Comparative Analysis of Neural Network Architectures on Fashion-MNIST

NAME: KARTHIK MADDINENI

P NO: P2889143

1. Introduction and Experimental Setup

The present manuscript provides an account of an experiment that aimed at comparing various neural network architectures for image classification using the Fashion-MNIST dataset. The ultimate purpose was to assess the performance of both a basic Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN), moreover, to see how much the performance could be improved by the advanced techniques like Batch Normalization and learning rate scheduling.

The input data consisted of 70,000 grayscale images (training 60,000, testing 10,000) divided into the 10 fashion categories. The input data underwent preprocessing by first converting into PyTorch tensors and subsequently normalizing. The training set was then split into training and validation subsets in the ratio of 80% and 20%. Creation of DataLoader objects allowed efficient batch processing during the model's training and evaluation stages. Moreover, the experiment was set up to utilize a GPU if one was available, thus, securing the best possible computational power.

2. Model Architectures: ANN vs. CNN

Two base model architectures were created and put into practice:

Simple Artificial Neural Network (ANN): This model was a straightforward implementation of a Multi-Layer Perceptron (MLP). The 28x28 input image was transformed into a 1D vector and subsequently processed through a stack of fully connected (linear) layers, interspersed with ReLU activation functions. This setup represents the image as a flat collection of pixels and neglects their spatial relationships.

Simple Convolutional Neural Network (CNN): This model was primarily focused on the 2D aspect of the images. It consisted of convolutional layers that were able to detect spatial features like edges and textures, and then non-linearities and max-pooling layers for subsampling were applied. The features obtained in this way were then flattened and sent to a last fully connected layer for classification. CNNs are naturally more advantageous when it comes to image data as the sharing of parameters and spatial hierarchy are the main reasons behind this.

3. Initial Training, Results, and Performance Analysis

Both models were trained utilizing the Cross-Entropy loss function alongside the Adam optimizer. The performance of the models was subsequently measured using the test set which was kept aside initially.

Results: The CNN model was a clear winner over the simple ANN.

ANN Test Accuracy: 88.58%

CNN Test Accuracy: 90.34%

Confusion together with Matrix evaluation: Each model was evaluated for performance class wise by generating the confusion matrix for each model. The analysis showed that both models, especially the ANN, had difficulty telling apart the classes that were visually similar. For instance, the CNN even though it had higher overall accuracy, it was very much confused between 'Shirt', 'Coat' and 'Pullover'. This suffered by the dataset which required fine distinctions to be made.

4. Enhancement with Batch Normalization and Learning Rate Scheduling

The initial models were improved with by the introduction of two enhancements:

Batch Normalization (BN): New model classes, MLP_BN and CNN_BN, were created by integrating Batch Normalization layers (BatchNorm1d for MLP, BatchNorm2d for CNN). By normalizing the inputs to each layer, BN reduces internal covariate shift and, as such, it stabilizes and speeds up the training process.

Learning Rate Scheduler: A StepLR scheduler was applied to the training process. The learning rate is decreased systematically through this scheduler after a certain number of epochs, so it will be the model that will be able to make very small weight adjustments during the constant nearing of the minimum.

By means of the new training function, the scheduler and the training loop for the improved models were taken care of.

5. Final Evaluation and Summary of Findings

The enhanced models (MLP_BN and CNN_BN) were trained and evaluated, yielding the following results:

MLP_BN Test Accuracy: 89.27%

CNN_BN Test Accuracy: 92.15%

The CNN architecture gave a huge benefit for the image classification task because it had the capability of naturally and correctly classifying images by keeping the spatial order of pixel data. In addition, the use of Batch Normalization and a learning Rate scheduler led to a considerable performance boost for both models, with the improved CNN reaching an impressive 92.15% test accuracy. Nevertheless, there were these improvements, and a problem of great concern continued to exist; confusion matrices indicated that all models had bad results in distinguishing very similar classes, such as 'Shirt,' 'Coat,' and 'Pullover,' they thus revealed the difficulty of fine-grained classification even with the top architectures and techniques.

In conclusion, the experiment corroborated the efficacy of CNNs and at the same time demonstrated the value of such training techniques as Batch Normalization and learning rate scheduling through the entire process of constructing and slowly improving neural network models for image classification.