Alexei Thorner, Adrien Capt, Flavien Valea

Report: Milestone 2, Machine Learning

INTRODUCTION

In this project, we aimed to train and evaluate various machine learning algorithms on the Fashion-MNIST dataset, which consists of images of fashion items categorized into ten different classes. The primary objective was to classify images of clothing items into their respective categories.

METHODOLOGY

Data Preparation

The dataset was standardized based on its mean and standard deviation to ensure consistent input for the models. Each experiment used 6/7 of the data for training, in this training data we used 1/5 to make a validation set to prevent fitting our hyper-parameters to the test data which consist of the 1/7 left. The test data is used to validate the models' performances at the end.

Multi-Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) has an input layer, hidden layers, and an output layer, using backpropagation for training to handle non-linear separable data.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) reduces dimensionality by transforming correlated variables into orthogonal principal components, prioritizing variance.

Convolutional Neural Network (CNN)

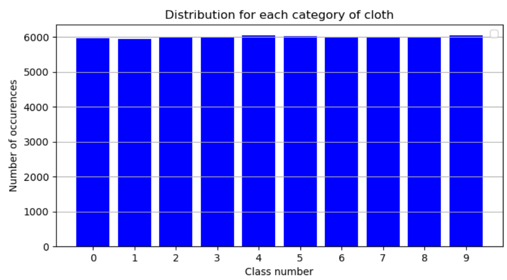
Convolutional Neural Networks (CNNs) analyze images through convolutions and extract the most information via the use of max pooling to reduce dimensionality. They work on the principle of spatiality and the final prediction uses fully connected layers just like MLP.

Transformers

Transformers are a model architecture using attention mechanisms to enhance training for tasks like language translation and text generation. Unlike RNNs, they handle long-range dependencies without sequential data processing, using self-attention in encoder and decoder blocks with multi-headed attention and feedforward layers.

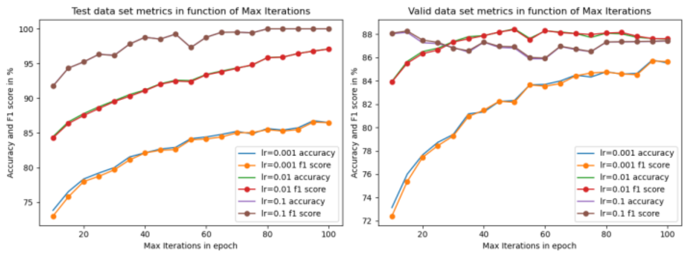
EXPERIMENT AND RESULTS

We trained our models to identify each picture. To determine the optimal hyperparameters, such as the learning rate and the number of epochs, we performed a grid search over a range of possible values for these parameters. The best combination was selected based on performance.

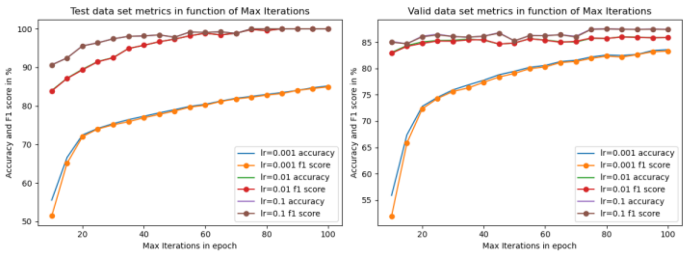


The dataset appears to be well-balanced, indicating no concerns about any class being underrepresented. No additional precautions are necessary; we can opt for a random 20% sample for the validation set, without fearing further unbalanced data.

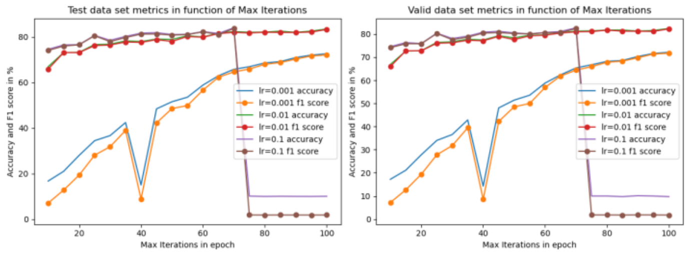
CNN



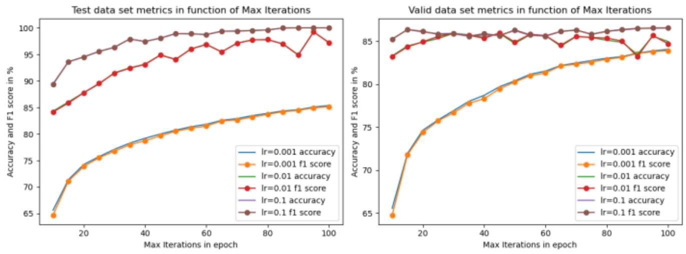
MLP



TRANSFORMER



MLP with PCA



We observe that if the model uses a higher learning rate, it will converge too quickly to the optimal solution, leading to overfitting after only a few iterations. Conversely, a smaller learning rate requires a greater number of iterations. Additionally, we note that after 75 epochs, the transformer's performance deteriorates significantly, with a substantial drop in accuracy and F1 score. For both CNN and MLP, learning rates of 1e-1 and 1e-2 cause rapid overfitting, resulting in poor validation set accuracy. While using PCA for the MLP does not significantly change its accuracy, it considerably reduces the number of parameters and the training time required.

To determine the hyperparameters within the models, such as the number of layers, activation functions, and the number of self-attention heads, we used the validation set and time needed to train/predict for selection. We tuned each parameter, trained the model, and retained any improvements. While cross-validation would provide a more robust solution, it significantly increases the training time.

We found that the best fitted model architecture for this kind of problem, i.e classification over an image, is CNN. It shows the most promising result based on time and number of parameters to train. The worst one was transformer. Even with less than 10 times the number of parameters of MLP, it still takes 15 times more time to train due to the overwhelming amount of computation a Transformer must perform. It is not a viable solution for this kind of problem.

CONCLUSION

In conclusion, for image classification, the best model is CNN. It outperforms the MLP and Transformer in terms of training and prediction speed, while also having fewer parameters than the MLP. Despite overfitting the training set, CNN achieved the highest performance, with an accuracy of 88.278% and an F1 score of 88.158%, making it the most effective model among the three evaluated.