Assignment 3: Classification of Image Data

April 2nd, 2024

1. Abstract

In this study, we explore different kinds of machine learning architectures employed in this task of classifying images of sign language. A Multilayer Perceptron (MLP) model was initially implemented to demonstrate basic neural network architecture and training via backpropagation. Further experiments indicate the benefits of network depth and nonlinear activations for performance. The role of L2 regularization in preventing overfitting has shown to provide best results for a given strength of regularization. The analysis further shows that ConvNets yield better results than MLPs and thus recommend their usage in most of the tasks related to images.

2. Introduction

Sign language serves as a fundamental communication bridge for individuals with hearing and speaking impairments, facilitating their interaction with the broader community (Toth, 2009). This project includes the implementation and further evaluation of several machine learning models on classification tasks with regard to image data, in particular, the MNIST dataset for sign language. An implementation of a Multilayer Perceptron (MLP) model is shown to demonstrate the basic aspects of the neural network architecture, such as the construction of layers, activation functions, and the training algorithm (the backpropagation). This leads to an exploration of different model configurations and their impact on performance. Firstly, we tested three neural network models on the Sign Language MNIST dataset: one with no hidden layers, one with one hidden layer, and one with two hidden layers, with varying numbers of hidden units (32, 64, 128, 256). We discovered that the depth of the network and the number of hidden units significantly impact model performance. Secondly, nonlinearity introduced by activation functions significantly impacts model performance. In the study, a neural network model with two hidden layers was tested using three activation functions: sigmoid, Leaky-ReLU, and ReLU. The ReLU and Leaky-ReLU activations resulted in higher accuracy (0.74) compared to the sigmoid (0.72) . Thirdly, we also learned the effect of L2 regularization on model performance; A two-layer neural network was tested with different levels of L2 regularization to prevent overfitting. The best accuracy was achieved with a mid-level regularization strength (λ=0.1). Fourthly, a Convolutional Neural Network was trained on a dataset with varying hidden units in the fully connected layers, achieving the highest accuracy with 128 units. Lastly, we compared MLPs with ConvNets and found that ConvNets were better than MLPs in an image classification task, which shows the superiority of ConvNets in spatial data.

3. Datasets

The dataset we used comes from the Modified National Institute of Standards and Technology (MNIST). It is a collection of pictures representing 24 out of 26 letters of the American Sign Language. The letters "J" and "Z" have been left out because they require movement of the hands. There are 27455 training samples and 7172 testing examples. The resolution of the pictures is 28x28. During pre-processing, pictures are flatten to an array of 784 pixels. The pictures are grayscale and have a value between 0-255. The letters with the least training examples are the letter "E" with 957 instances and the letter with the most instances is "R" with 1294 instances. During our pre-processing, we standardized our input features such that the mean of the grayscale value of the pixels was zero and the standard deviation one.

4. Results

4.1 Creation of a MLP with 0, 1, and 2 hidden layers

In this experiment, we formulated twelve distinct Multilayer Perceptron (MLP) architectures. We systematically generated each possible combination of models, encompassing configurations with 0, 1, and 2 hidden layers, alongside hidden unit counts of 32, 64, 128, and 256. As can be seen by figure x, The models converge a lot quicker when using 2 hidden layers and 256 hidden units then when using 0 hidden layers and 32 hidden units. These results generalize and we can see from our collab document that the more hidden layers and hidden units there are, the faster the model converges. This is unsurprising since the added layers and hidden units increase the flexibility of the model.

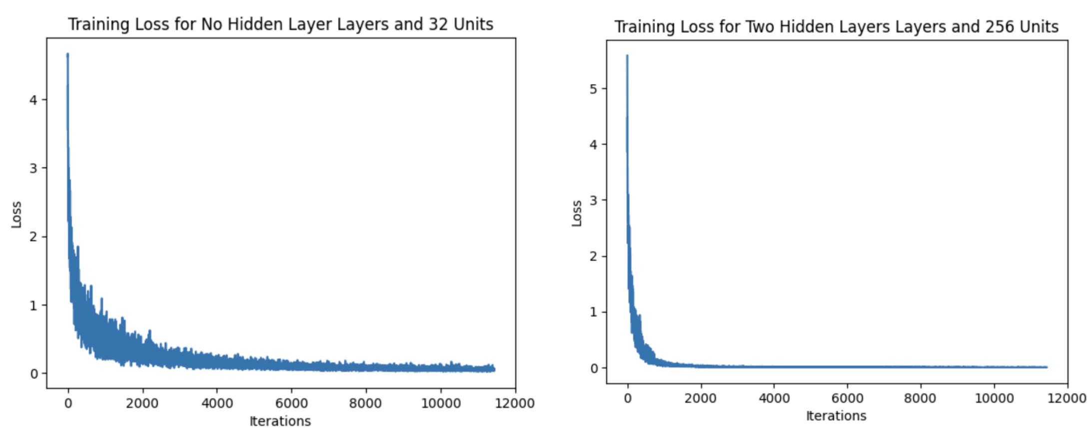

Figure x: Training loss given the number of iterations for the MLP with no hidden layer and 32 units (left) and two hidden layers and 256 units

However, this increase in flexibility can also create some overfitting, especially if there isn't enough training data. We can observe this happening in figure x, where the model with 1 hidden layer performs better than the one with 2 hidden layers when using 256 hidden units. When using 256 hidden units, the accuracy is 68%, 78%, and 77% for the MLP using no hidden layers, 1 hidden layer, and 2 hidden layers respectively.

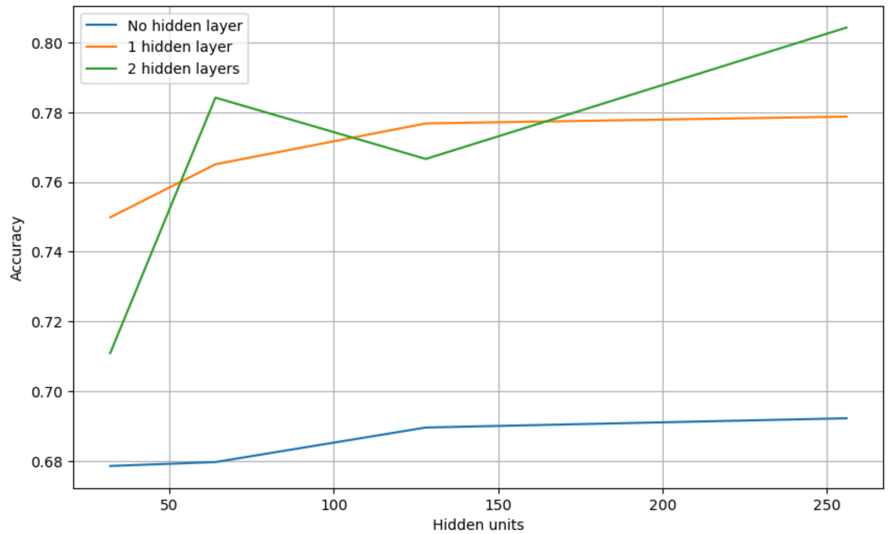

Figure x: Accuracy of the MLP with 0, 1, and 2 hidden layers given a different number of hidden units.

4.2 Creation of a 2-hidden layer MLP using Leaky-ReLu, ReLu and Sigmoid

As can be seen in figure x, the activation function that performs the best is ReLu. However, the accuracy between Leaky-ReLu and ReLu, is very similar at 74% and 76% respectively. This is slightly surprising as we would have expected Leaky-ReLu to perform better as it passes negative values to the next layer.

On the other hand, the sigmoid function is worse than the two others. It is possible that the sigmoid activation function performs slightly worse because of the vanishing gradient problem. This problem happens with the sigmoid function because it is a saturating function.

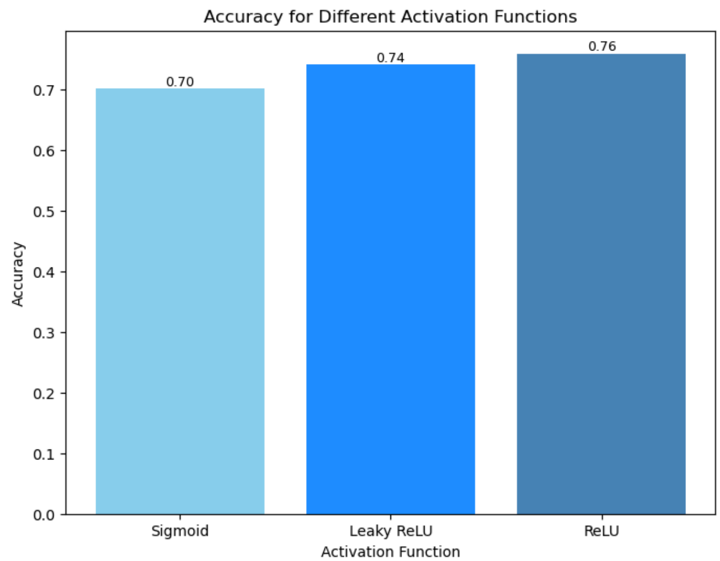

Figure x: Accuracy for Different Activation Functions

4.3 Incorporation of L2 regularization into a neural network model with two hidden layers

We modify the two-hidden-layer neural network model to incorporate L2 regularization. L2 regularization helps prevent overfitting by adding a penalty to the loss function with respect to the magnitude of the weights. As shown in figure x, the regularization strength, lambda (λ), was varied across four different values: 0.001, 0.01, 0.1, and 1. The results, shown in the graph, demonstrate that a moderate amount of regularization (λ=0.1) yielded the highest accuracy. Too little regularization led to the low accuracy, which could be due to overfitting, while too much regularization (λ=1) also reduced accuracy significantly, likely by underfitting the model.

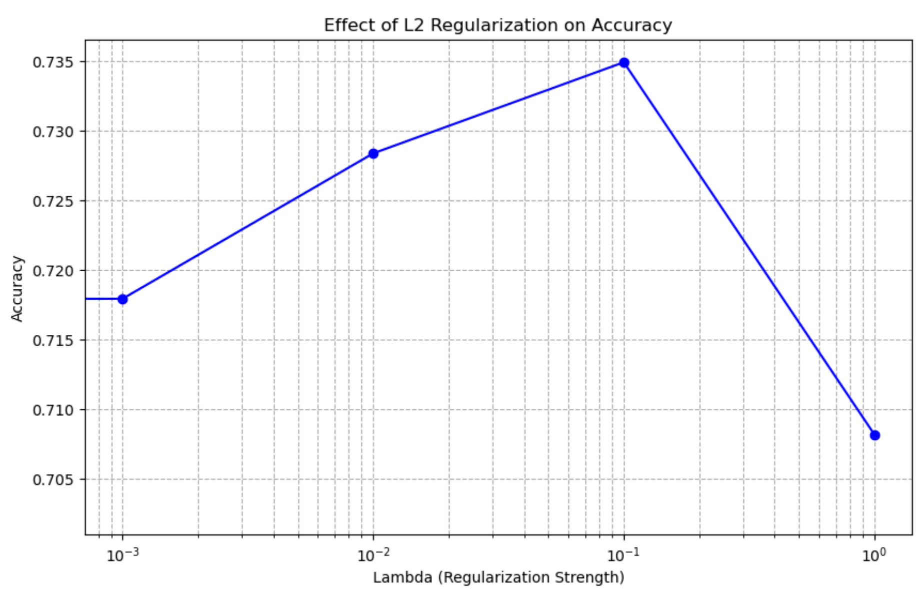

Figure x: Effect of L2 Regularization on Accuracy

4.4 constructing a Convolutional Neural Network

The task was to implement a Convolutional Neural Network (ConvNet) with 3 convolutional and 2 fully connected layers using the deep learning library. The number of hidden units for the fully connected layers tested was 32, 64, 128, and 256. All the activation functions applied ReLU for every one of the layers during training, and the same dataset was used for training. The bar graph (figure x) compares the test accuracies when the number of hidden units was varied in the fully connected layers of ConvEnt. Among the test accuracies plotted for the three best cases, it scored most for 128 hidden units (0.981), followed by 64 and 256, and least for 32 hidden units.

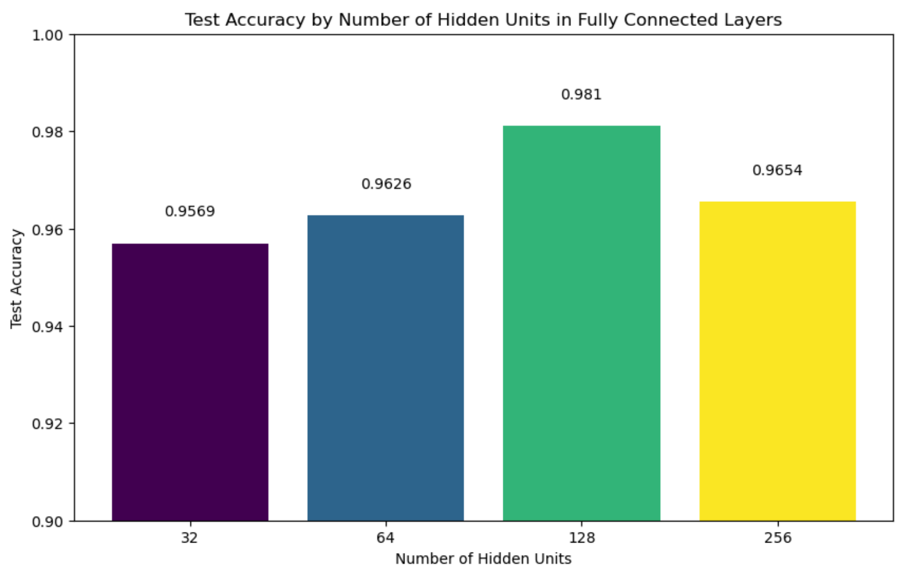

Figure x: Test Accuracy by Number of Hidden Units in Fully Connected Layers

4.5 Comparison between optimized Multi-Layer Perceptron and CNN

An improved Multi-Layer Perceptron (MLP) architecture was compared to the one mentioned earlier Convolutional Neural Network (ConvNet).The ConvNet achieved a higher test accuracy of 0.9650, whereas the MLP reached 0.8561. The results in this case clearly evidence that, despite attempts at tuning architectures every which way to bring a better improvement to the performance of the MLP model, the ConvNet was able to outperform the MLP. This is likely due to its specialized ability to process spatial features in image data which is crucial for tasks like the Sign Language MNIST dataset (Mei et al., 2017).

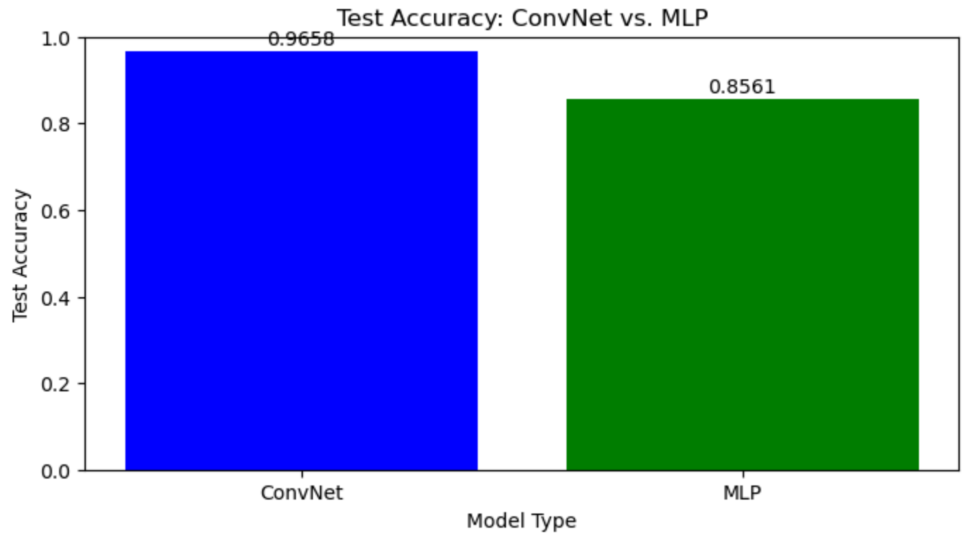

Figure x: Test Accuracy: ConvNet vs. MLP

5. Discussion and Conclusion

From this research, we have some key findings in the area of classifying sign language images using MLP. They lie in the effect of network depth and non-linear activations, such as ReLU, leaky-RELU, which contribute highly to the improvement in model performances. Also, L2 regularization is crucial for preventing overfitting and should be applied with a moderate strength of λ value. Notably, Convolutional Neural Networks (ConvNets) outperformed Multi-Layer Perceptrons (MLPs), highlighting ConvNets' superior capability in processing spatial data essential for image classification. This groundwork shows deep learning's potential in sign language classification, paving the way for advancements in assistive technologies. Based on these findings, future studies can be performed from directions including searching for advanced neural network architectures and finding better data augmentation techniques that could help in better generalization of the model.

6. Statement of Contributions

a. Guillaume Delmas-Frenette (260982554): MLP model implementation, pre-processing, experiment 1- 2, report: datasets, experiment 1- 2.
b. Maxx Railton(260966381): MLP model implementation, pre-processing, experiment 1-5
c. Hanzi Li (261008394): Work on the Abstract, Introduction, Results (4.3-3.5), and Discussion and Conclusion part of the report.

References

Mei, S., Ji, J., Hou, J., Li, X., & Du, Q. (2017). Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(8), 4520–4533. https://doi.org/10.1109/TGRS.2017.2693346

Toth, A. (2009). Bridge of Signs: Can Sign Language Empower Non-Deaf Children to Triumph over Their Communication Disabilities? *American Annals of the Deaf*, *154*(2), 85–95. https://www.jstor.org/stable/26234583?read-now=1