CNN Comparative Study on Fashion MNIST Dataset Using Various Optimizers

By: Mohamed Motaz and Ahmed Mostafa

Introduction

The study explores the application of different convolutional neural network (CNN) architectures on the Fashion MNIST dataset, using various optimizers including Adam, RMSprop, and SGD. The primary focus is evaluating these networks' performance under different conditions, such as with dropout and data augmentation.

Methodology

A total of 23 different models were tested encompassing a range from a basic neural network with no convolutional layers to more complex structures like the simple CNN and Net5 architectures. Each model was run with the aforementioned optimizers. Notably, data augmentation was employed specifically with the Net5 model to observe its impacts.

Results Summary

General Performance

Each model was structured to include variations such as dropout at a probability of 0.5 and the employment of data augmentation techniques. All models maintained a consistent learning rate of 0.001 and were run over 4 epochs.

Model Variations and Accuracies

1. **Fully Connected Neural Network:**
   - No convolution layers were used, consisting merely of 3 NN layers.
2. **Simple CNN:**
   - This architecture included 1 convolution layer, 1 pooling layer, and 3 fully connected layers. It achieved the highest validation accuracy of 95.56% and a testing accuracy of 90.60%.
3. **Advanced CNN Configurations:**
   - Addition of more convolution and pooling layers provided varied results, but generally, less efficacy than the simple CNN setup.
4. **Net5 Architecture:**
   - The application of this architecture with and without modifications like sigmoid activation demonstrated substantial diversity in performance outcomes, particularly when juxtaposed with data augmentation techniques.

Optimizers Performance

- Adam optimizer frequently provided higher validation accuracies across different models compared to RMSprop and SGD.
- The influence of optimizer choice on model performance indicated significant variances, with certain architectures responding better to specific optimizers.

Activation Functions Performance

- ReLU activation function often achieved higher validation accuracies across different models compared to Sigmoid and Tanh.
- The choice of activation function significantly influenced model performance, with different architectures demonstrating better compatibility with specific functions.
- Tanh functions are similar to sigmoid functions but have a range [-1, 1], making them zero-centered.

Pooling Techniques Performance

- Max Pooling generally offered better performance in feature detection, leading to higher accuracy in more complex CNN architectures.
- The selection between Max Pooling and Average Pooling showed considerable impact on model behavior, with some architectures benefiting from the gentle characteristic of Average Pooling for information preservation.

Conclusions

The study clearly illustrated that simpler CNN architectures could outperform more complex setups depending on the dataset and model tuning specifics like the use of dropout or data augmentation. The optimal choice of optimizer also varied significantly between different network architectures, suggesting a need for careful selection based on the specific requirements and behavior of each model type.

The achievements highlight the effectiveness of targeted data augmentation and parameter tuning in improving the accuracy of CNNs tailored to specific tasks like image classification on the Fashion MNIST dataset. This serves as a pivotal insight for further research and application in related fields.