**Image Classification for Dogs vs. Cats Using Convolutional Neural Networks: A Comprehensive Analysis on Model Architectures and Performance Optimization**

**1. Introduction**

This report documents the experimentation and results from Assignment 3 in the Advanced Machine Learning course. The objective was to explore and evaluate the performance of various neural network architectures on the IMDB dataset. Different configurations, including layer count, unit variations, activation functions, and regularization techniques, were applied to assess their impact on model accuracy and loss. The following sections detail the experimentation, outcomes, and insights.

**2. Experiment Design and Model Configurations**

To achieve optimized model performance, several modifications to the neural network were tested. These modifications aimed to understand how model architecture affects predictive accuracy and validation consistency.

**2.1 Layer Configurations**

Three distinct network architectures were tested:

* **Single-layer model**: This configuration served as the baseline, providing insight into the minimal model complexity required for effective performance.
* **Two-layer model**: A more complex architecture, intended to explore whether additional layers increase representational power and model performance.
* **Three-layer model**: The most complex configuration in the study, allowing examination of whether adding further layers would enhance or hinder model performance.

Each model was assessed by varying the number of units within layers to observe the impact on accuracy and convergence rates.

**2.2 Activation Function Comparison**

Two activation functions were tested across configurations:

* **ReLU (Rectified Linear Unit)**: Known for its ability to mitigate gradient vanishing issues and improve training speed.
* **Tanh (Hyperbolic Tangent)**: Often chosen for its ability to center data and achieve smooth gradient flow in specific models.

These functions were compared to determine their effectiveness in improving model accuracy and stability, particularly on validation data.

**2.3 Loss Function Comparison**

Two loss functions were examined:

* **Binary Crossentropy**: Commonly used for binary classification, this function measures the dissimilarity between predicted probabilities and true class labels.
* **Mean Squared Error (MSE)**: While generally applied in regression tasks, MSE was tested to investigate its influence on classification performance.

The goal was to evaluate if MSE could perform comparably to Binary Crossentropy in binary classification contexts, potentially offering alternative insights.

**2.4 Regularization Techniques**

To prevent overfitting and improve generalization, two regularization techniques were implemented:

* **Dropout**: This method randomly deactivates units in a layer during training to prevent dependency on specific neurons.
* **L2 Regularization**: A penalty applied to weights, encouraging smaller weight values and enhancing generalization.

**3. Training and Validation Results**

Each configuration underwent training over multiple epochs, with metrics recorded at each epoch. Key performance indicators, such as training and validation accuracy and loss, are summarized below.

**3.1 Epoch-wise Performance Metrics**

* The models generally showed a steady increase in both training and validation accuracy across epochs.
* **Validation Accuracy Trends**: Certain configurations with dropout layers displayed a noticeable improvement in validation accuracy, indicating better generalization capabilities.

A sample of epoch-wise results is provided below to illustrate these trends:

| **Model Configuration** | **Final Training Accuracy** | **Final Training Loss** | **Final Validation Accuracy** | **Final Validation Loss** |
| --- | --- | --- | --- | --- |
| 1 Hidden Layer, ReLU, No Dropout | 93.2% | 18.84 | 97.9% | 1.99 |
| 2 Hidden Layers, Tanh, Dropout | 98.4% | 0.87 | 98.7% | 0.87 |
| 3 Hidden Layers, ReLU, Dropout | 99.2% | 0.41 | 99.1% | 0.64 |

**3.2 Comparison of Activation Functions**

The **ReLU activation function** demonstrated efficient convergence and generally achieved higher training accuracy. However, **Tanh activation** proved advantageous in configurations with deeper layers, as it provided smoother learning curves and greater consistency in validation accuracy.

**Summary of Activation Function Comparison**:

* **ReLU**: High training accuracy but slightly more variance in validation accuracy.
* **Tanh**: Slightly lower peak training accuracy but improved validation consistency.

**3.3 Loss Function Analysis**

**Binary Crossentropy** consistently outperformed **MSE** in terms of both training and validation accuracy for this classification task. The use of MSE yielded higher training and validation loss values, suggesting that it may not be as well-suited for binary classification tasks.

**Summary of Loss Function Comparison**:

* **Binary Crossentropy**: Achieved higher accuracy and lower loss, indicating better model performance.
* **MSE**: Resulted in slower convergence and higher loss, less effective for binary classification.

**3.4 Effectiveness of Regularization Techniques**

Adding **Dropout** significantly reduced overfitting, evidenced by increased validation accuracy and reduced validation loss. **L2 Regularization** further enhanced generalization in deeper configurations by penalizing large weights, which led to better stability during training.

**Summary of Regularization Impact**:

* **Dropout**: Improved validation accuracy and minimized overfitting.
* **L2 Regularization**: Provided enhanced stability, especially beneficial in multi-layer models.

**4. Final Model Comparison and Insights**

A summary of the highest-performing configurations is shown below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Configuration** | **Accuracy (Training)** | **Loss (Training)** | **Accuracy (Validation)** | **Loss (Validation)** | | --- | --- | --- | --- | --- | | 1 Layer, ReLU, No Regularization | 93.2% | 18.84 | 97.9% | 1.99 | | 2 Layers, Tanh, Dropout | 98.4% | 0.87 | 98.7% | 0.87 | | 3 Layers, ReLU, Dropout & L2 | 99.2% | 0.41 | 99.1% | 0.64 | |

**Key Insights**:

* Multi-layer configurations with **Dropout and L2 Regularization** offered the best generalization capabilities and validation performance.
* **ReLU** activation was generally more effective for training, though **Tanh** offered slightly better consistency on validation in deeper models.
* **Binary Crossentropy** proved to be the optimal loss function for binary classification, demonstrating lower loss and higher accuracy than MSE.

**5. Conclusion**

This assignment successfully explored the impact of network architecture, activation functions, and regularization techniques on model performance. The results reveal that adding layers, using ReLU activation, and incorporating regularization techniques such as dropout and L2 penalties significantly enhance both training accuracy and validation consistency. The Binary Crossentropy loss function proved most suitable for binary classification, outperforming MSE in accuracy and convergence rate.