National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name: 楊雅淳 **Student ID: 113036535** **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.

In Lab 2, two hyper-parameters were selected for experimentation: **learning rate** and **number of hidden layers**. Three different values were chosen for each:

- **Learning Rate**:
  - 0.001 / 0.01 / 0.1
- **Number of Hidden Layers**:
  - 1 hidden layer / 2 hidden layers / 3 hidden layers

The model was trained with these different combinations, and performance was measured across training, validation, and test datasets. The results are summarized in the table below:

Learning Rate	Hidden Layers	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Test Accuracy
0.001	1	0.30	88.88	0.55	74.07	67.74
0.001	2	0.35	82.01	0.76	77.77	70.96
0.001	3	0.40	78.83	0.63	74.07	70.96
0.01	1	0.30	88.35	0.48	82.71	77.41
0.01	2	0.26	86.77	0.45	83.95	83.87
0.01	3	0.26	86.24	0.74	77.77	70.96
0.1	1	0.69	52.38	0.68	59.25	48.38
0.1	2	0.68	53.96	0.68	55.55	48.38
0.1	3	0.69	52.38	0.68	59.25	48.38

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.)

The results show a clear relationship between the learning rate, number of hidden layers, and the model's performance. As the learning rate increased from 0.1 to 0.01, the training and validation accuracy improved, suggesting that the model was able to learn faster and achieve better generalization. Additionally, increasing the number of hidden layers consistently improved both training and test accuracy, but only to a point. The test loss and test accuracy showed improvements with up to two hidden layers, but beyond that, the returns became diminishing.

This supports the idea that the model benefits from deeper networks and higher learning rates, but there is an optimal number of hidden layers. Overfitting is also a potential issue when the model becomes too complex.

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.)

The observed discrepancy between training and test accuracy can be attributed to **overfitting**. This occurs when the model learns to perform exceedingly well on the training data but fails to generalize to unseen data.

Overfitting can happen when the model is too complex relative to the amount of data, or the model has too many parameters that lead it to memorize rather than generalize. Additionally, **insufficient regularization** or improper **data augmentation** during training can exacerbate this gap in accuracy. The test set, being unseen data, often highlights the generalization capabilities of the model.

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 is the process of identifying and choosing the most relevant input variables for machine learning models. It is crucial because irrelevant or redundant features can decrease model accuracy, increase training time, and lead to overfitting.

Methods for feature selection include **filter methods**, which rank features based on statistical tests; **wrapper methods**, which evaluate subsets of features by training the model on them; and **embedded methods**, which perform feature selection during the model training process itself.

For instance, decision trees provide feature importance scores that help in selecting the most important features.

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.)

雖然人工神經網絡（ANN）非常受歡迎，但它們並不總是處理表格數據的最佳選擇，因為它們容易 overfitting 並切需要大量的數據。另一個更適合的替代模型是 TabNet，這是一種專門為表格數據設計的深度學習模型。TabNet 使用注意力機制在每一步決策過程中選擇相關特徵，這使得它在結構化數據中更加可解釋且有效。TabNet 模型專注於學習哪些特徵最為重要，以及如何使用這些特徵進行決策，從而比傳統的 ANN 提供更高的準確度和效率。TabNet 的一大優勢是它能更好地處理稀疏數據和小型數據集，因為它不像深度神經網絡那樣依賴大量的預處理或訓練集。

Reference Data :“TabNet: Attentive Interpretable Tabular Learning” by Arik 和 Pfister（2020）。