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.

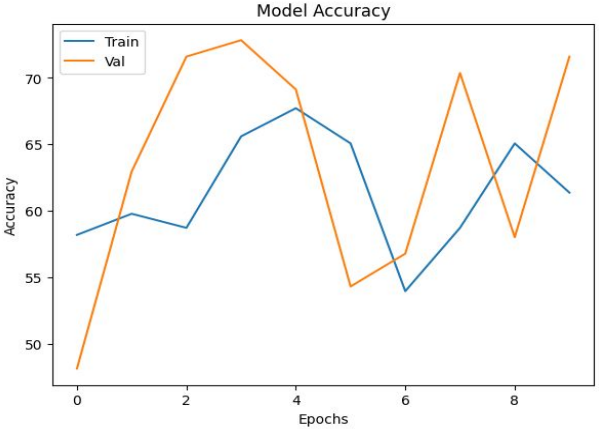
- 2 hyper-parameters:
  - batch size: 16, 32, 64
  - learning rate: 1e-1, 1e-2, 1e-3

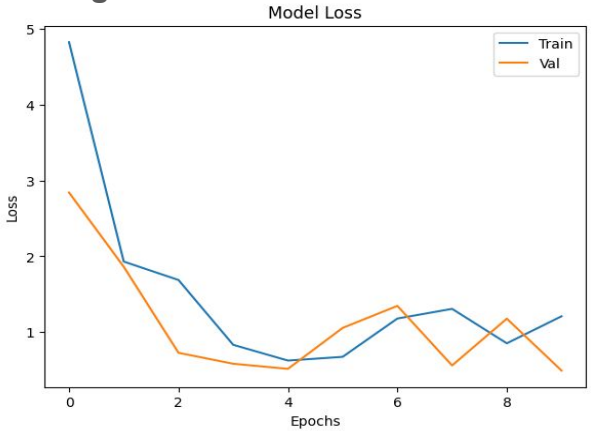
batch size	learning rate	Train loss	Train acc	Val loss	Val acc	Best Val loss	Best Val acc	Test accuracy
16	1.00E-01	0.6935	52.91%	0.6706	58.02%	0.6703	58.02%	48.39%
32	1.00E-02	0.5928	68.78%	0.5467	71.60%	0.5334	75.31%	61.29%
64	1.00E-03	1.0525	66.14%	1.7349	59.26%	1.1351	60.49%	58.06%
16	1.00E-02	0.6275	64.55%	0.502	79.01%	0.458	79.01%	64.52%
32	1.00E-03	0.5868	68.25%	0.6673	64.20%	0.5745	67.90%	64.52%
64	1.00E-01	0.7229	55.56%	0.7058	51.85%	0.6858	51.85%	48.39%
16	1.00E-03	0.7033	66.67%	0.483	74.07%	0.483	76.54%	64.52%
32	1.00E-01	0.6893	55.03%	0.6913	53.09%	0.6913	53.09%	48.39%
64	1.00E-02	0.5978	67.20%	0.5956	64.20%	0.5392	71.60%	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.)

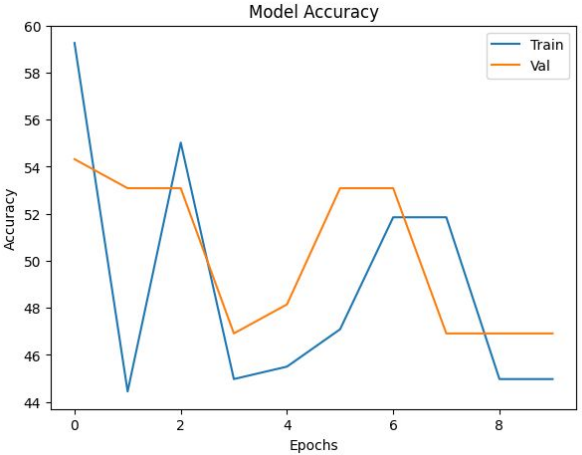
- When the batch size is 16, it is easier to obtain higher test accuracy.
- When the learning rate is 1e-3, it is easier to obtain higher test accuracy.
- According to the experimental results, it is known that the **smaller** the batch size and learning rate, it is easier to have higher test accuracy.
- A larger batch size cause the model to get stuck in a local optimal solution, which may be easier to escape with a smaller batch size.
- A larger learning rate may cause the model to be more susceptible to underfitting because the model may not be able to adequately learn the features of the training data.

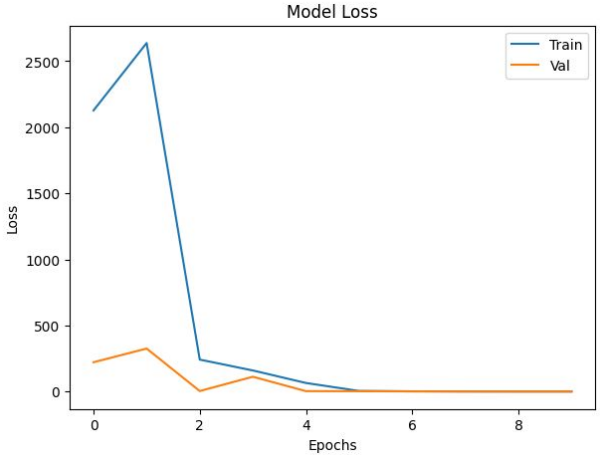
batch size = 16, learning rate = 1e-3





batch size = 64, learning rate = 1e-1





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 difference between training and test datasets is large, causing the model to encounter **too many unseen situations**, so accuracy is not very ideal.
- The learning ability of the model is so good that it is in an **overfitting** state. It can only accurately answer the data it has seen in the training datasets, but it cannot answer the test datasets it has not seen through learning.
- If the **hyperparameters of the model (such as learning rate, batch size, etc.) are not chosen properly**, it may also lead to accuracy differences between the training set and the test set.

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.)

- Algorithm for selecting features

  1. The Full Search method has an operation complexity of **2 to the Nth power**.(When there are many features, the amount of calculation will be too large.)

  2. The Greedy Forward Selection method has an operational complexity of **N square**.(Although the amount of calculations is reduced a lot compared to the previous method, the amount of calculations is still relatively large.)

  3. The Stepwise Forward Selection method has an operational complexity of **2N**.(Filtering through set criteria can effectively reduce the amount of calculations.)

  4. The Simplified Greedy Forward Selection method has an operational complexity of **N**.(The least amount of calculation among the above four methods.)

參考資料:
https://flag-editors.medium.com/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92lesson-4-%E7%89%B9%E5%BE%B5%E9%81%B8%E6%93%87-bf6caf08ac36

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.)

- Choosing appropriate features to train the model can **greatly increase the accuracy of the model**, because those features are the **important factors** that affect the results.
- Including irrelevant or redundant features **may lead to overfitting**. Feature selection helps alleviate overfitting by reducing model complexity.
- Models with fewer features are generally easier to interpret and understand. Because other irrelevant noise is reduced, the **interpretability of the model is improved**.

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.)

- **TabNet (Tabular Attention Network)** is a deep learning model designed for tabular data.
- TabNet uses **sequential attention** to select features to reason about at each decision step to achieve interpretability and more efficient learning. This method enables learning to focus on the most representative features, thereby improving learning efficiency.
- TabNet has advantages similar to decision trees (**high interpretability**), and at the same time has the **high performance** of deep learning, so it has great benefits in processing tabular data.
- TabNet utilizes **gradient-based** learning algorithms for model training, enabling efficient optimization of model parameters.
- TabNet combines **a sparse feature selection mechanism** to encourage the sparsity of feature activation, and can effectively process high-dimensional data while maintaining computational efficiency.

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.)

Reference:

- https://arxiv.org/abs/1908.07442
- https://blog.csdn.net/m0_43455312/article/details/123297929
- https://blog.csdn.net/deephub/article/details/109044022