Methods in Artificial Intelligence Homework 6

Christoffer Brevik

February 28, 2025

1 Introduction

This report documents the results obtained from training a fully connected neural network (MLP) and a convolutional neural network (CNN) on the MNIST dataset. The primary task was to modify training parameters to achieve the required test accuracy.

2 Model adjustments

To solve the task I did not require to change the layer size or count, or any other major architectural changes. The only necessary modifications were changing the number of epochs from 1 to 50 and reducing the learning rate from 2.0 to 0.05. While increasing model size would mean better results, it would also result in long training times.

3 Final Results

The final test accuracy achieved for each model is as follows:

- MLP: **0.9488** (94.88%)
- CNN: **0.9664** (96.64%)

These results meet and exceed the required benchmarks of 90% for MLP and 96% for CNN.

4 Comparison of MLP and CNN Performance

From my results, we see that the CNN outperformed the MLP, achieving a higher accuracy. This can be attributed to:

- The convolutional layers extract spatial features effectively, preserving local patterns in the image data.
- CNNs use parameter sharing through convolutional filters, reducing the number of parameters while capturing essential features.
- The total number of trainable parameters in the CNN is generally lower than a fully connected MLP with similar expressive power, leading to better generalization.

In contrast, the MLP treats each pixel independently, which may not capture spatial hierarchies as efficiently as a CNN.

5 Categorical Cross Entropy and Softmax Output

Both models use a softmax output layer with 10 units, corresponding to the 10 digit classes in MNIST. The categorical cross-entropy loss function is used to train the model by computing the difference between predicted probabilities and true labels. Mathematically, the loss function is defined as:

$$L = -\sum_{i=1}^{10} y_i \log(\hat{y}_i) \tag{1}$$

where y_i is the true label (one-hot encoded) and $\hat{y}_i$ is the predicted probability from the softmax function:

$$\hat{y}i = \frac{e^{z_i}}{\sum j = 1^{10} e^{z_j}} \tag{2}$$

This formulation encourages the model to assign high probability to the correct class while pushing down the probabilities for incorrect classes, thus enabling classification learning.

6 Conclusion

By adjusting only the number of epochs and learning rate, both the MLP and CNN achieved the required accuracy. The CNN performed better due to its ability to capture spatial relationships with fewer trainable parameters. The use of categorical cross-entropy with softmax effectively trained both models to classify MNIST digits