EECE5644 Assignment 4 Report

Dawei Wang
NUID: 002842604

November 29, 2025

Question 1: SVM and MLP Classification

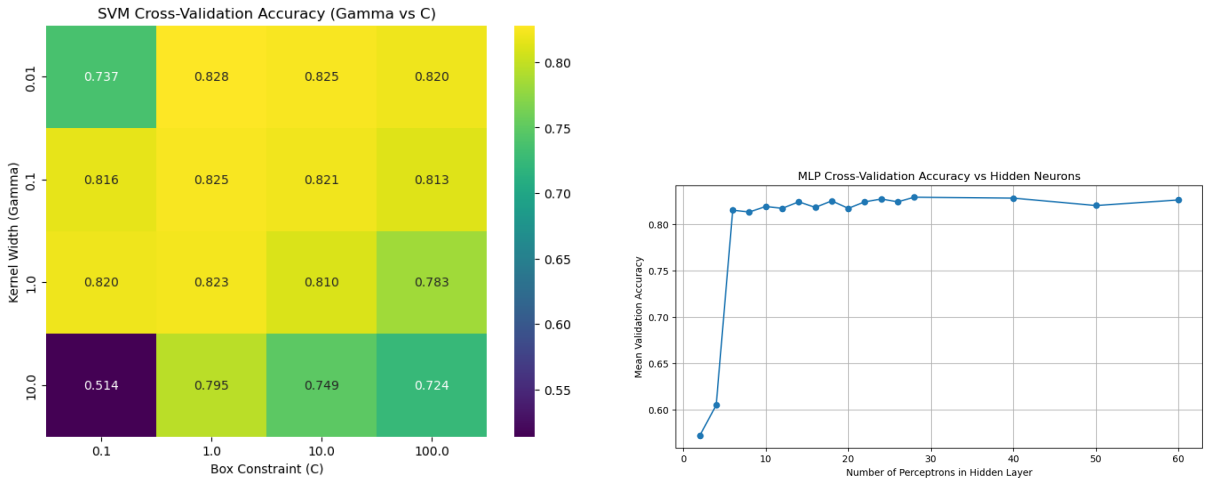
1.1 Data Generation

Samples Ire generated for class $l \in \{-1, +1\}$ using concentric distributions: $x = r_l[\cos(\theta), \sin(\theta)]^T + n$, where $r_{-1} = 2, r_{+1} = 4$, and noise $n \sim \mathcal{N}(0, I)$. Training set size $N_{train} = 1000$, testing set size $N_{test} = 10000$.

1.2 Hyperparameter Tuning (K-Fold Cross-Validation)

I utilized 5-fold cross-validation to select optimal hyperparameters.

- **SVM:** I searched for Box Constraint (C) and Kernel Width (γ) using a Gaussian RBF kernel.
- **MLP:** I optimized the number of perceptrons in a single hidden layer (ReLU activation).



(a) SVM Grid Search (Accuracy)

(b) MLP Neuron Selection (Accuracy)

Figure 1: Visual demonstration of the K-fold cross-validation process.

1.3 Test Performance and Decision Boundaries

The optimized models Ire evaluated on the test set. The decision boundaries approximate the theoretical circular boundary betIen the classes.

Table 1: Final Model Performance on Test Data ($N = 10000$)

Model	**Best Hyperparameters**	**Accuracy**	**Probability of Error**
SVM (RBF)	$C = 1, \gamma = 0.01$	82.80%	0.1712
MLP	Hidden Neurons = 28	82.90%	0.1685

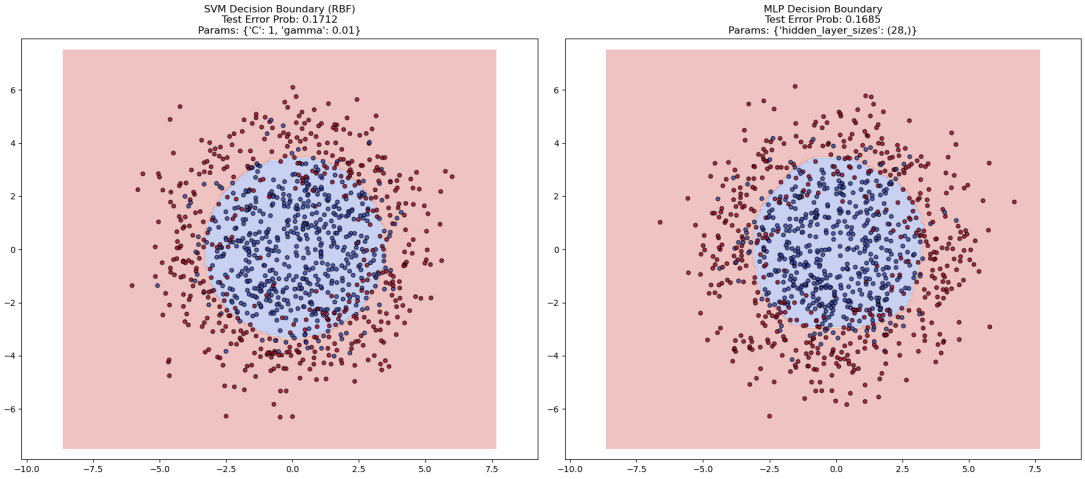


Figure 2: Classification boundaries and error estimation on test data.

Question 2: GMM-based Image Segmentation

2.1 Methodology

A color image was processed to generate 5-dimensional feature vectors $v = [r, c, R, G, B]^T$, normalized to $[0, 1]$. A Gaussian Mixture Model (GMM) was fitted to these vectors.

- **Model Selection:** I used K-fold cross-validation to maximize the average validation log-likelihood to determine the optimal number of components K.

2.2 Model Order Selection

Figure 3 illustrates the log-likelihood scores for different component counts.

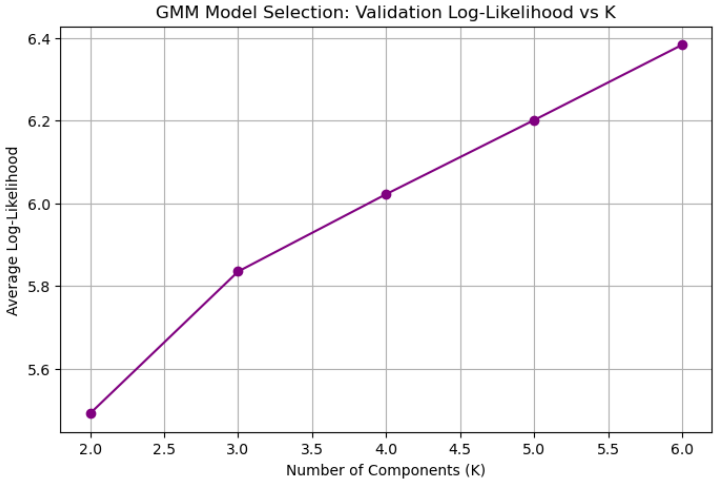


Figure 3: Average Validation Log-Likelihood vs. Number of Components (K). The peak indicates the optimal model order.

2.3 Segmentation Results

Pixels Ire assigned to the most likely component based on posterior probabilities. Figure 4 shows the segmentation outcome using the optimal K.

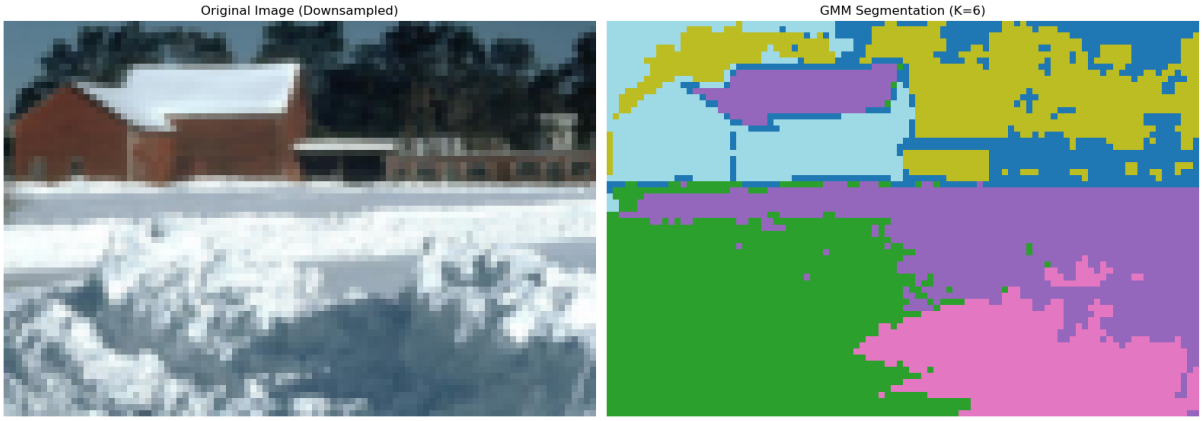


Figure 4: Original Image vs. GMM Segmentation.

Figure 5: GMM Segmentation Labels (K=6)

Appendix: Code

The code for this assignment is available at the following repository: `https://github.com/wang-dawei1/EECE5644/tree/main/Assignment4`