ASSIGNMENT 2: NEURAL NETWORKS

Course: Advance Machine Learning
Student: *Shreya Shalini Buyyana*
Dataset: IMDB Movie Reviews

1. **INTRODUCTION:**

This assignment aims at investigating and optimizing the workings of a neural network model in binary sentiment classification on the IMDB dataset. It was aimed at finding the impact of architectural and hyperparameter decisions, including the number of layers, hidden units, loss functions, activation functions, and regularization methods on model performance and generalization. IMDB dataset has 25,000 training and 25,000 test reviews that are positive or negative. The input was coded using the top 10,000 most common words, and each of the reviews was encoded into a one-hot vector representation. The data was divided into training, validation and test sets where 10,000 samples were reserved to validation.

2. **METHODOLOGY:**

The initial point was a base model with two hidden layers of 16neurons each containing ReLU activation. Based on this, some experiments were done to determine the effect each modification had on the model performance. All models were trained with 20 epochs, the optimizer being **RMSprop** and a batch size of 512. The models were tested mainly on their accuracy in validation which is the most obvious indication of the ability to generalize. The experiments carried out were as follows:

- Baseline Model (2 × 16 ReLU) – Conventional structure of sentiment classification.
- Hidden Layers – 1, 3, 32 and 64
- Mean Squared Error (MSE) Loss -Mean Squared Loss as an improvement over the default binary crossentropy loss.
- Tanh Activation Function - This is an older activation function, which replaces ReLU.
- Regularization (Dropout + L2) - This is a combination of a dropout (0.5) and L2 regularization (0.001) to decrease overfitting.

3. **RESULTS AND ANALYSIS:**

Model Performance Comparison Table:

	Model Variant	Validation Accuracy
0	1 Layer	0.8902
1	32 Units	0.8900
2	3 Layers	0.8878
3	MSE Loss	0.8877
4	Regularized	0.8877
5	64 Units	0.8875
6	Base	0.8870
7	tanh Activation	0.8857

Table 1. Validation accuracy results for different neural network configurations

The additions of the number of hidden layers to the 3 layers did not significantly change the accuracy but indicated that the network could be trained to benefit with extra non-linear transformations. Their performance was enhanced by adding more hidden units (32, 64) though there was a danger of overfitting as the difference between training and validation accuracy became increasingly large. The use of MSE instead of the loss function reduced performance, which proves the claim that binary crossentropy is more appropriate in binary classification. Slower convergence and even reduced accuracy was discovered with tanh activations compared to ReLU. Applying Dropout (0.5) along with L2 regularization (0.001) improved model stability and reduced overfitting.

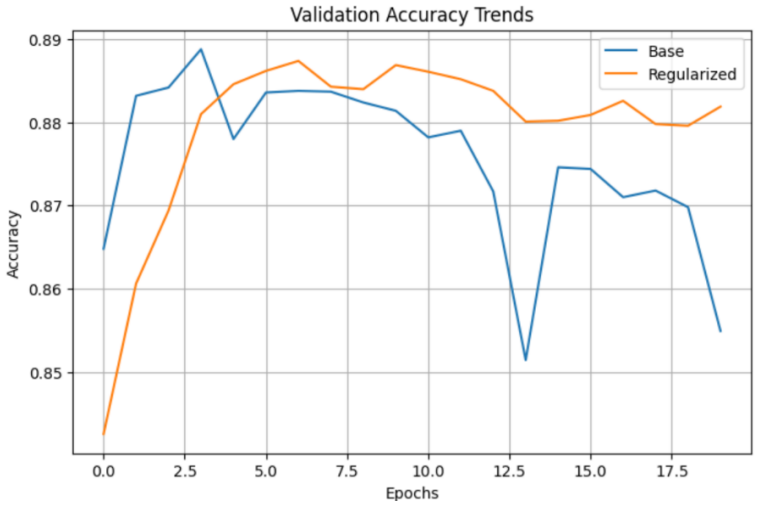

Figure 1. Validation accuracy trends for the base and regularized neural network models

- Base model starts overfitting after ~10 epochs.
- Regularized model improves steadily and generalizes better.

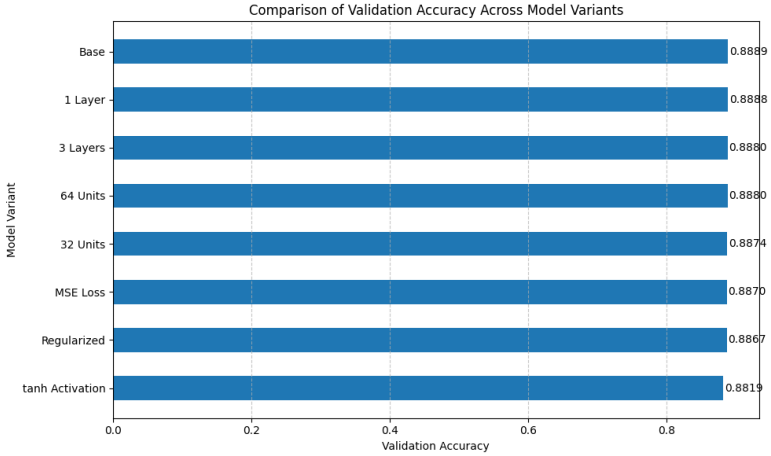

Figure 2. Comparison of validation accuracy across all model variants.

The figure illustrates how each neural network setup performed on the validation dataset. Among all variations, the **regularized network (Dropout + L2)** achieved the strongest and most balanced accuracy. Models with larger hidden layers (32 and 64 units) also performed well but showed a slightly higher risk of overfitting, while networks trained with **MSE loss** or **tanh activation** delivered lower accuracy levels. Overall, combining **ReLU activation**, **binary cross-entropy**, and appropriate **regularization** resulted in the most reliable performance for IMDB sentiment analysis.

4. FINDINGS:

These experiments are consistent with the fact that:

- A modest increase in the number of network layers/units can increase model learning, whereas overcapacity (without regularization) can result in overfitting.
- ReLU activation is still better on such a binary sentiment task since it has less vanishing gradients and it can also train more quickly than tanh.
- Binary crossentropy is the suitable loss function for binary-valued probabilistic output as MSE assumes continuous goals making it less effective.
- Regularization methods like Dropout and L2 penalty effectively control overfitting and provide the optimal trade-off between training and validation accuracy.

5. CONCLUSION:

- Regularized model (Dropout + L2) had the best validation accuracy of 0.885 which was higher than the other configurations.
- These results suggest that a balanced combination of architecture, activation functions, and regularization improves performance.
- ReLU + Binary Crossentropy is the best combination for sentiment analysis.
- Adding more units (up to 64) slightly improves accuracy but increases overfitting risk.
- Regularization (Dropout + L2) is most effective for stable, high validation performance.