National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name:林郁希 Student ID:312350045

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.

Run 1	Run 2	Run 3
Activation functions: ['ReLU']	Activation functions: ['ReLU']	Activation functions: ['ReLU']
Batch_size: 32	Batch_size: 64	Batch_size: 128
Train acc: 83.5978835978836	Train acc: 79.8941798941799	Train acc: 75.13227513227513
Train loss: 0.3891709993282954	Train loss: 0.4312191108862559	Train loss: 0.49659501016139984
Val acc: 83.95061728395062	Val acc: 83.95061728395062	Val acc: 67.90123456790124
Val loss: 0.4342486759026845	Val loss: 0.4920649528503418	Val loss: 0.5863717794418335
Test acc: 83.87096774193549	Test acc: 67.74193548387096	Test acc: 58.064516129032256
Test loss: 0.5403211067520803	Test loss: 0.6636876533589056	Test loss: 0.9805642732328945
Run 4	**Run 5**	**Run 6**
Activation functions: ['Tanh']	Activation functions: ['Tanh']	Activation functions: ['Tanh']
Batch_size: 32	Batch_size: 64	Batch_size: 128
Train acc: 92.5925925925926	Train acc: 82.01058201058201	Train acc: 75.66137566137566
Train loss: 0.23092448835571608	Train loss: 0.37702303131421405	Train loss: 0.46024127304553986
Val acc: 75.30864197530865	Val acc: 65.4320987654321	Val acc: 75.30864197530865
Val loss: 0.5528976917266846	Val loss: 0.6419929265975952	Val loss: 0.5072206854820251
Test acc: 77.41935483870968	Test acc: 67.74193548387096	Test acc: 58.064516129032256
Test loss: 0.5495924027336221	Test loss: 0.7159139501952356	Test loss: 0.6758817655424918
Run 7	**Run 8**	**Run 9**
Activation functions: ['Sigmoid']	Activation functions: ['Sigmoid']	Activation functions: ['Sigmoid']
Batch_size: 32	Batch_size: 64	Batch_size: 128
Train acc: 86.24338624338624	Train acc: 76.71957671957672	Train acc: 76.71957671957672
Train loss: 0.34154412895441055	Train loss: 0.45830220977465314	Train loss: 0.4696127623319626
Val acc: 77.77777777777777	Val acc: 70.37037037037037	Val acc: 69.1358024691358
Val loss: 0.5348519186178843	Val loss: 0.6827734410762787	Val loss: 0.6509851217269897
Test acc: 67.74193548387096	Test acc: 67.74193548387096	Test acc: 67.74193548387096
Test loss: 0.5672047609883931	Test loss: 0.6782981376013448	Test loss: 0.6483308634450359

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.)

Number of samples in train and validation are 189 and 81.

較小的 Batch size 會對於訓練模型有更大的成效，考慮到 Iteration 和 Distribution，訓練資料共有 189 筆，32 的 batch size 可以 iteration 六次，多於其他兩者，並且觀察結果，數量有覆蓋到資料原始的 distribution。而比較 ReLU、Tanh、Sigmoid，Sigmoid 的形狀固定難以透過迭代顯著改變，Tanh 和 ReLU 都具有較高的學習能力。綜觀下來，這筆資料用小的 Batch size 和較彈性的 ReLU 會有較好的表現。

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.)

Number of samples in train and validation and test are 189 and 81 and 32.

以呈現結果來看，Validation和Testing表現遠不如Training，正常會判斷為overfitting或mismatch的一種，但我認為純粹是資料數太少，Batch size不論32,64,128，對上Val: 81, Test: 32能進行的Iteration驗證都太少，不足以反應Training的訓練成果，不知道是否算資料不平衡的一種。總上所述，資料太少而導致呈現出有巨大GAP的成果。

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.)

feature selection對於訓練預期方向是有幫助的，能減少相當大程度的雜訊，並且能以較少的資料數達到更高的成效，另外能減少模型的規模。方法很多元，排序features重要程度、用algorithm(ex: Lasso)，或是TabNet都行。重點是篩選出與訓練模型預期符合和相關性高的features達成更好的效率或是效果。(Borisov et al., 2024).

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 為一個適合 Tabular data 的 alternative deep learning model，原因為 TabNet 具有幾個為 Tabular data 設計的特性，1. sequential attention mechanism, 2.sparse feature masking,3. self-supervised pretraining，1 的作用是能針對每筆資料進行動態的特徵選擇，並保有一定的彈性，2 的功能為能精確地找出少量關鍵特徵用於訓練，3 的功能為在標註資料有限的情況下，也能學習到有效的特徵。最重要的是 TabNet 具有可解釋性，能幫助我們理解結果呈現的邏輯。(Borisov et al., 2024).