National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name: 黃皓澤 Student ID: 113034519

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.

Epochs	Train Accuracy	Train Loss	Validation Accuracy	Validation Loss	Test Accuracy
50	77.77%	0.4449	82.72%	0.4712	67.74%
100	83.06%	0.3828	80.25%	0.4030	77.419%
150	85.71%	0.3362	85.19%	0.4299	83.87%

Epochs	Train Accuracy	Train Loss	Validation Accuracy	Validation Loss	Test Accuracy
1e-4	75.132%	0.4866	80.25%	0.5084	64.516%
1e-3	83.06%	0.3828	80.25%	0.4030	77.419%
1e-2	86.77%	0.3254	86.42%	0.3483	77.419%

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.)

Increasing epochs generally improves performance, but validation accuracy fluctuates. 100 epochs showed overfitting, but 150 epochs balanced training and generalization well.

1e-4 is too small, leading to slow convergence and poor generalization.

1e-3 is a reasonable choice, balancing accuracy and loss.

1e-2 may be overfitting, given its high validation accuracy but stagnant test accuracy.

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 accuracy gap between training and test data suggests overfitting, where the model learns noise instead of general patterns. Higher training accuracy (e.g., 86.77%) with lower test accuracy (77.419%) indicates poor generalization. Possible causes include data leakage, high model complexity, insufficient data, or distribution shift between train and test sets. Solutions include early stopping, dropout, regularization, reducing model complexity, and ensuring a proper train-test split. Increasing data or using augmentation can also help. Addressing these issues will improve test performance and prevent the model from memorizing training data.

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 a critical step in preparing tabular datasets for machine learning models, as it helps in reducing dimensionality, removing noise, and improving model performance.

Feature selection methods can be broadly categorized into three main types: Filter Methods, Wrapper Methods, and Embedded Methods.

Filter Methods:

Filter methods evaluate the relevance of features based on their statistical properties, independently of any machine learning algorithms. Common techniques include the Chi-Square Test, correlation coefficients, and mutual information. These methods are computationally efficient and often used for preliminary feature selection.

Wrapper Methods:

Wrapper methods assess feature subsets based on the performance of a predictive model. Techniques such as Recursive Feature Elimination (RFE) and forward or backward feature selection fall under this category. Although they tend to provide better performance, they are more computationally intensive compared to filter methods.

Embedded Methods:

Embedded methods perform feature selection during the model training process, balancing computational efficiency and performance. Lasso regression, which utilizes L1 regularization to eliminate less important features, is a prominent example of this approach.

https://www.prepbytes.com/blog/data-mining/feature-selection-techniques-in-machine-learning/

https://codefinity.com/blog/Feature-Selection-Techniques-in-Machine-Learning

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.)

Alternative Deep Learning Model for Tabular Data: Sparse Hierarchical Table Ensemble (S-HTE)

Model Overview: The Sparse Hierarchical Table Ensemble (S-HTE) is a deep learning model specifically designed for tabular data, utilizing ferns (oblivious decision trees) instead of traditional neurons. This model is particularly advantageous in environments with limited computational resources, such as IoT, drones, or NUI applications.

Design Rationale: S-HTE is structured to be dense at the beginning of the training process and gradually becomes sparse through an annealing mechanism. This design allows the model to efficiently learn useful internal representations and benefit from increased depth, which is crucial for handling the complexities of tabular data.

Key Features:

- Efficient Learning: The model starts dense and becomes sparse, optimizing computational resources while maintaining accuracy.
- Internal Representation: S-HTE learns meaningful internal representations, which is a significant advantage over traditional ferns that do not typically benefit from increased depth.
- Lower Computational Complexity: Compared to neuron-based networks, S-HTE offers comparable accuracy with reduced computational demands, making it suitable for applications with limited processing power.

Advantages:

- Resource Efficiency: S-HTE is designed to operate efficiently in low-computation environments, making it ideal for applications where computational capacity is a constraint.
- Competitive Accuracy: The model achieves accuracy comparable to other deep learning alternatives while maintaining lower computational complexity, which is beneficial for practical applications.

Farjon, G., & Bar-Hillel, A. (2022). Sparse Hierarchical Table Ensemble–A Deep Learning Alternative for Tabular Data. IEEE Access, 10, 75376-75384. https://doi.org/10.1109/access.2022.3190537.