ASSIGNMENT 2: Neural Networks (BA-64061-001)

Nishanth Gurujala
014-10-02025

Introduction

The report analyses the performance of a binary classification model on the IMDB dataset by testing the various neural network architectures. The primary aim was to study the effect of modifications to the model structure (including the number of layers and units, the choice of activation functions and use of dropout) on its loss and accuracy. The baseline model that is going to be compared to consists of 2 hidden layers of 16 units, ReLU activation functions and binary cross-entropy as the loss function.

Data Overview

The dataset in this study is the IMDB Movie Reviews Dataset which has 50,000 reviews divided into training, validation and test sets. Every review is respectively deemed either positive or negative, and thus can be classified through binary sentiment. The model allows sentences of words to be inputted to forecast sentiment. The data has been already prepared to contain the 10,000 most common words only; each review being represented as a sequence of integers indicating the words in this vocabulary.

Process

The neural network models were built using the standard IMDB codebase, and the following steps were followed in the model development process:

1. Import necessary libraries.
2. Build the model.
3. Compile the model.
4. Prepare the validation set.
5. Train the model
6. Retrain the model from scratch.
7. Evaluate the model.
8. Make predictions.

Methodology

The framework was oriented at the systematic manipulation of the various elements of the neural network structure to see the effect of each alteration on the model performance-loss and accuracy. The model configurations tried included nine different model configurations which have been summarized in a table that gives the number of layers, number of units, the activation functions, loss functions and the dropout rates.

In each configuration, the parameters were varied at a time, and all the other parameters were kept constant. The changes were in terms of dropout, activation functions, number of hidden layers and units per layer.

The first model was all trained on 20 epochs to provide a baseline. The epoch, according to the lowest validation loss per model was next determined by analysis and re-training of each model was performed on a clean slate in order to capture the best performance. The ultimate loss and accuracy values were obtained upon retraining, which guaranteed a just and reliable comparison of all the models.

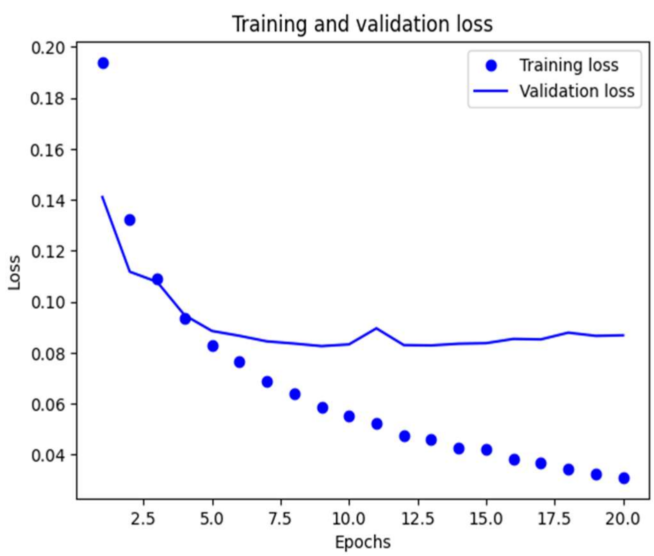
FINAL RESULTS

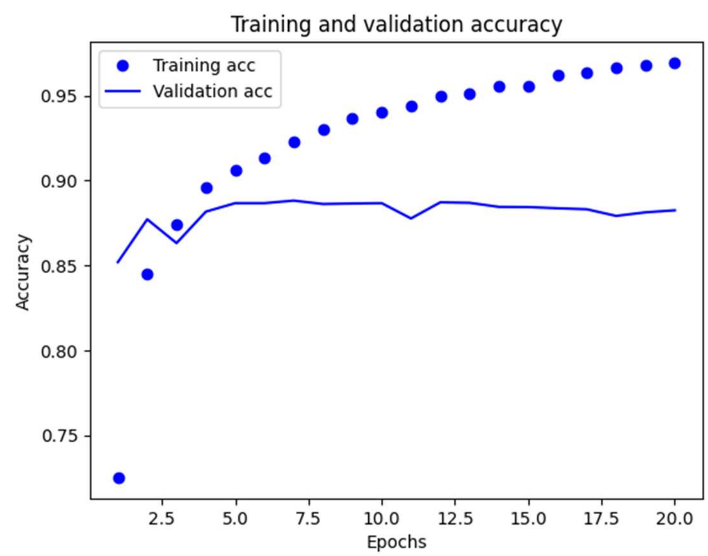
Model No.	Layers	Units per Layer	Activation	Loss Function	Regularization	Validation Accuracy	Validation Loss	Testing Accuracy	Testing Loss
1	2	16	relu	binary_crossentropy	None	87.11%	0.0992	88.46%	0.2896
2	1	16	relu	binary_crossentropy	None	88.27%	0.2966	88.93%	0.27754
3	3	16	relu	binary_crossentropy	None	85.32%	0.4412	87.62%	0.3145
4	2	32	relu	binary_crossentropy	None	87.15%	0.3781	88.37%	0.2948
5	2	64	relu	binary_crossentropy	None	85.02%	0.4514	88.40%	0.293
6	2	16	tanh	mse	None	86.65%	0.1058	87.60%	0.0916
7	2	16	relu	mse	None	87.38%	0.0948	88.37%	0.0862
8	1	16	relu	mse	L2 regularization	87.31%	0.0969	88.87%	0.0825
9	1	16	relu	mse	Dropout (0.5)	88.79%	0.0822	88.45%	0.0879
10	1	32	tanh	mse	Dropout (0.5)	88.35%	0.0863	88.47%	0.0844

<u>Model Performance Analysis-Model 9</u>

The best validation accuracy of all the models (88.79) was reached with Model 9, and it has one hidden layer with 16 units, ReLU activation, MSE loss, and dropout rate of 0.5. Both training and validation accuracy improved steadily in the course of training, which shows that learning was continually occurring. The low distinction between training and validation accuracies indicates high generalization. Further, training and validation loss curve exhibited a progressive and slow decrease and leveled off after a few epochs. This pattern proves that the model efficiently learned without overfitting and detected meaningful patterns, at the same time being stable.

Training and Validation Loss & Training and Validation Accuracy





1. Training vs. Validation Accuracy

There was a steady increase in training accuracy with every epoch, and this proves the effectiveness with which the training data is learned. The same upward trend was observed in the validation accuracy that showed a stable value of 88.45, which corresponds to excellent generalization. The gap between the training and validation accuracy is low, indicating that the model has not been overfitted in the process of training.

2. **Training vs. Validation Loss**

The training loss as well as the validation loss reduced gradually which meant that the model minimized the error and learned important patterns. The fact that there are no sharp peaks in the validation loss curve implies even further the stability of the model and its resilience to overfitting.

Architectural Impact

The performance of the model was sensitive to the number of hidden layers. Interestingly more complex or deep networks did not necessarily perform better. For example:

- Model 2 (1 hidden layer, 16 units) had 88.27% validation accuracy.

- Model 3 (3 hidden layers, 16 units each) had marginally worse validation accuracy (85.32%), indicating that the improvement in accuracy with depth failed.

- Model 5 (2 layers, 64 units per layer) lowest performance gave 85.02 percent, which could be attributed to overfitting the model because of the large capacity.

These findings underscore the need to select balanced architecture and not just deepen or widen.

Activation Function Effect

The convergence and stability were influenced by the activation function. ReLU models tended to be much more stable in the training phase and more accurate in validation than Tanh models. For instance:

- Model 1 (ReLU) - 87.11% validation accuracy.

- Model 6 (Tanh, same structure) - 86.65% validation accuracy.

Whereas Tanh may be good in some situations, Tanh is more likely to be saturated making it slower to learn than ReLU.

Loss Function Comparison

The classification performance was affected by the loss function to a great extent. Binary Crossentropy models (Models 1-5) all had better validation accuracies than the MSE models (Models 6-10):

- Model 2 (Binary Crossentropy, ReLU) - 88.27%
- Model 7 (MSE, ReLU) - 87.38%

This is in accordance with best practice because Binary Crossentropy would be more applicable to binary classification whereas MSE would be more applicable to regression tasks.

Regularization: L2 vs. Dropout

Regularization was successful in reducing overfitting. For example:

- Model 8 (L2 regularization) - worst validation accuracy of 87.31 - it is possible that regularization may inhibit learning when regularized too much.
- Model 9 (Dropout 0.5) - the model has the best validation accuracy of 88.79 and shows that dropout can be used to enhance generalization due to the lack of co-adaptation of neurons.

The model 10 (Dropout 0.5, Tanh) performed with 88.35, which justifies the assumption that dropout is effective, yet the selection of the activation function also has a great influence on the performance.

Conclusion

Model 9, which used a single dropout (0.5), had one hidden layer of 16 units, ReLU activation, and MSE loss, and was the most successful model. It recorded the best validation accuracy of 88.79, showing that the use of a rather basic architecture and dropout can be used to achieve good generalization. On the whole, these results help to highlight the necessity of balancing the complexity of the architecture and carefully adjusting regularization measures rather than pay attention only to deeper or more complex models.