National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 2

Name: 黃家興　　　Student ID: 111003806

Due on 2024.03.21

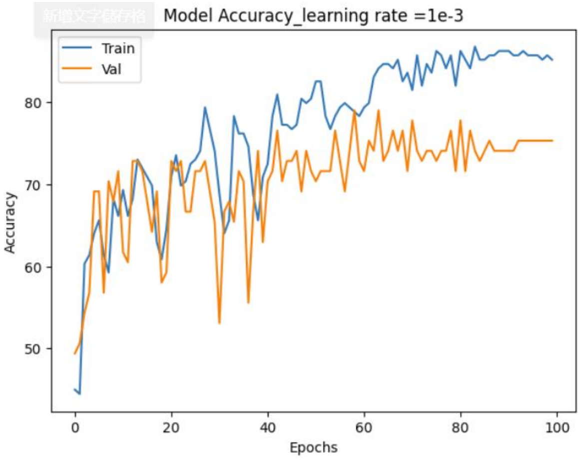
1. (20 pts) Select 2 hyper-parameters of the artificial neural network used in Lab 2, and set 3 different values for each. Perform experiments to compare the effects of varying these hyper-parameters on the loss and accuracy metrics across the training, validation, and test datasets. Present your findings with appropriate tables.

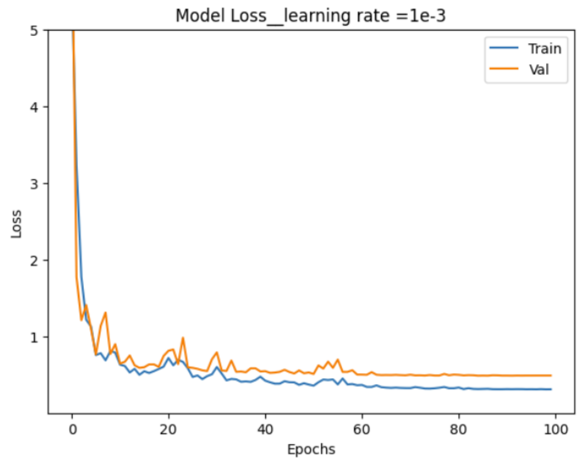
Try to set 3 different values for batch size (Exp. 1) and learning rate (Exp. 2).

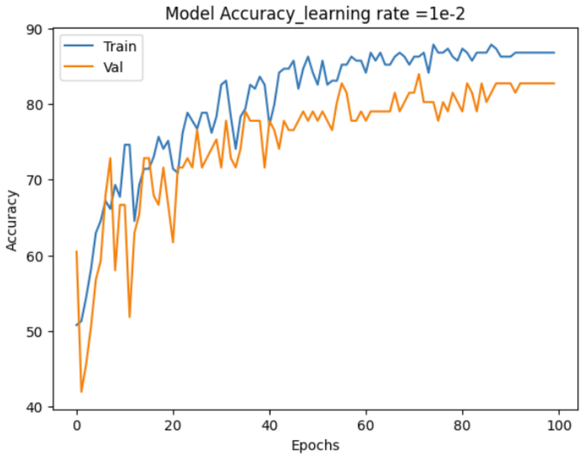
	Batch size	Learning rate	Training		Validation		Test
			Loss	Accuracy	Loss	Accuracy	Accuracy
Exp. 1	32	1.00E-03	0.3325	87.83%	0.5501	71.60%	74.19%
	320	1.00E-03	0.5158	72.49%	0.6685	64.20%	64.52%
	8	1.00E-03	0.2916	87.30%	0.454	76.54%	83.87%
Exp. 2	32	1.00E-03	0.3311	88.89%	0.5491	70.37%	77.78%
	32	1.00E-02	0.3368	85.19%	0.4568	82.72%	70.97%
	32	1.00E-04	0.502	77.25%	0.5993	65.43%	61.29%

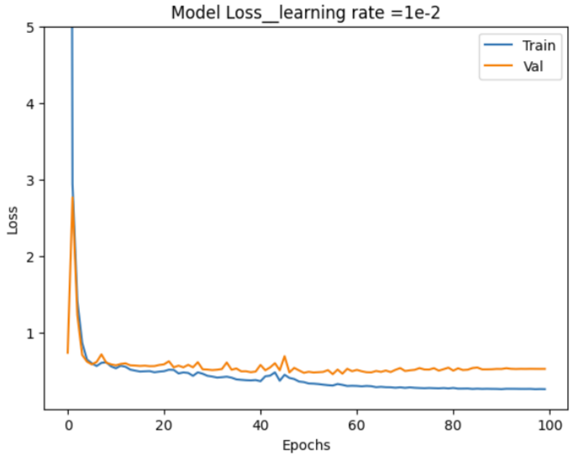
2. (20 pts) Based on your experiments in Question 1, analyze the outcomes. What differences do you observe with the changes in hyper-parameters? Discuss whether these adjustments contributed to improvements in model performance, you can use plots to support your points. (Approximately 100 words.)

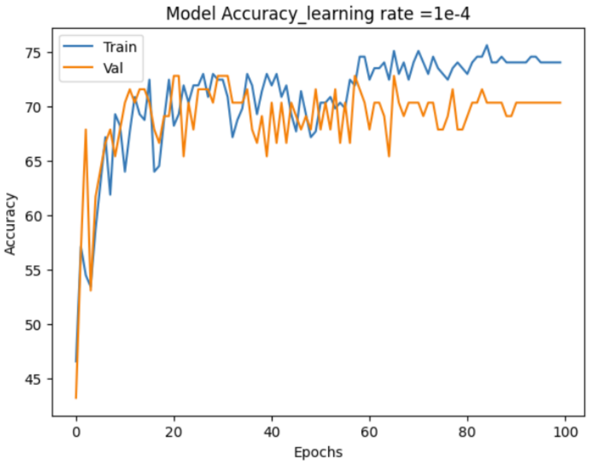
In Exp. 1 show the smaller batch sizes can provide better accuracy and the batch size set 8 is the best in this case. On the other hand, large batch sizes get lower accuracy and overfitting. When we compare the difference between from the model loss in Exp.2, a higher learning rate can offer faster convergence initially but may suffer from instability. A lower learning rate provides more stable optimization but albeit at the cost of slower convergence.

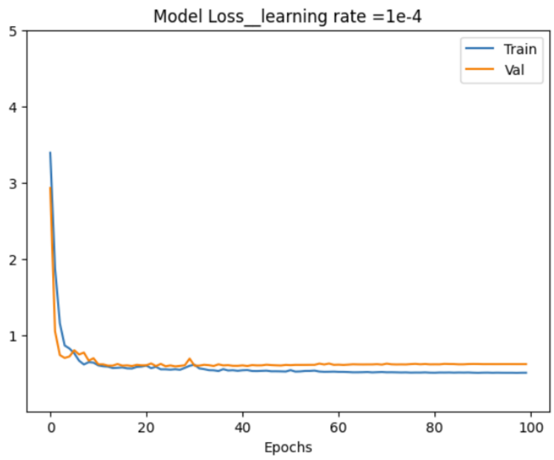












3. (20 pts) In Lab 2, you may have noticed a discrepancy in accuracy between the training and test datasets. What do you think causes this occurrence? Discuss potential reasons for the gap in accuracy. (Approximately 100 words.)

The gap in accuracy between the training and test datasets, known as the training-test discrepancy, that the potential reasons come from over fitting or under fitting. The Model optimization is not good enough. Hyperparameter tuning can lead to models that are optimized for the training data but fail to generalize well. Insufficient training data may not capture the true underlying patterns, leading to poor generalization to unseen 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 task is one kind of data preprocessing task where the redundant and irrelevant features have been omitted to improve the classifier's accuracy and time complexity. Methods for selecting relevant features include filter methods (e.g., correlation analysis), wrapper methods (e.g., forward/backward selection), and embedded methods (e.g., regularization). The tabular comparison within these three models shown that filter approach is robust against overfitting and there is a chance to fail, the selection of useful features. Wrapper methods is very prone to overfitting, it selects the most useful features but has massive time complexity. Embedded approach is less prone to overfitting, comparatively lower time complexity. Reference: [1]

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.)

   The ensemble of deep models and XGBoost is better than ANNs suited for tabular data. Deep learning models excel at capturing complex patterns in data, while XGBoost (an implementation of gradient boosting decision trees) is highly effective at handling structured/tabular data and mitigating overfitting. By combining these two approaches, the ensemble leverages the strengths of both methods to achieve superior performance. XGBoost provides interpretability through feature importance rankings and decision paths. This combination ensures that the ensemble maintains a balance between model complexity and interpretability, crucial for understanding the underlying patterns in tabular data. Ensemble methods inherently offer robustness against overfitting and noise in the data. By aggregating predictions from multiple models, the ensemble reduces the risk of individual model biases and variance, leading to more stable and reliable results. The ensemble's ability to generalize well on unseen data is enhanced by combining diverse models trained on different aspects of the data. This ensures that the ensemble can capture a broader range of patterns and relationships present in the tabular dataset. Reference: [2] [3]

[1] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Applied Intelligence*, vol. 52, no. 4, pp. 4543-4581, 2022.

[2] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci, "Deep neural networks and tabular data: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[3] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84-90, 2022.