

SVM vs Softmax for Classification

Abrhame Habtom

August 2025

Abstract

This document presents a comparison between a Softmax (CNN + Softmax) and the SVM method for multiclass classification.

1 Introduction

Convolutional Neural Networks (CNNs) are commonly used for feature extraction in image classification tasks. The final layer of a CNN can be followed by a **Softmax** output layer or an **SVM** classifier. This paper compares the outputs and performance of both approaches.

2 Methodology

2.1 Softmax

The CNN model is trained on the MNIST dataset for handwritten digit classification. For the Softmax classifier, the final fully-connected layer produces raw scores z_i for each class $i \in \{1, \dots, C\}$. These scores are converted into probabilities using the Softmax function:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}$$

The predicted probabilities \hat{y}_i and the true labels y_i are used to compute the cross-entropy loss:

$$L_{CE} = - \sum_{i=1}^C y_i \log(\hat{y}_i)$$

Here, C denotes the number of classes, y_i is the ground truth label for class i ($y_i \in \{0, 1\}$), and \hat{y}_i is the predicted probability of class i .

The cross-entropy loss inherently incorporates the Softmax operation before computing the log-loss, effectively transforming logits into a probability distribution. The model is optimized using the Adam optimizer with a learning rate of 0.01 and trained for 3 epochs.

2.2 SVM

The Support Vector Classifier (SVC) model from `scikit-learn` is used for multi-class classification. After feature extraction using a similar CNN model as in the Softmax classifier, the raw feature vectors are vertically stacked to create a single training feature matrix, and the labels are horizontally stacked to create the corresponding label vector. These vectors are passed to the SVC model for multi-class classification using linear kernels.

Originally, the Support Vector Machine (SVM) was designed for binary classification, but the SVC extends this to multi-class classification by training:

$$\frac{K(K-1)}{2}$$

binary classifiers, where K is the number of classes to classify.

3 Results

3.1 Accuracy Comparison

Classifier	Accuracy (%)	Training Time (s)
CNN + Softmax	98.95	294
CNN + SVM (Linear)	99.03	7.13

Table 1: Performance comparison on MNIST test set.

3.2 Visualization

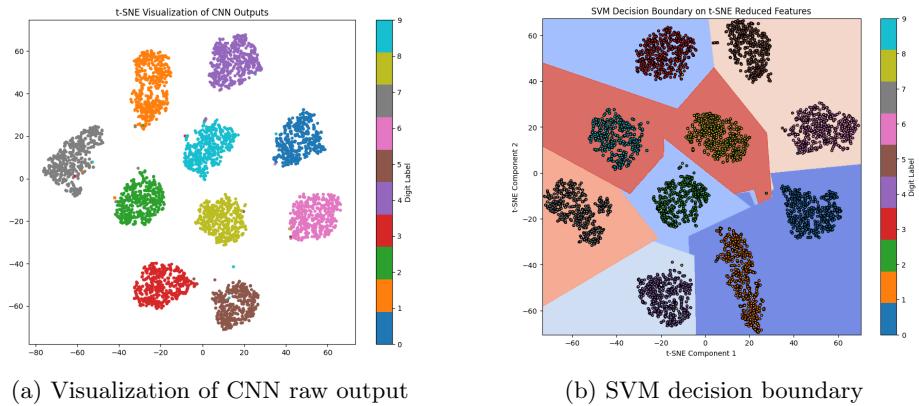


Figure 1: Comparison of CNN and SVM outputs