National Tsing Hua University

Fall 2023 11210IPT 553000

Deep Learning in Biomedical Optical Imaging

Homework 2

AUTHOR

柯志明

Student ID: 111003804

1. Coding

1.1 Coding task A: transitioning to Cross-Entropy Loss

See hw2.ipynb.

1.2 Coding task B: creating an Evaluation Code

See hw2.ipynb.

2. Report

2.1 Report task A: performance between BCE loss and BC loss

The model which we use to perform the comparison of the performance between BCE loss and CE loss is a fully connected neural network. Its architecture is designed with the following specifications:

- Layers: the model includes the input layer, a hidden layer, and the output layer.
- Width: the hidden layer has a width of 64, i.e., 64 neurons in the hidden layer.
- Activation Function: the activation function used in the hidden layer is the Rectified Linear Unit (**ReLU**) activation function.
- Dropout: the dropout rate is set to 0.0, which means no dropout is applied.
- Normalization: the **batch normalization** is applied in input layer and hidden layer.
- Initial Learning Rate: the initial learning rate is set to 0.01.
- Learning Rate Scheduler: the **CosineAnnealingLR** scheduler is applied.
- Epochs: the number of epochs is 30. Besides, a criterion of early stopping is also applied

to compare the convergence between the two loss modes.

First, we conducted training with 32 epochs for both loss modes. In Fig. 1(a), the chart displays the validation (solid line) and testing (dashed line) accuracies for both BCE mode (green line) and CE mode (red line). All the validation accuracies hover around 91% to 96%. While we can observe that the green line generally stays above the red line, it's important to note that we don't have enough evidence to conclusively claim that BCE loss mode consistently outperforms CE loss mode based on a single experiment. It's even challenging to discern which of the two dashed lines represents better testing performance.

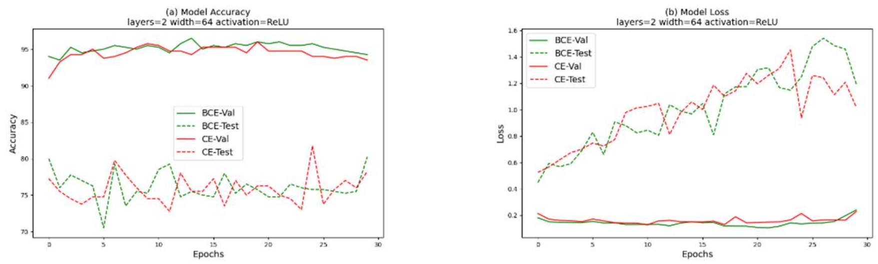

Fig. 1 (a) and (b) display the accuracies and losses, respectively, of validation and testing for both the BCE loss mode and the CE loss mode.

In Fig. 1(b), the chart displays the validation (solid line) and testing (dashed line) losses for both BCE mode (green line) and CE mode (red line). However, relying solely on the dynamics of losses does not yield a clear distinction between the two modes. Nonetheless, a noticeable trend emerges as testing losses increase with the number of epochs, while validation losses remain relatively stable. This observation suggests that training for 30 epochs may lead to overfitting with respect to the testing dataset, highlighting the need for an early stopping mechanism.

We conducted ten rounds of training, each with an early stopping mechanism. Figures 2(a) and 2(b) display the results for accuracies and losses, respectively. In the figures, lines that end before reaching 30 epochs indicate that the training was stopped early at the final epoch. Notably, we observed that more training sessions with the CE loss mode experienced early stopping compared to those with the BCE loss mode. However, in Figure 2(b), we also observed that the testing losses of the CE loss mode are, on average, higher with greater variation than those of the BCE loss mode. If we consider a model with better performance regarding the testing data we used, then, based on our experiments, the BCE loss mode should be more robust.

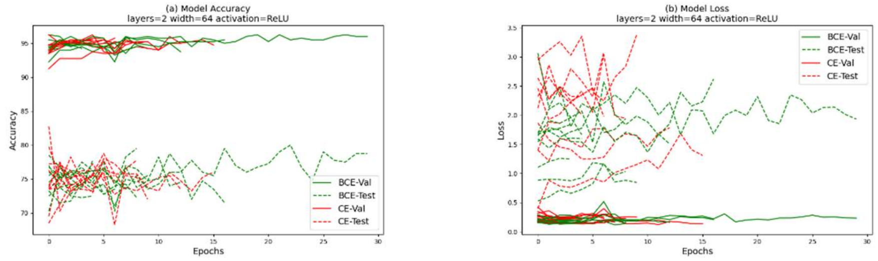

Fig. 2 (a) and (b) display the accuracies and losses of validation and testing obtained from ten individual trainings for both BCE loss mode and CE loss mode, respectively.

2.2 Report task B: performance between Different Hyperparameters

We use the same model described in Section 2.1, Task A, with the BCE loss mode as the base model for Task B. We select the batch size and the dropout rate as hyperparameters, and for each, we choose three different values. We conduct the training and compare their performances.

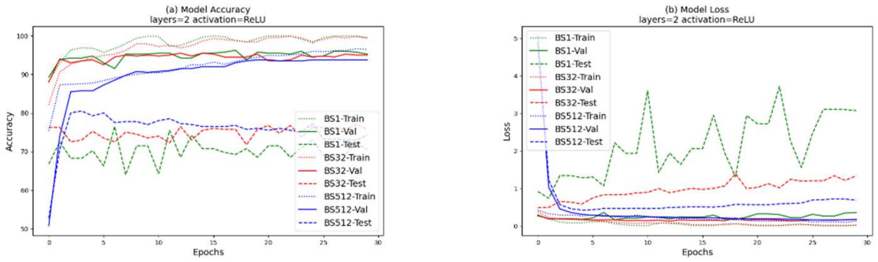

Fig. 3 (a) and (b) display the accuracies and losses according to training data (dotted line), validation data (solid line), and testing data (dashed line) during training processes conducted by three different batch sizes: 1 (BS1 in green), 32 (BS32 in red), and 512 (BS512 in blue).

Batch size - In Fig. 3(a), the chart displays the accuracies for three different batch sizes: 1, 32, and 512. For the batch size of 1, as batch normalization cannot be applied to a single sample, we employ layer normalization instead. When examining the training and validation accuracies, it is evident that BS1 and BS32 converge to a high accuracy at an earlier epoch compared to BS512, while BS512 exhibits a smoother evolution. When we check the testing accuracies, we can find that BS512 and BS32 have better and stabler performance compared to BS1.

In Fig. 3(a), the chart displays the losses for three different batch sizes: 1, 32, and 512. Fig. 3(b) provides insight into the testing losses, revealing that BS1 exhibits significant fluctuations in losses across epochs. Notably, training with BS1 requires the longest time among these three cases. From this comparison, we conclude that either BS32 or BS512 is suitable for training the model, while BS1 is not recommended.

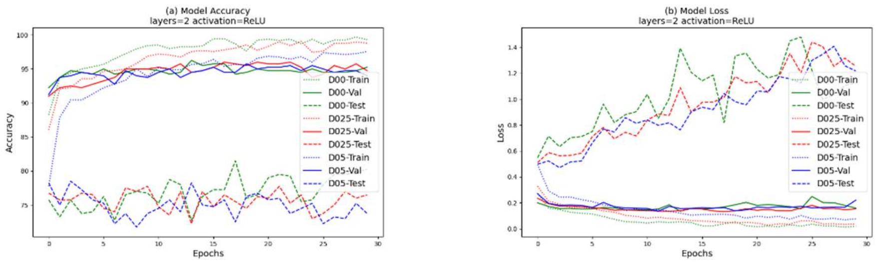

Fig. 4 (a) and (b) display the accuracies and losses according to training data (dotted line), validation data (solid line), and testing data (dashed line) during training processes conducted by three different dropout rate: 0.0 (D00 in green), 0.25 (D025 in red), and 0.5 (D05 in blue).

Dropout rate - In Figures 4(a) and 4(b), the charts display the accuracies and the losses, respectively, for three different dropout rates: 0.0, 0.25, and 0.5. Although the training accuracy evolutions are different in these three cases (D00 is the highest and D05 is the lowest), we cannot obtain a clear observation to distinguish these their performances based on the validation accuracy.

When comparing the testing accuracies and losses, an interesting phenomenon emerges: although case D00 exhibits the highest test losses on average among the three cases, it also achieves the highest test accuracies at most epochs.

2.3 Conclusion

In our study, we conducted experiments with different batch sizes, dropout rates, and loss modes, analyzing their impact on model training. Key observations include the convergence behaviors of different batch sizes, the influence of dropout rates on training and validation, and the performance of BCE and CE loss modes. Notably, our findings suggest that for our specific model and dataset, batch sizes of 32 and 512 exhibit better and more stable performance compared to a batch size of 1. Additionally, while different dropout rates affect training accuracy, they do not significantly impact validation and testing accuracies. Lastly, the BCE loss mode is shown to be more robust than CE loss mode in our experiments.

In summary, our results provide valuable insights into hyperparameter choices for training neural networks and highlight the importance of careful experimentation to select the most suitable settings for a given task.

Furthermore, it's worth noting that I observed a discrepancy in the distribution of the testing data when compared to the training and validation data. This mismatch in data distribution likely contributed to the significant drop in performance on the testing data. Ensuring that the testing data samples randomly from the same distribution as the training and validation data is crucial for obtaining reliable and consistent evaluation results.