**Report**

**Introduction**

The dataset of IMDB containing 25000 movie review with the classes positive and negative were loaded and the pre-processing included limiting the number of words to 10000. It is to be noted that each of these reviews was first encoded as a sequence of integers which corresponds to the index of words in a vocabulary. Specifically, the data was pre-processed to fit a neural network algorithm and was one hot encoded with each review represented as the neural network input, a vector of 10000 wherein only the indices that correspond to the words used in the reviews were set as one while all other indices were set to zero. This altered the training and testing data shapes into the binary matrices of (25,000 X 10,000). Also the labels that were saved as integer (0 for negative and 1 for positive), they were then changed to float 32 to match the model.

**Model Building**

**Model built with in class:**

The Model is sequential.

1. First Hidden Layer: 16 neurons with ReLU activation

2. Second Hidden Layer: 16 neurons with ReLU activation

3. Output Layer: 1 neuron with Sigmoid activation (for binary classification)

**Training Summary**

The model used RMSprop as an optimizer and binary cross-entropy loss function

There are distinguished validation set, consisting of 10,000 samples and training set; the number of samples in it amounts to 15,000.

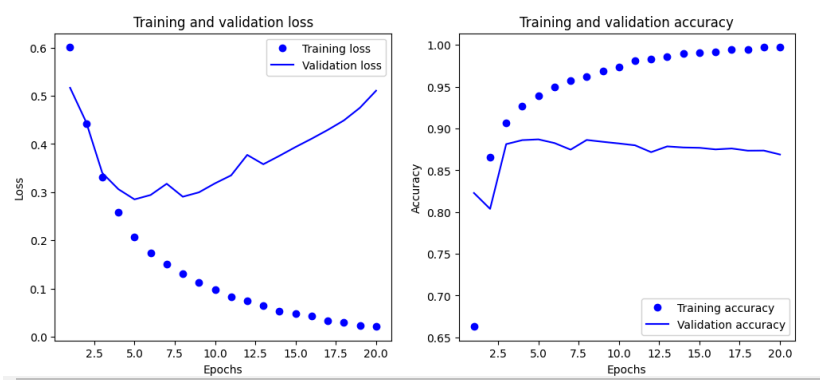
Here, the model was trained for 20 cycles with 512 batches.

**Performance Observations**

The training efficiency was also gradually improved during the epochs from the beginning of the training to the final epoch which is 58.39% to 99.71% respectively.

The validation accuracy was range bound mainly stayed high at 88.61% up to the fourth epoch and thereafter decreased and stabilized at 86.89%.

The validation loss at first declined and then, from epoch 6 it rose up, hence, pointing toward overfitting.

** Best performing Epochs:**

Examining both maximum validation accuracy and minimum validation loss until the overfitting stage allows identification of the best-performing epochs.

From your results:

Epoch 4: Validation Accuracy: 88.61%, Validation Loss: 0.3062 (Peak Accuracy)

Epoch 5: Validation Accuracy: 88.70%, Validation Loss: 0.2851 (Lowest Validation Loss)

**Model with 1 hidden layer and 3 hidden layers**

Measurement of validation accuracy at 88.59% occurred during Epoch 6 when the model produced a validation loss of 0.2751. The model starts to overfit after Epoch 4 because the validation loss rises without any corresponding change in accuracy. The model with three hidden layers achieved its highest performance during Epoch 1 because it delivered 87.54% validation accuracy together with a validation loss of 0.3981. The model begins overfitting rapidly beginning from Epoch 3 when validation loss starts to rise even though accuracy levels off or decreases. Among the presented models the basic model achieves better generalization but the deeper model shows rapid overfitting behavior.

**Comparison**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Best Epoch** | **Validation Accuracy** | **Validation Loss** | **Overfitting Trend** |
| 1 Hidden Layer | 6 | 88.59% | 0.2751 | Overfitting starts after epoch 6 |
| 2 Hidden Layers | 5 | 88.70% | 0.281 | Overfitting starts after epoch 6 |
| 3 Hidden Layers | 1 | 87.54% | 0.3981 | Overfitting starts after epoch 3 |

**Model with Tanh and mse Loss Function with 32 units:**

The research shows that the single-hidden-layer network produced outstanding results due to its highest validation accuracy of 88.59% alongside test accuracy of 85.36%. The findings point to a conclusion that basic architecture designs produce superior outcomes in this situation because they minimize overfitting problems. A neural network with two hidden layers using the tanh activation function together with MSE loss experienced poor performance as it reached only 48.65% test accuracy though training results were good. The high level of overfitting demonstrates that MSE is inappropriate as a loss function for classification tasks since it produces incorrect gradient updates for binary classification tasks. The test data accuracy from the three-hidden-layer model exceeded the two-hidden-layer model by a slight margin while remaining subject to overfitting at 50.03%. New hidden layers did not enhance the overall performance and resulted in inferior generalization capabilities. The dataset shows a better response when a neural network design with fewer layers is used for the specified task.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Best Epoch** | **Validation Accuracy** | **Validation Loss** | **Test Accuracy** |
| 1 Hidden Layer | 6 | 88.59% | 0.2751 | 85.36% |
| 2 Hidden Layers (tanh, MSE) | 1 | 86.83% | 0.6041 | 48.65% |
| 3 Hidden Layers | 1 | 87.54% | 0.3981 | 50.03% |

**Model with regularization Technique**

The combination of dropout together with L2 regularization succeeded in providing better generalization when contrasted with standalone models that used no such techniques. Both Dropout (0.5) and L2 regularization (0.001) functioned to prevent overfitting since they maintained consistency between the test accuracy (85.86%) and validation accuracy. The training procedure began with 86.49% validation accuracy that stayed constant across epochs because of proper regularization techniques. The maximum training accuracy achieved 96.60% without showing signs of major overfitting behavior thus indicating that dropout effectively stopped train-data memorization.

The use of Mean Squared Error as the loss-function appears suboptimal for a classification challenge compared to Binary Cross-Entropy. The most favorable choice for binary classification tasks is the implementation of cross-entropy since it delivers superior gradient updates. The implemented model showed solid results although using cross-entropy as the loss function could possibly improve both performance and convergence speed. The simplified algorithm has an optimal performance at epoch 16 but shows a reduction of validation accuracy that could indicate the dropoutrate value of 0.5 is too high because it potentially causes underfitting. The optimal dropout rate should fall between 0.3 and 0.4 because it would achieve the best balance between regularization and learning capacity.

**Summary/Overall Comparison**

The newest model design affords increased generalization capabilities when compared to previous versions. Training accuracy levels for the non-regularized models soared but they concurrently generated inadequate validation and test accuracy measures. This regularized model displayed consistent performance metrics between all dataset divisions including training, validation and testing. These dropout layers protected the model from developing a strong dependence on particular neurons which improves its stability when applied in actual usage. Using this decision made the model slightly less accurate during training to achieve superior generalization results. The model can be improved through using the Adam optimizer instead of RMS prop to accelerate its convergence while enhancing stability and optimizing the dropout rate to strike an ideal balance between underfitting and overfitting.

