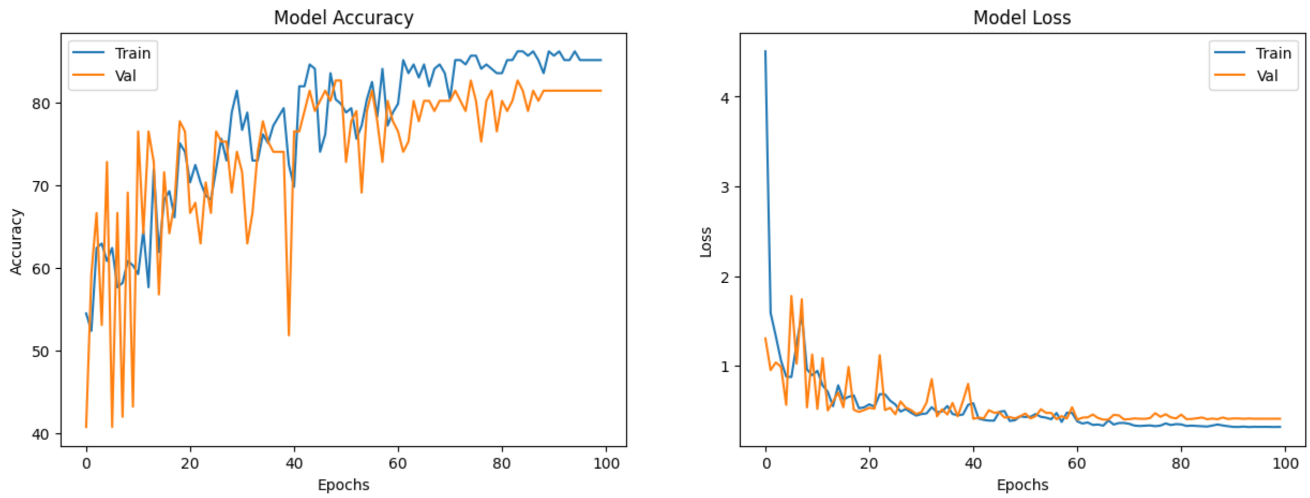
2. (20 pts) Based on your experiments in Question 1, analyze the outcomes. What differences do you observe with the changes in hyper-parameters? Discuss whether these adjustments contributed to improvements in model performance, you can use plots to support your points. (Approximately 100 words.)

在問題一我分別調整了超參數 batch_size 以及 epoch 的大小，接下來我會各自探討其影響。

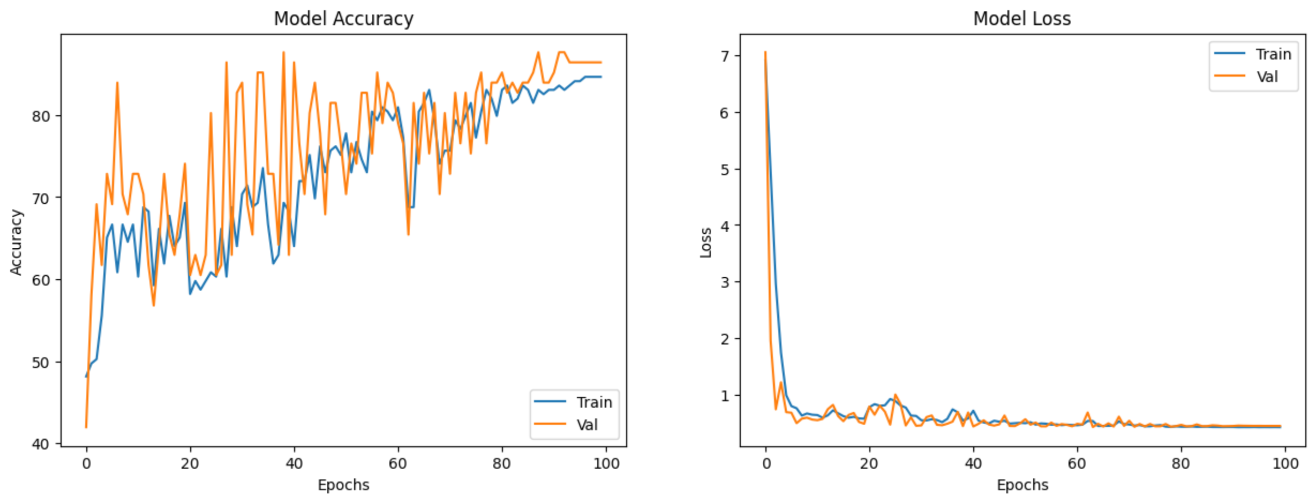
1. **Batch Size**：
    - 較小的 batch size（8、16）導致了更高的訓練準確率，但是驗證集上的性能相對較差。
    - 較大的 batch size（128）導致了較差的模型性能，尤其在驗證集和測試集上的表現明顯下降。
    - 在選擇 batch size 的大小時，太大或太小都不適當，過小的 batch size 可能導致模型難以收斂，而過大的 batch size 可能導致模型的泛化能力下降。
    - 在此模型中 batch size = 32 較為合適。



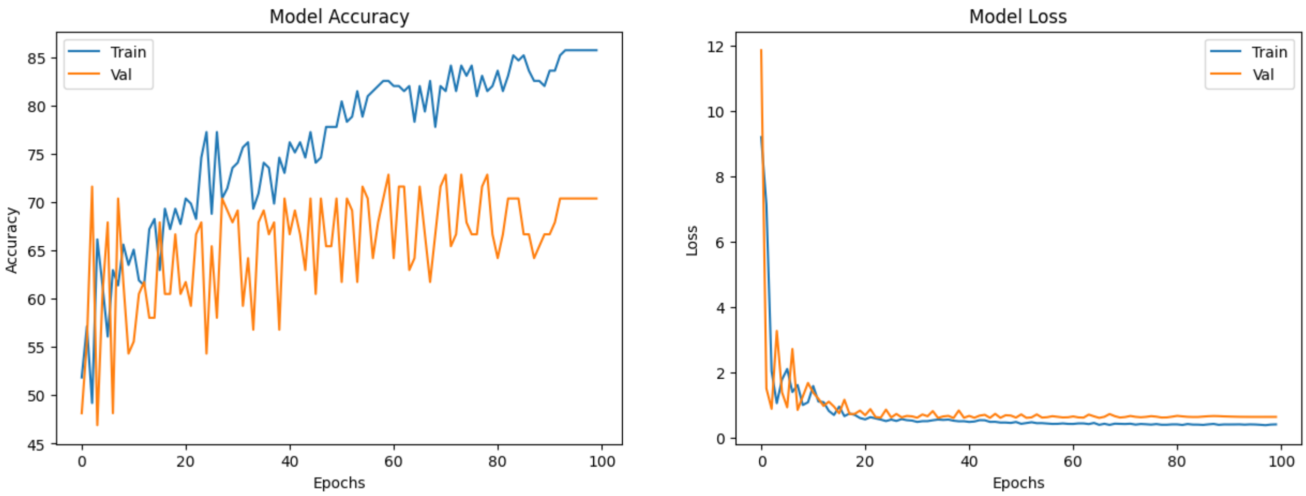
圖一：batch_size = 8



圖二：batch_size = 16



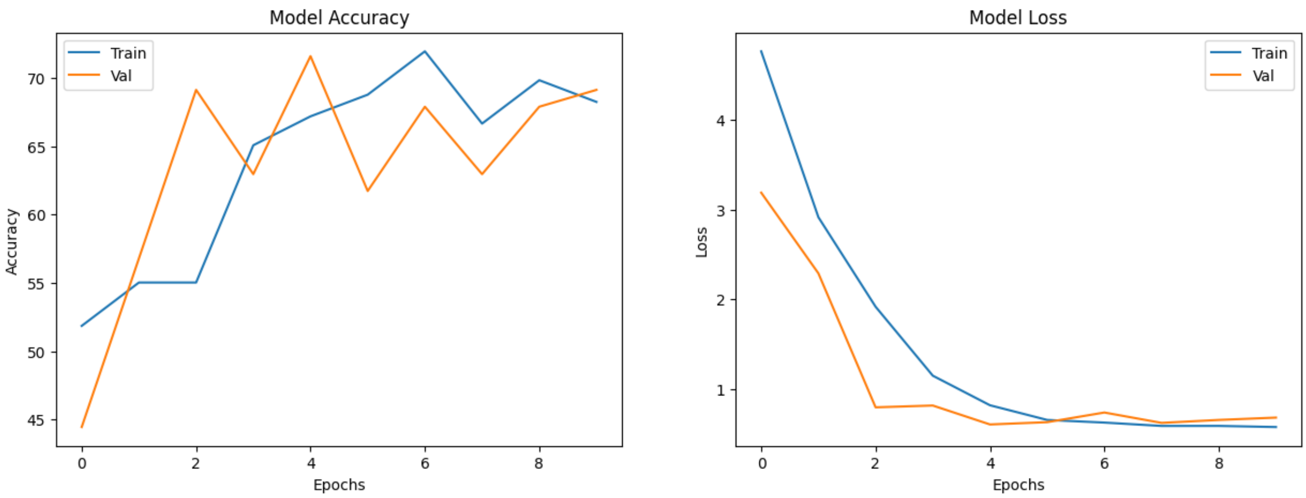
圖三：batch_size = 32



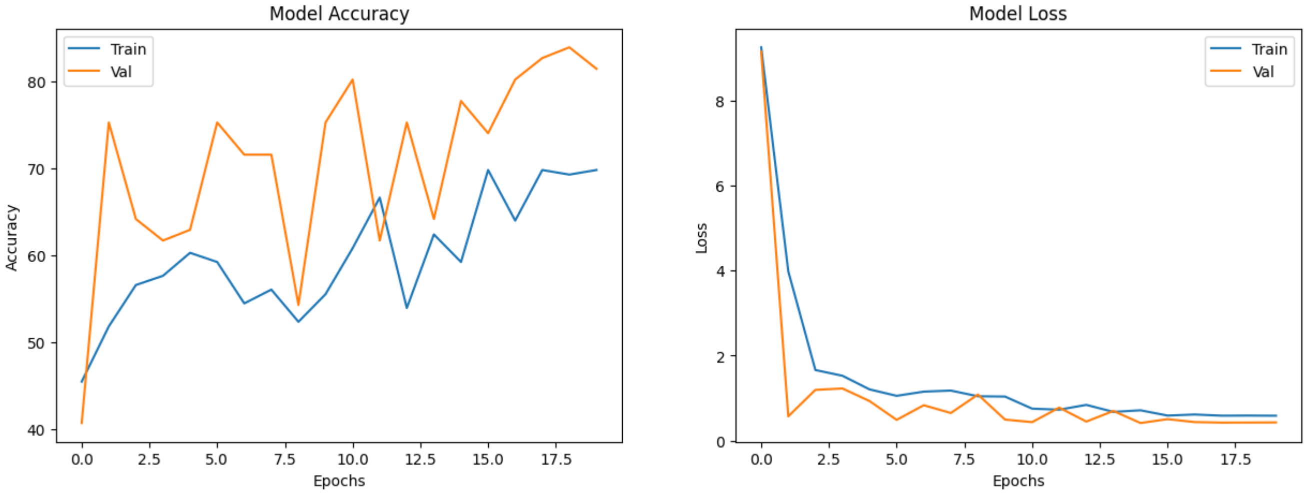
圖四：batch_size = 128

2. **Epoch**：

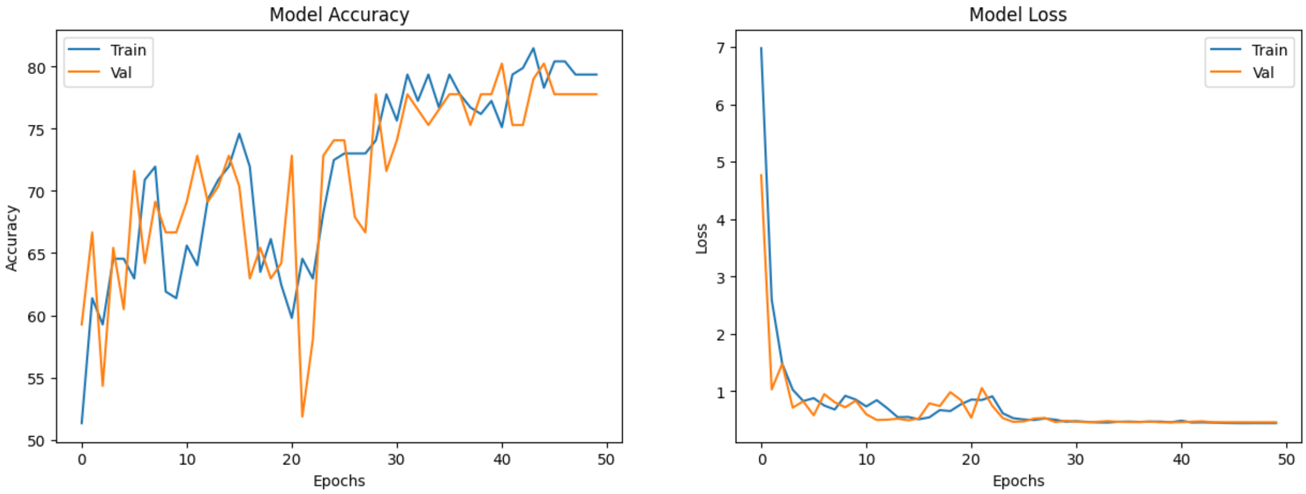
- 當 epoch 數設定的太小 (10、20、50)，模型可能還沒有充分學習到數據中的模式和特徵，無法捕捉到其中較複雜的結構，從而導致訓練和測試準確率都較低。
- 相對的，隨著 epoch 數的增加，訓練準確率和測試準確率通常會提高，但是如果太多，可能會使模型記住訓練集的細節，導致過擬合，而無法泛化到新的數據上。
- epoch = 100 過擬合的狀況還沒那麼明顯，相對於 10、20 和 50 欠擬合的情況，此設定更適合一點。
- 除此之外，增加 epoch 數會增加訓練時間，如同我在第一題的表格中整理的一樣。雖然這次的網路相對較簡單，體感上沒有太大的差別，但若應用在非常深、非常複雜的網路中，運行時間和準確率將成為一個trade-off，是需要仔細考慮的問題。



圖五：epoch = 10



圖六：epoch = 20



圖七：epoch = 50

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.)

   1. **Overfitting**：模型可能過度學習訓練數據，導致在訓練集上有很高的準確率，但對未見過的數據泛化能力較差。
   2. **Data Mismatch**：訓練集和測試集之間的可能存在差異，例如分佈不同或有噪音存在，導致性能表現有所不同。
   3. **Model Complexity**：如果模型相對於數據集的大小過於複雜，則可能會捕捉到噪音或無關(不重要/不正確)的規律，導致過擬合。
   4. **Hyperparameter Tuning**：超參數設置不好可能導致模型在訓練數據上表現不錯，但無法泛化到新數據上。

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.)

特徵選擇對於提升機器學習模型的性能和可解釋性至關重要，其目標是找到一個最好的特徵子集，替除掉一些不相關的特徵，進而達到提高模型精確度，減少運行時間等目的。

目前常見的方法有以下幾種：

1. Filter Methods：根據統計指標（如相關性、方差等）來選擇特徵。
   - Information-Based Methods：使用信息增益（Information Gain）、互信息（Mutual Information）等指標來評估特徵的信息量。
2. Wrapper Methods：通過反覆構建模型來評估特徵子集的性能，例如遞歸特徵消除（Recursive Feature Elimination）。
3. Embedded Methods：在模型訓練過程中進行特徵選擇，例如LASSO。
   - Model-based Feature Selection：使用機器學習模型來評估特徵的重要性，例如梯度提升樹或隨機森林。
   - Feature Importance Evaluation：計算模型中各特徵的重要性得分，通常在已經訓練好的機器學習模型中使用。

Reference:

1. https://medium.com/ai%E5%8F%8D%E6%96%97%E5%9F%8E/%E7%89%B9%E5%BE%B5%E5%B7%A5%E7%A8%8B%E4%B9%8B%E7%89%B9%E5%BE%B5%E9%81%B8%E6%93%87%E6%A6%82%E5%BF%B5-ca11745db63c
2. https://ithelp.ithome.com.tw/articles/10264846

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 to reference any external sources you consult. (Approximately 150 words, , excluding reference.)

TabNet 是一種基於注意力機制的神經網絡架構，專門為表格數據而設計。它特別設計的注意力機制，可以有效地從表格資料中提取關鍵特徵，並具有較強的解釋能力。具體來說，它透過加性模型以及注意力機制實現了 instance-wise 的特徵選擇，兼顧了樹模型以及NN的優點。與傳統的 ANN 相比，TabNet 在處理稀疏和高維度特徵時表現更好，同時能夠更好地處理缺失值。除此之外，TabNet 通常需要較少的數據進行訓練，因此對於數據量有限的情況下非常有用。

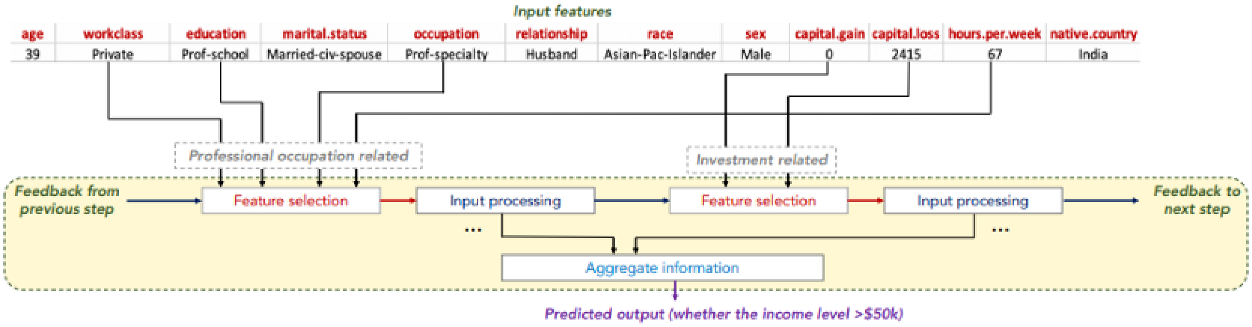


Figure 1: Depiction of TabNet's sparse feature selection for Adult Census Income prediction (Dua & Graff, 2017). TabNet employs multiple decision blocks that focus on processing a subset of input features for overall decision making. Two decision blocks shown as examples process the features that are professional occupation related and investment related in order to predict the income level.

Reference:

3. Arik, S. Ö., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 8, pp. 6679-6687).
4. https://blog.csdn.net/deephub/article/details/109044022