National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name: 姚國慶　　Student ID: 112003802

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.

Ans.

選擇 batch size 超參數(learning rate 固定為 1.00E-03)，調整 4, 32, 128 三種大小，模型正確率(train/val/test)結果如下表。

batch size	learning rate	training accuracy	validation accuracy	test accuracy
4	1.00E-03	87.30%	83.95%	**80.65%**
32	1.00E-03	85.71%	74.90%	69.89%
128	1.00E-03	78.48%	70.37%	65.59%

【註】表中實驗結果皆為 3 次 run 的平均值

在 learning rate 固定為 1.00E-03 的條件下，Batch size=4 得到最高的正確率 80.65%

選擇 learning rate 超參數(batch size 固定為 4)，調整 1.00E-02, 1.00E-03, 1.00E-04 三種大小，模型正確率(train/Val/test)結果如下表。

batch size	learning rate	training accuracy	validation accuracy	test accuracy
4	1.00E-02	85.54%	84.36%	76.34%
4	1.00E-03	87.30%	83.95%	**80.65%**
4	1.00E-04	83.77%	72.84%	68.82%

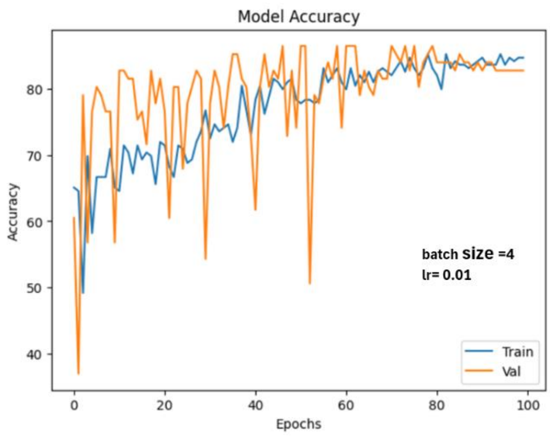
【註】表中實驗結果皆為 3 次 run 的平均值

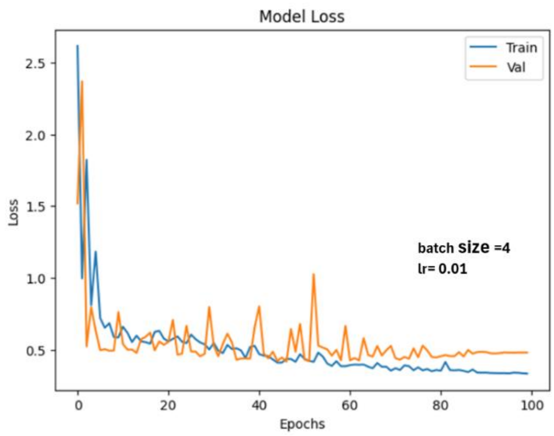
在 Batch size 固定為 4 的條件下，learning rate=1.00E-03 得到最高的正確率 80.65%

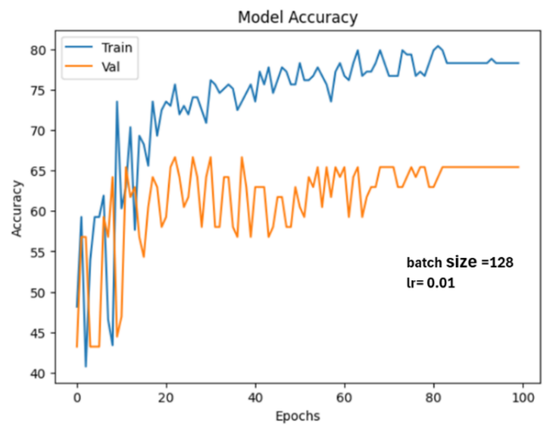
2. (20 pts) Based on your experiments in Question 1, analyze the outcomes. What differences do you observe with the changes in hyper-parameters? Discuss whether these adjustments contributed to improvements in model performance, you can use plots to support your points. (Approximately 100 words.)

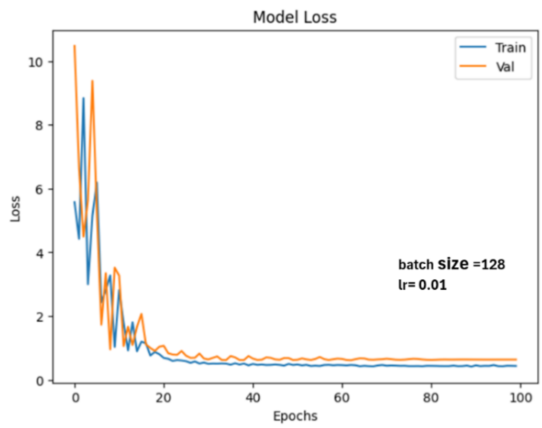
Ans.

batch size 是 global 和 local 的 trade-off，例如 larger batch size 較能找到 global trend，但相對會跳過最佳化點。此作業的模型 batch size=128 在 train 與 val 時 acc 有較大差異就是學習時 loss 掉一些資料的特性導致 test 的 acc 不好。此時設定 smaller batch size 相對可以逐步找到資料的真實特性，提升 test 的 acc。不過 Smaller batch 在訓練時會有 acc 震盪較大的情形，必須觀察震盪情形是否收斂，否則 model 會不 robust。









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.)

Ans.

以實驗參數 batch size = 128 為例，train acc 是 78.48%, test acc 是 65.59%，差異約 13%，這就是在 128 的情形下學習跳太大 loss 掉資料的一些重要訊息， 此時將 batch size 調小有助於學習細部的資料訊息得到 train 和 test 較一致的結果，如本作業 batch size= 4 時，train acc 是 87.30%, test acc 是 80.25%，差異縮小一半約 7%。

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.)

   Ans.

   features selection 是 machine learning 流程中的關鍵步驟，影響從數據集中識別和選擇最相關的特徵以建構最適的預測模型。features selection 在改善模型性能、減少過度擬合和提高可解釋性上有非常重要的作用。以下舉例一些特徵工程的方法：

   (1)基於機器學習模型的特徵選擇技術，例如決策樹：基於決策樹的模型，如決策樹(decision tree)中的隨機森林(random forest)，通過選擇最具訊息性的特徵來劃分節點。

   (2)基於領域和專業((domain & expertise)知識，可以藉由專家定義 rule-based rule 來協助模型選擇特徵。當處理特定於領域的數據集時，某些特徵可能與目標變量具有已知的關聯，這種方法尤其有用。

   Resource: Guyon, I., & Elisseeff, A. (2003). "An Introduction to Variable and Feature Selection". Journal of Machine Learning Research, 3, 1157-1182.

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.)

   Ans.

   一個更適合處理表格數據的深度學習模型是 TabNet 模型。TabNet，全名 Tabular Network，專為表格數據而設計，結合了深度學習和傳統決策樹的特性。它採用了一種 sequential attention mechanism，逐步選擇信息豐富的特徵並進行預測。與傳統的 ANN 不同，TabNet 不需要特徵工程或預處理，因此非常適合具有異質特徵類型和缺失值的表格數據。其可解釋的 sequential attention mechanism 使得模型透明化，並能進行特徵重要性分析。

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