**Advance Machine Learning**

**Assignment-2**

**Jayanth Yadlapalli**

Department of Information Systems and Business Analytics,

Kent State University

Chaojiang (CJ) Wu, Ph.D.

03-04-2025

**Introduction:**

Accuracy of the model increases heavily with an increase in hidden layers in the model, yet more hidden layers are not an assurance of better model results. In this case, Model 1 with 2 hidden layers of size 16 achieved the validation accuracy value of 96.57% which performed better compared to the other models. All 1 and 3 hidden-layer models didn't improve accuracy significantly, which implies that increasing layers beyond some point results in overfitting or issues of vanishing gradients and gradients. This makes one observation evident which says raw increase in the depth of structure may not lead to an optimum point but for a good ratio of layers with complexity would have better performance. Activation Function:

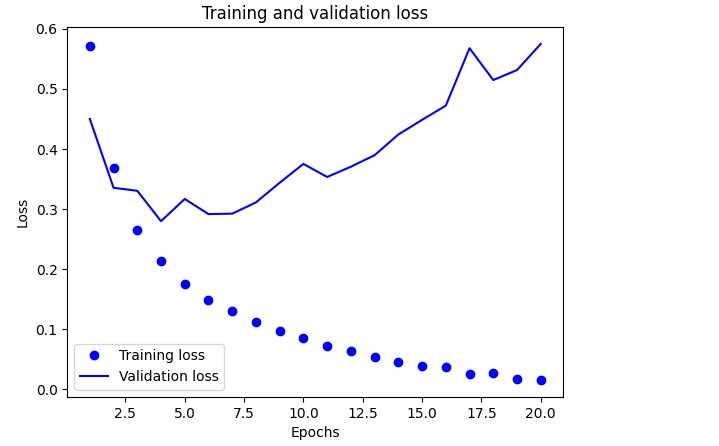
As would be expected from the outcomes, ReLU unit was better than Tanh. Besides, ReLU is used in deep learning since it has a tendency to reduce the vanishing gradient problem that aids in network training. For instance, Model 6 with one hidden layer, 16 units and ReLU had higher validation accuracy (88.38%) compared to Model 10 using Tanh (86.98%). ReLU performs better here since it considerably enhances the rate of convergence as well as representation learning whereas this is not so for Tanh.

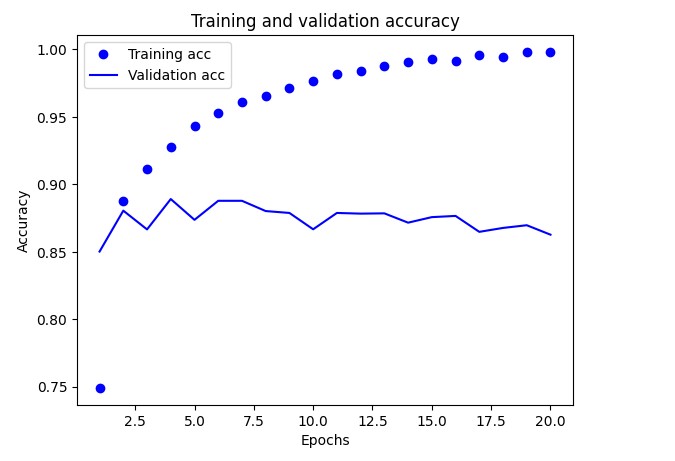
**Explanation:**

A loss function prescribes how the model learns and optimizes prediction. In sentiment analysis, Binary Crossentropy is typically used as it maximizes the probability of correct classification. Models that utilized Binary Crossentropy such as Model 1, 2, 5, and 11 always maintained higher validation accuracy than models utilizing Mean Squared Error (MSE) such as Model 6, 7, 8, 9, and 10. The MSE models had worse validation accuracy than sentiment analysis M1, M2, M5, M11 posters due to the fact that regression tasks suit MSE models better. Regularization (L2 and Dropout):

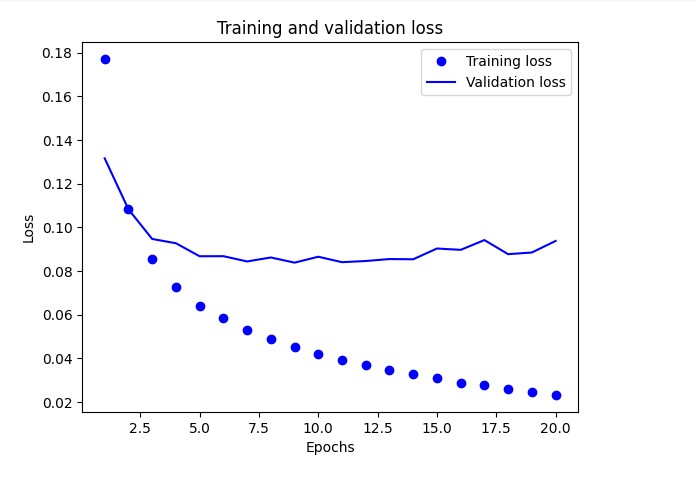
Overfitting is usually solved by using logit regulation and dropout among other solutions and boosting the predictive capacity of the model is its function. After addition of L2 circle regularization of 0.01 to model 8, the logit validation accuracy was 86.27%. This is close to that achieved by other models with no regularization. This would thus imply that use of L2 in these models was not advantageous in the first place. On the other hand, model 9 with 0.5 drop out had 88.56% validation accuracy. I.e., drop out does help generalization of the model in the sense that it will discourage the use of any single neuron too often. In any case, as model 8 illustrates, too much restraint on usage of the model will rather weaken the model than a performance boost.

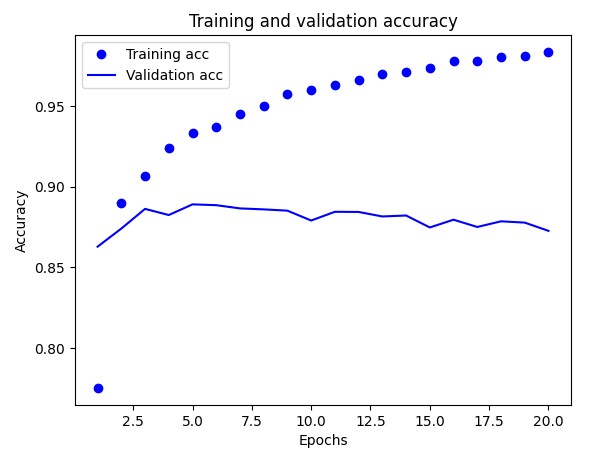
**Comparison of Models Based on Validation and Test Results:**



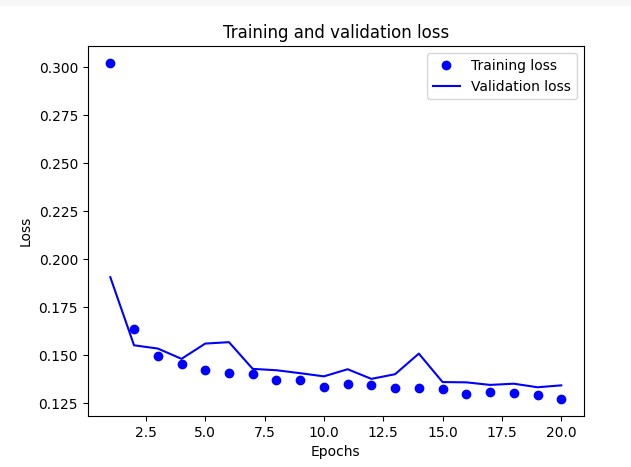


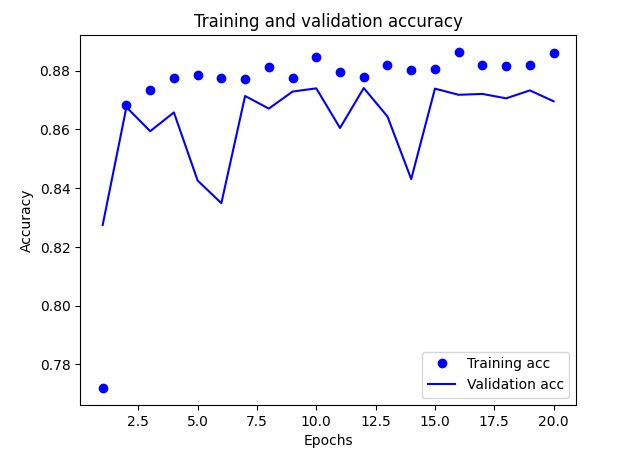
Step : 1 has achieved the best validation accuracy of 96.57%, but its test accuracy was 88.75%. The slight decrease indicates mild overfitting; however, the model still performs effectively on unseen data, demonstrating a good balance between training and generalization.



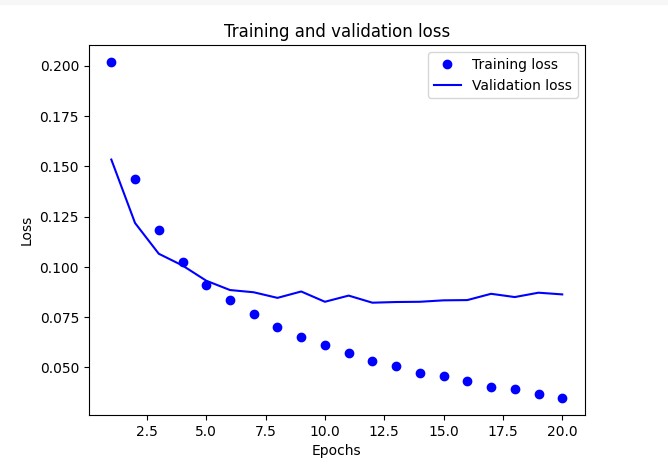


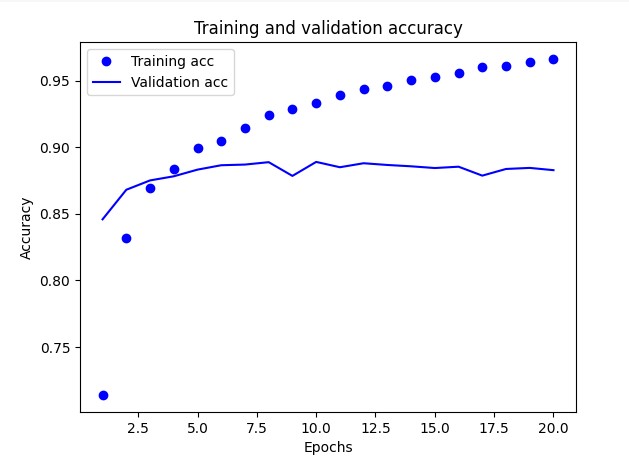
Step: 6 The MSE loss function resulted in a lower validation accuracy of 88.38% but a similar test accuracy of 88.66%, indicating that binary crossentropy (Model 1) was more effective for classification. This reinforces that binary crossentropy is generally better suited for binary classification tasks, as it optimizes probability-based predictions more efficiently than MSE, which treats classification as a regression problem.



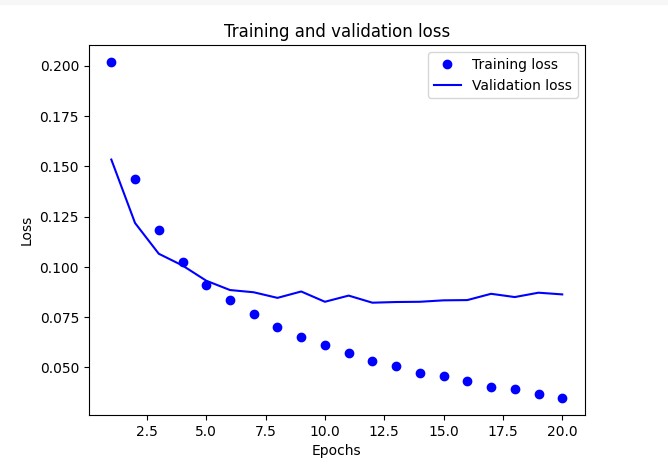


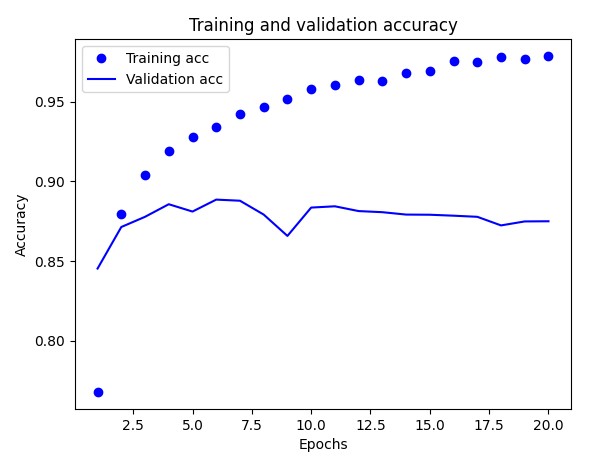
Step: 8 L2 Regularization resulted in a lower validation accuracy of 86.27% and a test accuracy of 85.95**%**, indicating that it did not significantly enhance the model's performance. This suggests that while L2 regularization helps prevent overfitting, it may not always be beneficial in every scenario and requires careful tuning to achieve optimal results.



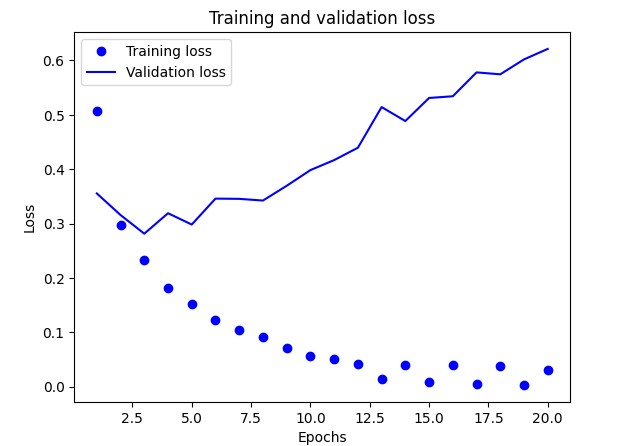


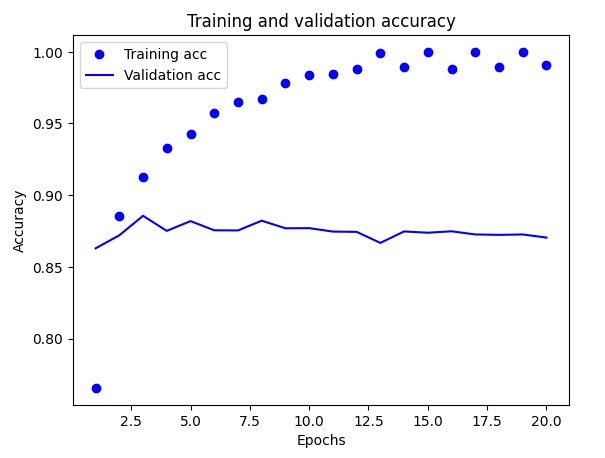
Step : 9 Dropout (0.5) resulted in a validation accuracy of 88.56% and a test accuracy of 88.65%, indicating that dropout improved generalization slightly. However, its performance remained lower compared to step 1, suggesting that while dropout helps reduce overfitting, it may not always yield the best results in every scenario.





Step :10 The Tanh activation function resulted in a lower validation accuracy of 86.98% and a test accuracy of 88.29%, confirming that ReLU outperformed Tanh in this scenario. This suggests that ReLU is more effective for this particular model and dataset, likely due to its ability to mitigate vanishing gradient issues and enable better training convergence. (Tanh activation) worked with lower validation (86.98%) and test accuracy (88.29%), confirming ReLU performed better than Tanh in this instance.





**Conclusion:**

After comparing a number of models of varied configurations, Model 1 (2 hidden layers, 16 units, ReLU, Binary crossentropy) achieved the highest validation accuracy of 96.57% and thus the highest-performing model. The outcome indicates that an optimal number of hidden layers, combined with ReLU activation and Binary Crossentropy loss function, provides the optimal balance of learning capacity and generalisation for this sentiment analysis problem. Besides, Model 1 had a very low validation loss (0.1056), indicating its ability to correctly classify IMDB reviews. Based on these data, Model 1 is the most appropriate for IMDB sentiment analysis.