Assignment 1: Neural Networks

Purpose:

The goal of this assignment was to investigate and enhance a neural network model for binary classification using the IMDB movie review dataset. We experimented with different hyperparameters and techniques to evaluate their effects on validation and test accuracy, including:

- Adjusting the number of hidden layers.
- Altering the number of hidden units.
- Trying out various loss functions (like `binary_crossentropy` compared to `mse`).
- Evaluating different activation functions (`relu` versus `tanh`).
- Implementing dropout regularization.

Data Preprocessing:

1. **Dataset**:
    - The IMDB dataset includes 50,000 reviews that are evenly divided into "positive" and "negative" sentiments.
    - To minimize dimensionality, only the top 10,000 most frequently used words were taken into account.
2. **Text Representation**:
    - Reviews were transformed into sequences of integers, with each integer corresponding to a specific word index.
3. **Padding**:
    - All reviews were adjusted to a consistent length of 500 words to ensure uniform input size.
4. **Train-Test Split**:
    - The dataset was split into 25,000 training samples and 25,000 test samples.

Experiments and Results

Experiment	Validation Accuracy	Test Accuracy	Key Observations
Base Model (2 hidden layers)	~88%	~87%	Baseline performance using two dense layers, each with 16 units and employing `relu` activation.
1 Hidden Layer	~87%	~86%	Slightly lower accuracy indicates a reduced ability to learn complex patterns.
3 Hidden Layers	~88%	~87%	Performance is similar to the base model, with no notable improvements.
32 Hidden Units	~86%	~86%	The lower capacity resulted in slightly diminished performance compared to the base model.
64 Hidden Units	~88%	~87%	Increased capacity showed performance on par with the base model, without signs of overfitting.
Loss Function (mse)	~85%	~86%	Accuracy was slightly lower than `binary_crossentropy`, as `mse` is not ideal for classification tasks.

Activation Function (tanh)	~87%	~86%	Performance was comparable to `relu`, but training took longer due to vanishing gradients.
Dropout Regularization	~86%	~86%	Dropout helped mitigate overfitting, but it required careful adjustment of the dropout rate for optimal results.

Key Findings

1. The base model, which consists of 2 hidden layers with 16 units each, uses `relu` activation and `binary_crossentropy` loss. This setup delivered a solid baseline performance, achieving a test accuracy of approximately 87%.
2. Increasing or decreasing the number of hidden layers (1 or 3 layers) had little effect on performance.
3. Changing the number of hidden units:
    - Using fewer units (32) slightly decreased accuracy because of limited learning capacity.
    - More units (64) kept performance similar without causing overfitting.
4. Using mean squared error `mse` as the loss function resulted in lower accuracy compared to `binary_crossentropy`, indicating that `binary_crossentropy` is more appropriate for binary classification tasks.
5. Switching from the `relu` to the `tanh` activation function produced similar results but slowed down training due to vanishing gradients.
6. Applying dropout regularization effectively reduced overfitting when a moderate dropout rate (20%-30%) was used. However, excessive dropout (50%) resulted in underfitting and poor performance.

Conclusion

The top-performing configuration was the base model, which reached a test accuracy of approximately 87% with:

• Two hidden layers.

• 16 units in each layer.

• `relu` activation function.

• `binary_crossentropy` loss function.

While dropout regularization enhanced generalization when fine-tuned, it did not significantly exceed the performance of the base model.