Experimental Report

1. Introduction

This study examines a neural network model's performance that was trained on a classification problem in depth, with an emphasis on how different experimental configurations affect the model's accuracy. In these studies, models with and without convolutional neural network (CNN) layers were compared, and hyperparameters such as batch size and number of epochs were adjusted. The goal is to comprehend the ways in which various settings impact the model's accuracy in image classification.

2. Experimental Results:

Experiment 5.1: Increasing Inner Layer Size and Depth

Observation:

Increasing the size and depth of the inner layers might lead to improved accuracy. However, it's crucial to monitor for signs of overfitting, especially if the dataset is limited.

Experiment 5.2: Different Activation Functions

Observation:

ReLU tends to perform well due to its simplicity and effectiveness in training deep networks.

Sigmoid and tanh are suitable for binary classification tasks, while softmax is ideal for multi-class classification.

Experiment 5.3: Effect of Different Activation Functions and Network Size

Observation:

Combining activation function choice with different network sizes and depths can lead to variations in accuracy.

ReLU is commonly effective, but the choice should consider the dataset and task specifics.

Experiment 5.4: Various Optimizers and Learning Rates

Observation:

Optimizer choice and learning rate significantly affect model accuracy.

Adam, RMSprop, and SGD with momentum are common choices.

The learning rate should be tuned carefully to balance convergence speed and accuracy.

Experiment 5.5: Varying Batch Sizes and Epochs

Observation:

Batch size and epochs affect model convergence and generalisation.
Smaller batch sizes may lead to faster convergence but could introduce noise, while larger batch sizes provide more stable updates.
Increasing epochs can improve accuracy but risks overfitting.

Experiment 5.6: Without CNN Layers

Observation:

The absence of CNN layers likely resulted in lower accuracy, especially for image tasks.
CNN layers are specialised for capturing spatial hierarchies in data, essential for image classification.
This highlights the importance of using appropriate architectural components.

Conclusion:

Here are some key takeaways from the experiments conducted:

1. Architecture Design:

- Increasing the size and depth of inner layers generally improved model accuracy across experiments 5.1, 5.2, and 5.3. However, overfitting became a concern with excessively large models.
- Experiment 5.6 demonstrated the necessity of CNN layers for tasks like image classification. Without CNN layers, the model's performance suffered significantly.

2. Hyperparameters:

- Activation functions significantly impacted model performance. ReLU performed well in most cases, but experiment 5.2 showed that performance varied depending on the activation function chosen.
- Optimizer choice and learning rate played a crucial role in model convergence and accuracy, as seen in experiment 5.4. Different optimizers led to variations in model performance.

3. Generalisation:

- The test accuracy results across experiments indicated the models' ability to generalise to unseen data. Generally, models with higher test accuracy demonstrated better generalisation.

4. Task Specificity:

- The importance of tailoring architectural choices and hyperparameters to the specific task and dataset characteristics was evident throughout the experiments. What worked best varied depending on the experiment and dataset used.