Generative Digit Classification and Inpainting

Using Gaussian Mixture Models

Tarun Ghutey
Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur

October 2025

Abstract

This project presents a generative framework for handwritten digit classification and image inpainting on the MNIST dataset. Each digit class is modeled using Gaussian Mixture Models (GMMs), with parameters learned via the Expectation-Maximization (EM) algorithm. The approach enables both robust multi-class classification and principled reconstruction of partially observed images through probabilistic inference.

1 Introduction

Discriminative models focus on learning decision boundaries between classes, whereas generative models explicitly learn the underlying data distribution. In this work, each handwritten digit is represented as a class-conditional probability distribution modeled by a Gaussian Mixture Model (GMM). This formulation allows the model to capture diverse writing styles and enables inference in scenarios with missing or corrupted input data.

2 Methodology

2.1 Gaussian Mixture Modeling

For each digit class $c \in \{0, 1, \ldots, 9\}$, the class-conditional distribution is defined as:

$$P(x \mid y = c) = \sum_{k=1}^{K} \pi_{c,k} \, \mathcal{N}(x; \mu_{c,k}, \Sigma_{c,k}), \tag{1}$$

where $\pi_{c,k}$ are the mixture weights satisfying $\sum_k \pi_{c,k} = 1$, and $\mu_{c,k}$ and $\Sigma_{c,k}$ denote the mean and covariance of the k-th component.

2.2 Parameter Estimation via EM

The parameters of the GMM are estimated using the Expectation-Maximization (EM) algorithm:

- **E-Step:** Compute the posterior responsibility $\gamma_{n,c,k}$ of each mixture component for a given sample.
- **M-Step:** Update the mixture weights, means, and covariances using the computed responsibilities.

2.3 Handling Censored Data and Inpainting

For partially observed images, classification is performed by marginalizing over the missing dimensions. Image reconstruction (inpainting) is achieved using conditional Gaussian inference:

$$\hat{x}_{\text{missing}} = \mu_m + \Sigma_{mo}\Sigma_{oo}^{-1}(x_{\text{observed}} - \mu_o), \tag{2}$$

where subscripts o and m denote observed and missing pixel subsets, respectively. This enables principled reconstruction based on learned digit structure.

3 Results and Analysis

3.1 Classification Performance

- **Uncensored Data:** A peak accuracy of **89.17%** was achieved using $K = 10$ mixture components per class.
- **Censored Data:** With 50% of pixels missing, accuracy improved with increasing model complexity, reaching **79.8%** at $K = 20$.

3.2 Model Efficiency: Hybrid-K Strategy

Empirical analysis revealed that simpler digits (e.g., `1`) require fewer components ($K = 2$), while more complex digits (e.g., `2`, `7`) benefit from higher capacity ($K = 10$). A hybrid-K strategy achieved an overall accuracy of **88.43%** while significantly reducing the total number of model parameters.

4 Conclusion

This project demonstrates that Gaussian Mixture Models can effectively capture the structural diversity of handwritten digits. By leveraging Bayesian inference and the EM algorithm, the proposed system performs reliable classification and high-quality image inpainting even under severe data degradation. The results highlight the interpretability and robustness of generative modeling approaches for vision tasks.