National Tsing Hua University
1130IEEM 513600
Deep Learning and Industrial Applications
Homework 2

Name: 魏詠新   Student ID: 113034542

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.

	lr	Batch	Training dataset Loss	Training dataset Accuracy	Validation dataset Loss	Validation dataset Accuracy	Test dataset Loss	Test dataset Accuracy
1	0.01	32	0.2567	0.8783	0.5574	0.8025	0.5603	0.7419
2	0.005	32	0.3162	0.8677	0.5268	0.8025	0.6148	0.7097
3	0.001	32	0.3847	0.8306	0.5414	0.7407	0.6512	0.6774
4	0.01	16	0.3309	0.8783	0.3297	0.8395	0.4817	0.7097
5	0.005	16	0.3324	0.8624	0.3501	0.8519	0.4660	0.7419
6	0.001	16	0.3245	0.8677	0.4366	0.7654	0.5913	0.7742
7	0.01	8	0.2875	0.8783	0.5706	0.8024	0.4926	0.7097
8	0.005	8	0.2546	0.9101	0.6549	0.8148	0.4705	0.7419
9	0.001	8	0.3025	0.8730	0.4806	0.7912	0.5915	0.7742

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.)

由圖可見，當batch size較大時，學習率越大，其accuracy表現越佳，但在batch size較小時，則沒有明顯的趨勢。適當調整 lr 和 Batch 可提高模型性能，但對於不同的batch size，適合不同的學習率，否則過高的 lr 可能導致收斂不穩定，而過小的 lr 可能導致訓練過慢或過擬合。總體來說， lr 為 **0.005** 且批次大小為 **16** 時(第5)，驗證 Loss 較低，測試準確率穩定，顯示較好的泛化能力。
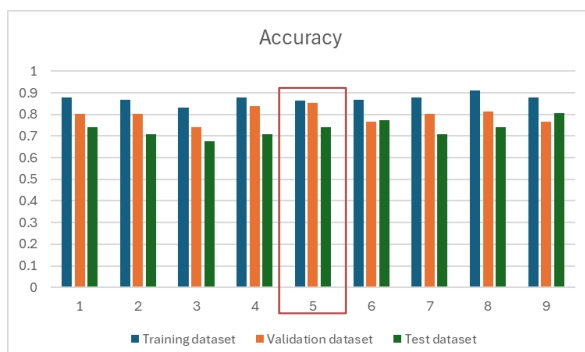
3. (20 pts) In Lab 2, you may have noticed a discrepancy in accuracy between the training and test datasets. What do you think causes this occurrence? Discuss potential reasons for the gap in accuracy. (Approximately 100 words.)

訓練數據集與測試數據集之間的準確率差異可能是由過擬合、訓練數據量不足或超參數選擇不當等因素導致的。模型在訓練數據上表現良好，但可能無法有效泛化到未見過的測試數據。當模型過於複雜或訓練次數過多時，可能會記住訓練數據的細節，而非學習可泛化的特徵。

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.)

以下是幾種特徵選擇方法：
- 單變量選擇：透過統計檢驗來評估特徵與目標變數的關聯性，篩選出重要特徵。
- 遞迴特徵消除（RFE）：逐步移除對模型影響較小的特徵，保留影響力較大的特徵子集。
- 隨機森林特徵重要性：利用隨機森林模型計算各特徵對預測結果的貢獻，選擇關鍵特徵。
- 主成分分析（PCA）：將原始數據轉換到新的坐標系，透過降維技術保留最具資訊量的主要成分。

特徵選擇的目標是在大量變數中找出對模型預測最有價值的部分，從而提高準確度與運算效率，它也能降低模型複雜度，減少過擬合風險，並提升解釋能力。

參考資料: https://www.ai-prediction.info/single-post/%E6%95%B8%E6%93%9A%E5%B0%8F%E7%9F%A5%E8%AD%98-feature-selection?utm

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更適合用於表格數據。ANN 通常在處理表格數據時，必須進行大量的數據預處理，如特徵縮放、類別編碼、缺失值處理等。而TabNet不需要這些繁瑣的預處理，可以直接輸入原始表格數據，並利用基於梯度下降的端到端學習完成模型訓練。並且傳統ANN 並沒有內建的特徵選擇機制，但TabNet在每一個決策步驟中，會利用順序注意力機制自動選擇「最重要的特徵」，讓模型專注於與預測任務最相關的特徵，減少不必要的噪聲，提升學習效果。這個過程類似於決策樹逐步分裂的重要特徵選擇，但透過神經網絡的方式實現，兼具了決策樹的選擇性與深度學習的表徵能力。

參考資料: https://zhuanlan.zhihu.com/p/506945010?utm