National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 2

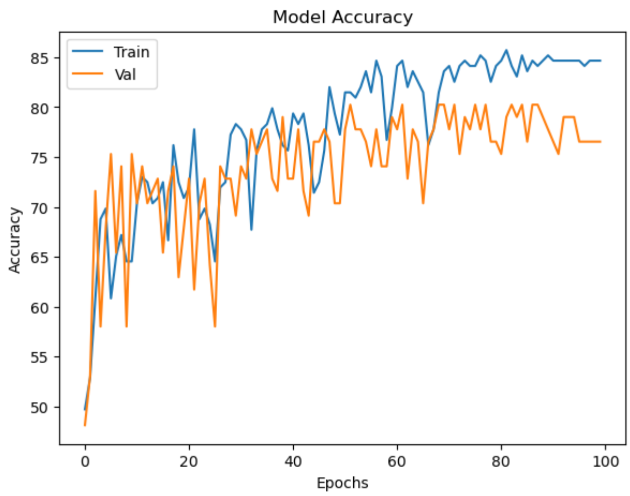
Name: 龔友謙　　　　Student ID: 112030512

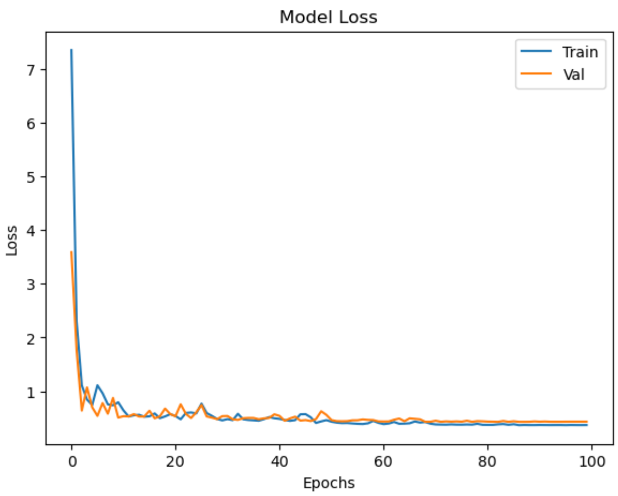
Due on 2024.03.21

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.

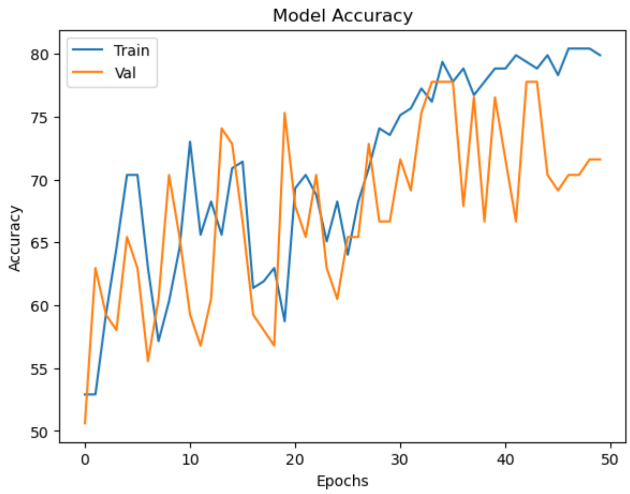
   First : Epochs (current=100), Second : learning rate (current = 0.001)

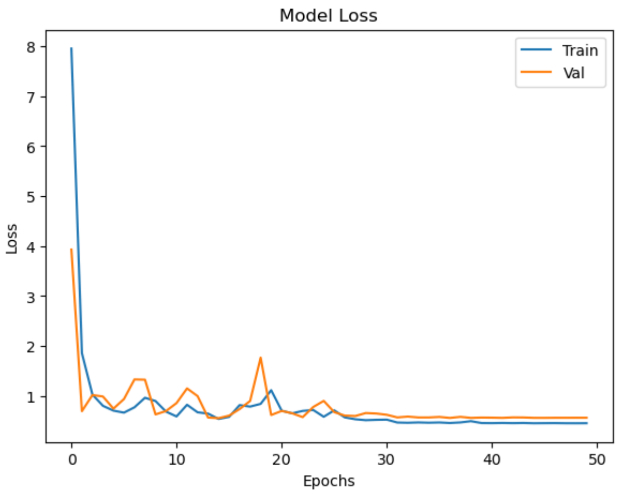
   Current: Test accuracy is 64.51612903225806%



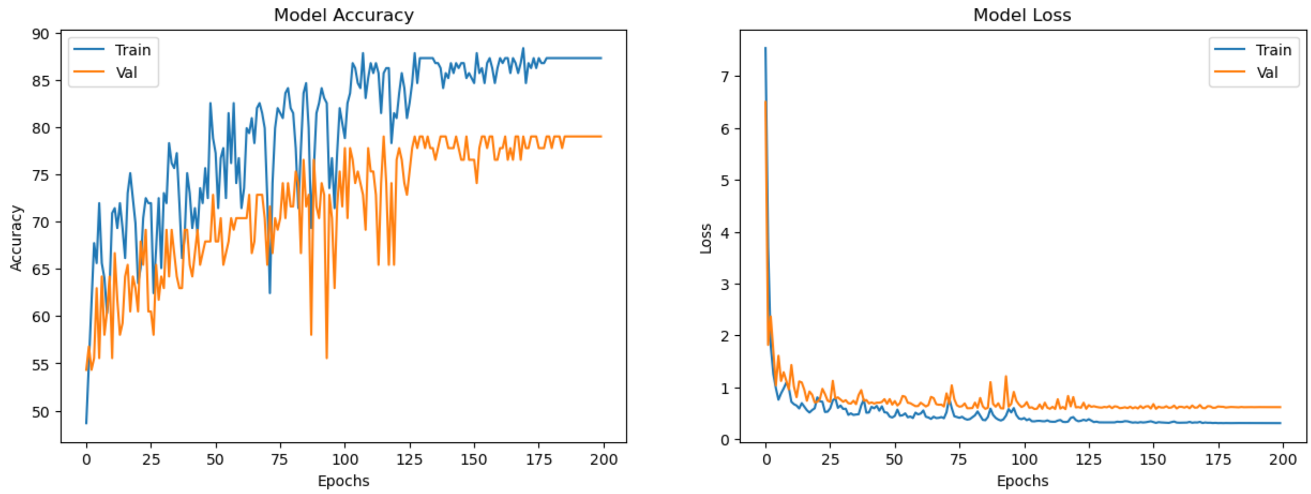


   Epoch = 50: Test accuracy is 58.064516129032256%

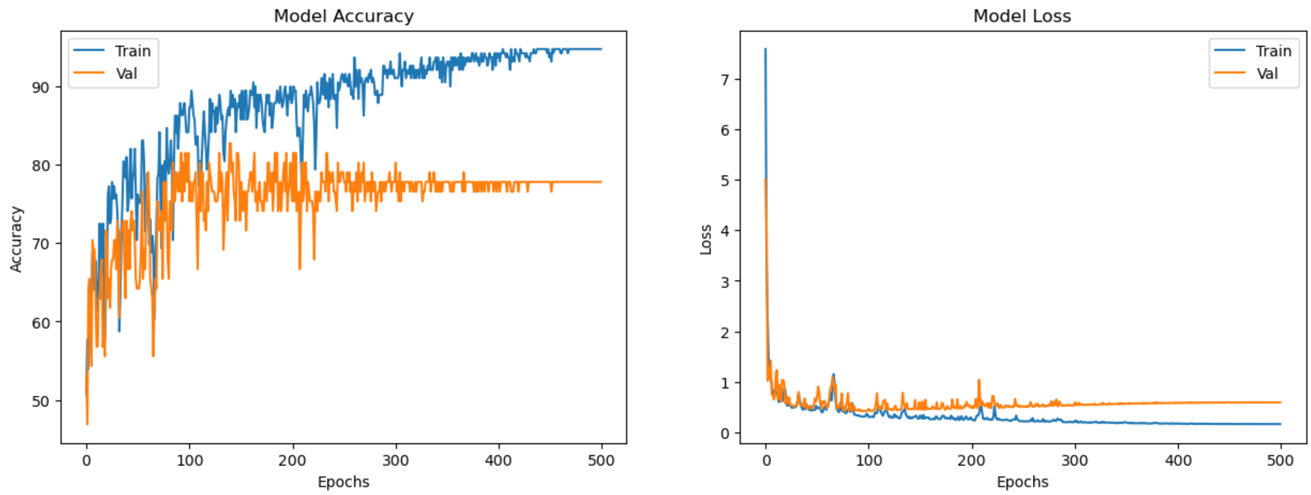




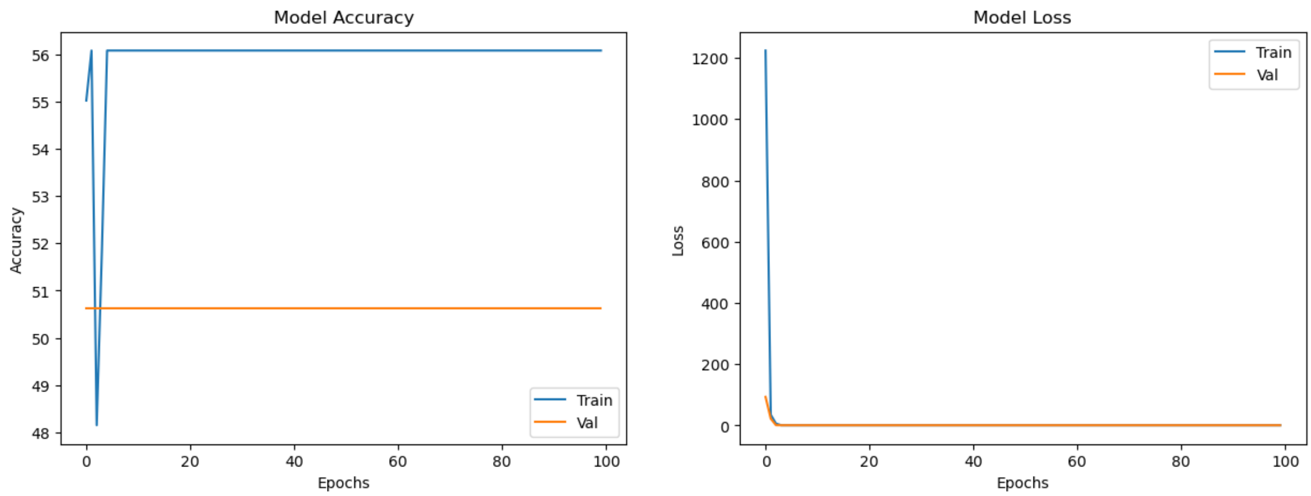
   Epoch = 200: Test accuracy is 70.96774193548387%



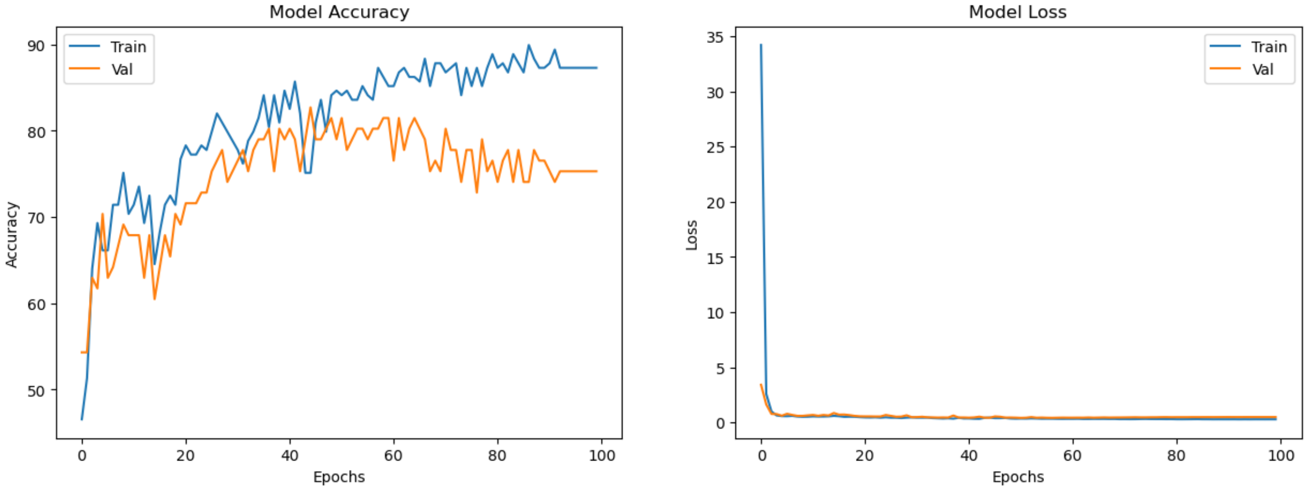
Epoch = 500: Test accuracy is 80.64516129032258%



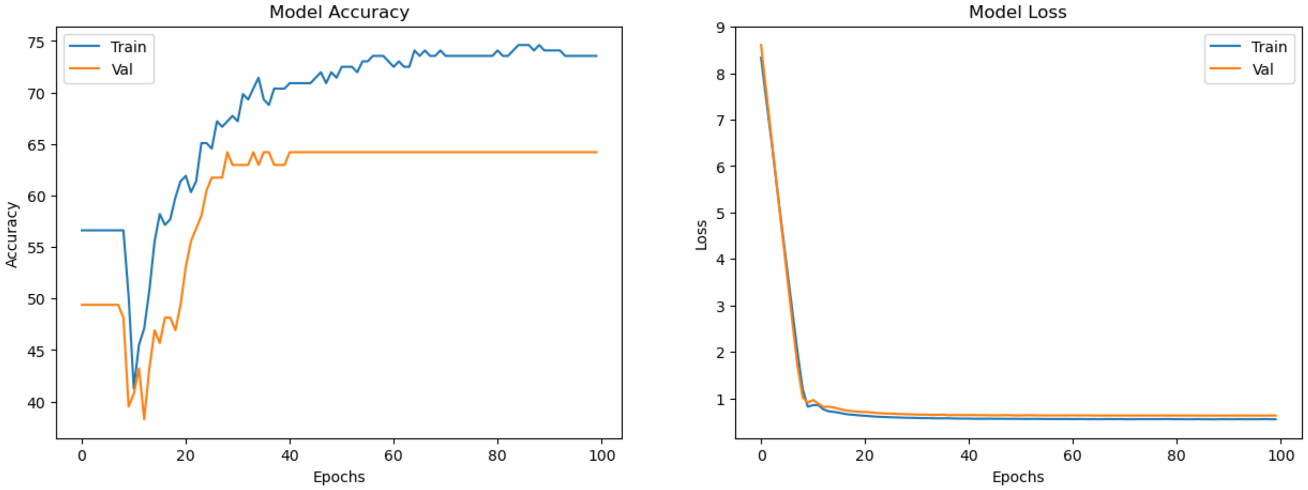
learning rate = 0.1: Test accuracy is 48.38709677419355%



learning rate = 0.01: Test accuracy is 77.41935483870968%



learning rate = 0.00001: Test accuracy is 64.51612903225806%



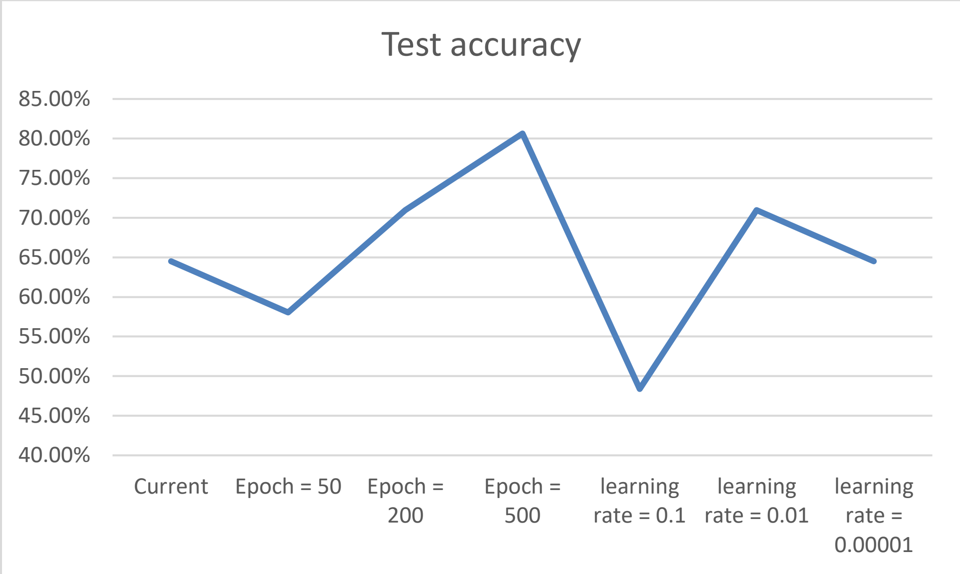
	Current	Epoch = 50	Epoch = 200	Epoch = 500	learning rate = 0.1	learning rate = 0.01	learning rate = 0.00001
Test accuracy	64.51 %	58.06 %	70.96 %	80.64 %	48.38 %	70.96 %	64.51 %

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.)

My findings:

Higher epochs result in higher test accuracy, since epochs are how many iterations the model will go through during training, it is reasonable to have higher scores when increasing the number of epochs.

Learning rate dose not particularly better for low or high values, but it can not be to high since we got the worst performance on learning rate = 0.1. We can try many different learning rates to find the optimal one.



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.)

There are many reasons that can cause discrepancy in accuracy between the training and test datasets. From the result of Lab 2 and based on my knowledge, the most common reason is overfitting, where the model learns to memorize training data rather than generalize, it happens more often when the dataset is small.

Some other reasons like data mismatch between training and test sets, data leakage or improper validation techniques can also cause this result.

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.)

Many machine learning may use different feature selection like feature importance ranking, L1 Regularization (Lasso), univariate feature selection. Popular algorithms like decision trees, random forests, or extreme gradient boosting provide a feature importance score. Features with higher importance scores are considered more relevant and are selected for the final model.

It can improve model accuracy by selecting only the most relevant features reducing noise and irrelevant information, excluding irrelevant features that can cause overfitting. It can also enhanced interpretability since models with fewer features are easier to interpret and understand and having a faster training time

Reference:

1. Guyon, I., & Elisseeff, A. (2003). "An Introduction to Variable and Feature Selection." Journal of Machine Learning Research, 3, 1157-1182.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." Springer Science & Business Media.
3. Raschka, S., & Mirjalili, V. (2019). "Python Machine Learning, 3rd Edition." Packt Publishing.

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.)

An alternative deep learning model better suited for tabular datasets is the Tabular model (e.g., Google's TabNet or Microsoft's Node). These models are designed specifically for tabular data, leveraging techniques like self-attention and masked multi-task learning to capture relationships between columns.

TabNet adopts a fully deterministic transformer-style architecture, using self-attention layers to effectively capture column interactions and assigning different importance to features during the decision process. Node introduces masked column-wise multi-task learning to reinforce the model's understanding of relationships between tabular columns.

Compared to traditional ANNs, these tabular models have significant advantages when dealing with high-dimensional sparse tabular data, better modeling column interactions to improve predictive accuracy. They also avoid manual feature engineering, simplifying the data processing pipeline.

Key advantages include their specialized architecture for tabular data, efficient modeling of column relationships, and automated feature learning capabilities.

Reference:

1. Arik, S. Ö., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning.
2. ChatGPT