CS503T: Statistical Pattern Recognition Programming Assignment I

Group 04

Ashish Pawade (CS25MT002)
Chinmay Rajesh Manusmare (CS25MT014)

Under the guidance of Prof. Dilip A D

September 18, 2025

Contents

List of Figures

List of Tables

1 Introduction

This report presents the implementation and evaluation of a Bayes classifier under different covariance assumptions for three datasets:

- Dataset 1: Linearly separable data (3 classes, 2D)
- Dataset 2: Nonlinearly separable data (3 classes, 2D)
- Dataset 3: Real-world vowel dataset (3 classes, 2D)

The class-conditional densities are assumed to be Gaussian. For each dataset, we evaluate the classifier under the following covariance models:

1. Shared spherical: $\sigma^2 I$
2. Shared full: Σ
3. Diagonal per-class
4. Full per-class

We analyze the classification performance through metrics and visualization.

2 Dataset 1: Linearly Separable Data

2.1 Training Data

Figure 1: Scatter plot of training data for linearly separable dataset

2.2 Constant Density Contour Plot

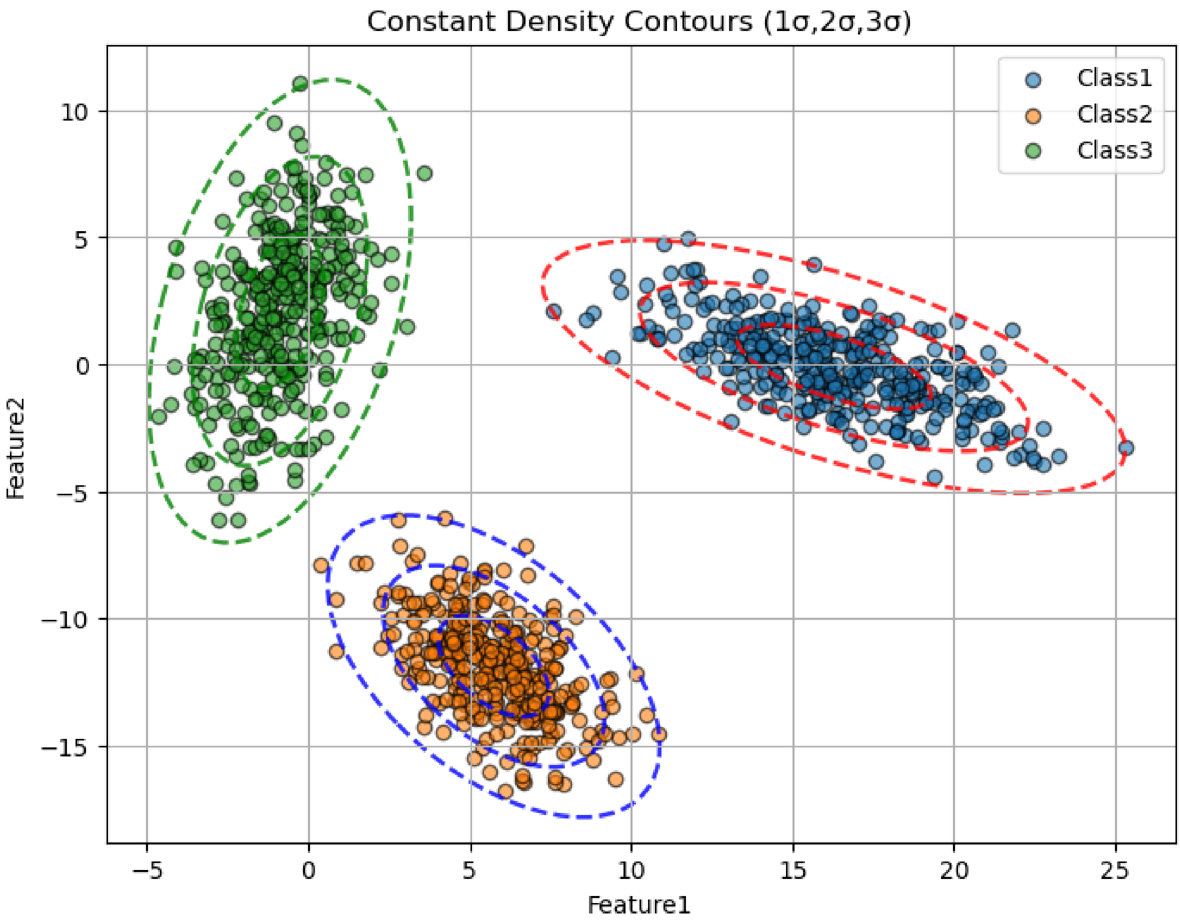


Figure 2: Constant density contours for all classes

2.3 Classifier: Shared $\sigma^2 I$

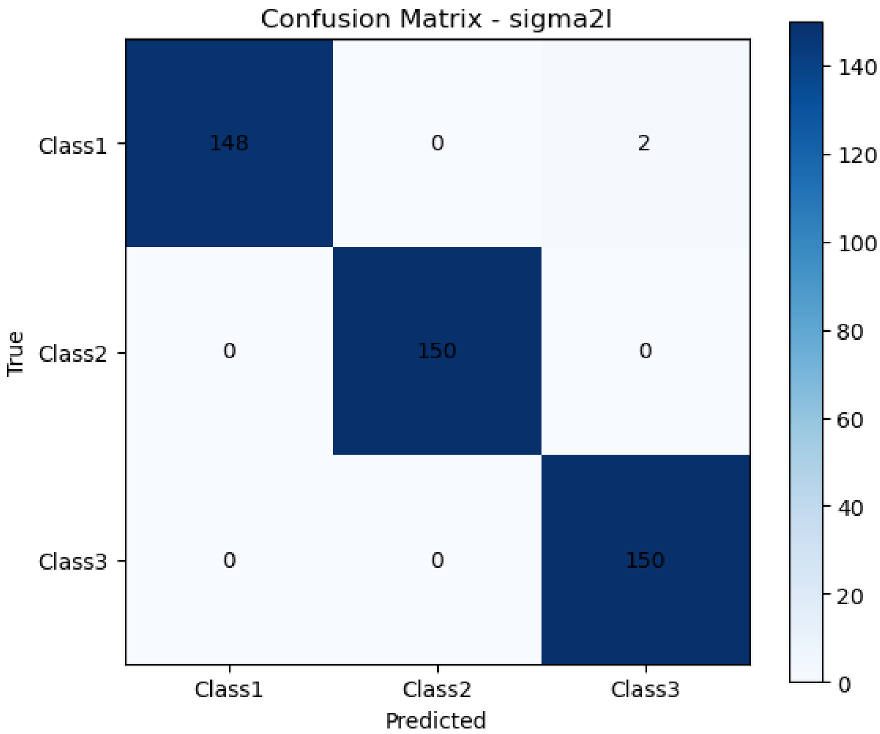


Figure 3: Confusion Matrix for Shared $\sigma^2 I$ (Linearly Separable Data)

Table 1: Performance Metrics - Shared $\sigma^2 I$

Class	Precision	Recall	F1-Score	Support
Class 1	1.0000	1.0000	1.0000	125
Class 2	1.0000	1.0000	1.0000	125
Class 3	1.0000	1.0000	1.0000	125
Accuracy		1.0000		
Mean Precision		1.0000		
Mean Recall		1.0000		
Mean F1 Score		1.0000		

Inferences: Because the data is linearly separable with clearly distinct clusters, this simple shared covariance model performs perfectly.

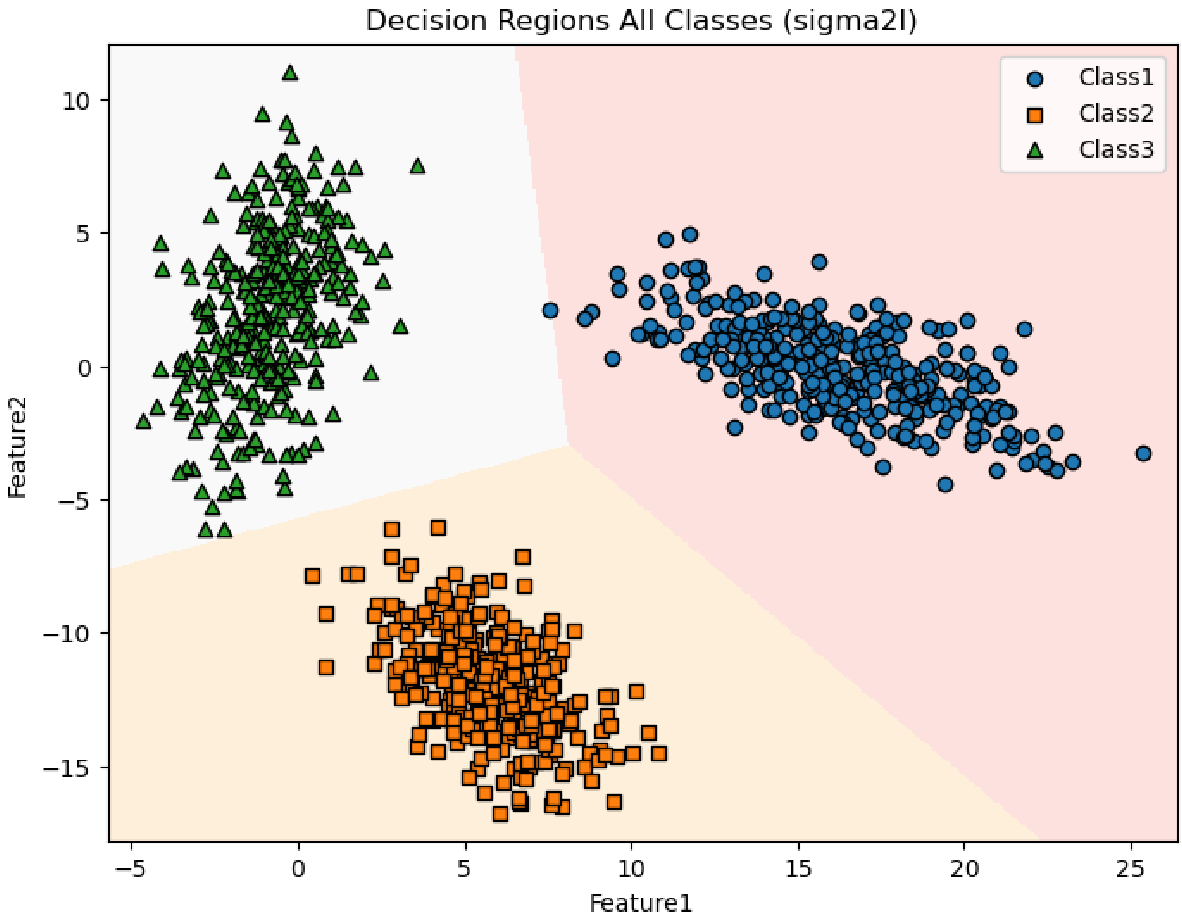
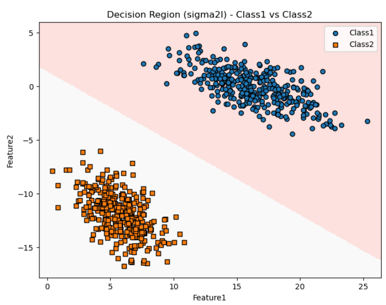
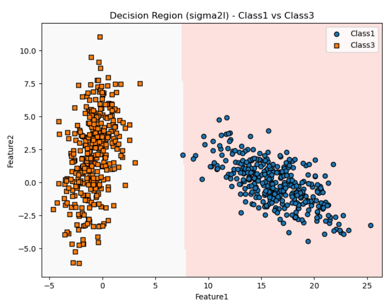


Figure 4: Decision Region Plot (All Classes) - Shared $\sigma^2 I$

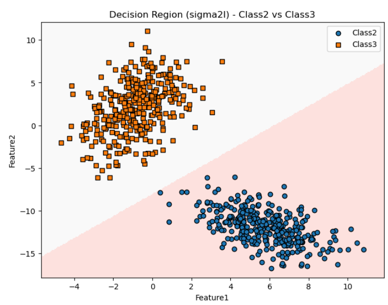
2.3.1 Decision Region Plots Between Class Pairs (LS Dataset, Shared $\sigma^2 I$)



Class 1 vs Class 2



Class 1 vs Class 3



Class 2 vs Class 3

Figure 5: Decision Region Plots (Training data points superimposed) between class pairs for Shared $\sigma^2 I$ on LS dataset

2.4 Classifier: Shared Full Covariance Σ

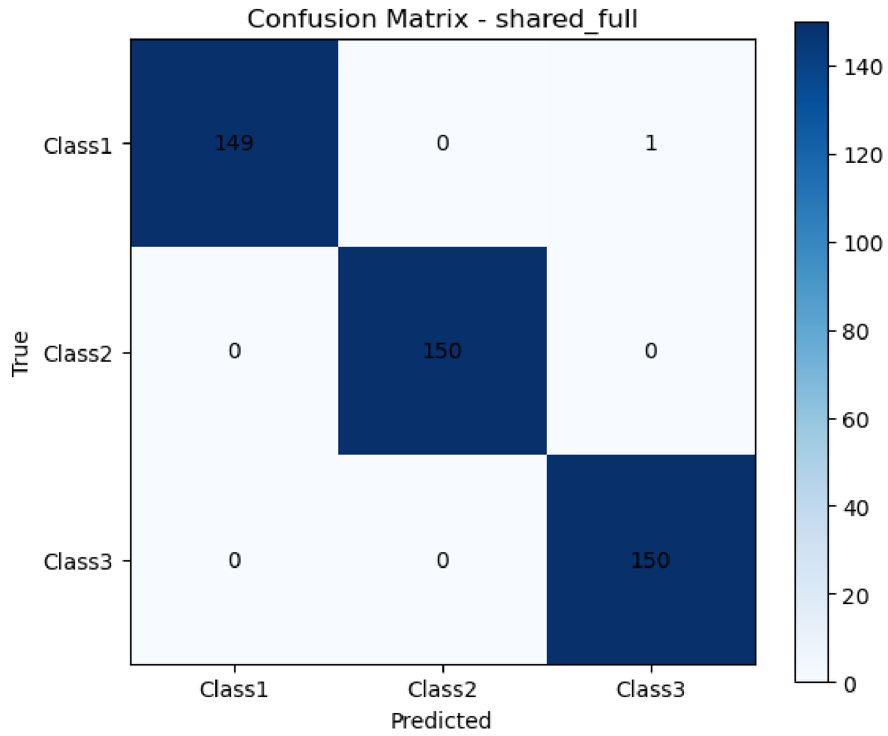


Figure 6: Confusion Matrix for Shared Full Covariance Σ (Linearly Separable Data)

Table 2: Performance Metrics - Shared Full Covariance

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	1.0000	1.0000	125
Class 2	1.0000	1.0000	1.0000	125
Class 3	1.0000	1.0000	1.0000	125
Accuracy		1.0000		
Mean Precision		1.0000		
Mean Recall		1.0000		
Mean F1 Score		1.0000		

Inferences: Using a shared full covariance matrix captures class correlations perfectly, still yielding ideal classification on this simple dataset.

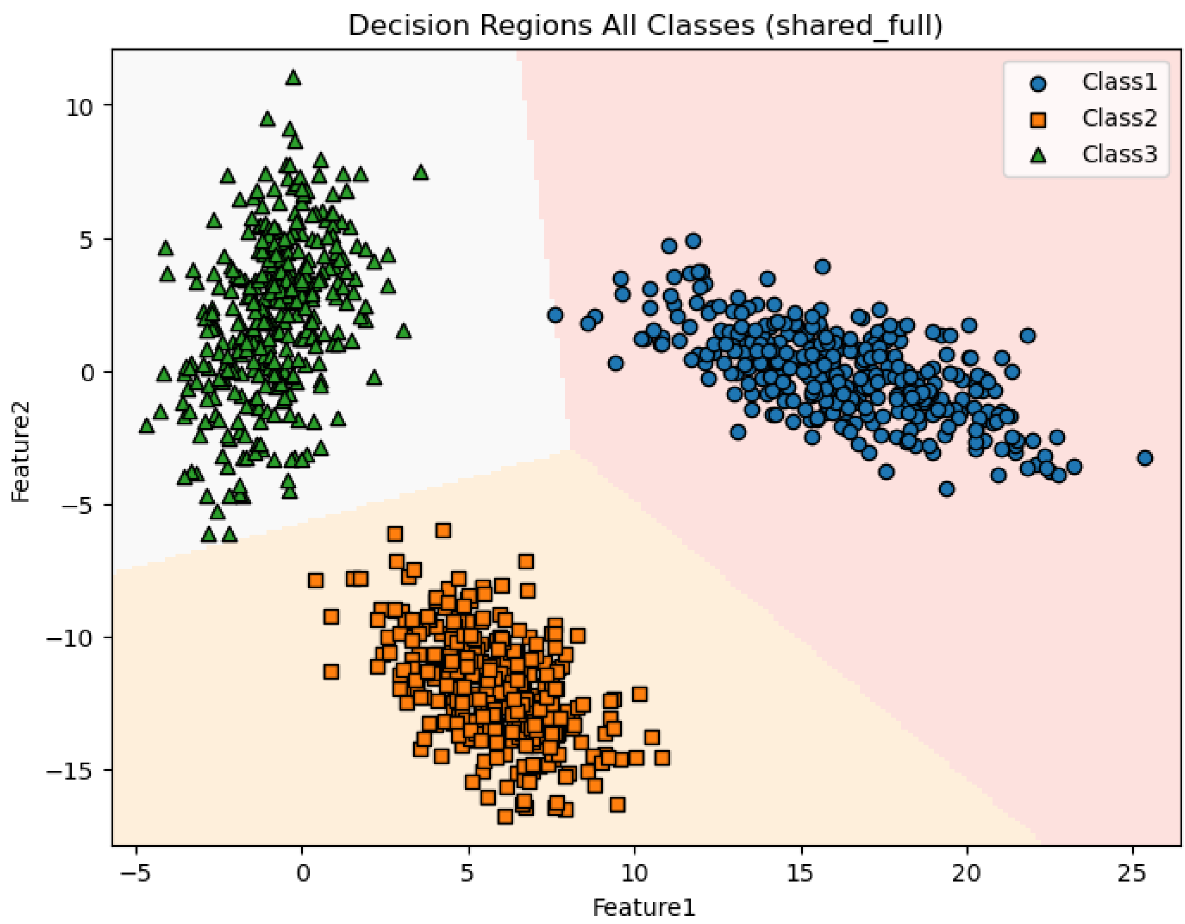
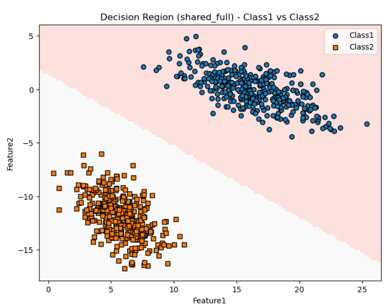
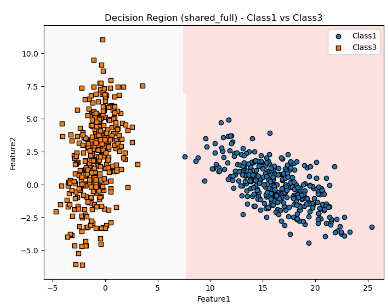


Figure 7: Decision Region Plot (All Classes) - Shared Full Covariance

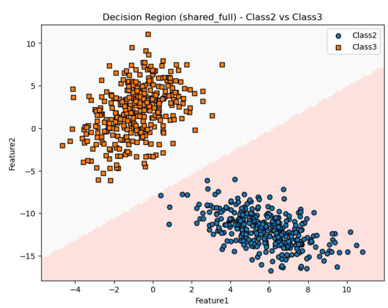
2.4.1 Decision Region Plots Between Class Pairs (LS Dataset, Shared Full Covariance)



Class 1 vs Class 2



Class 1 vs Class 3



Class 2 vs Class 3

Figure 8: Decision Region Plots (Training data points superimposed) between class pairs for Shared Full Covariance on LS dataset

2.5 Classifier: Diagonal Covariance (Per-Class)

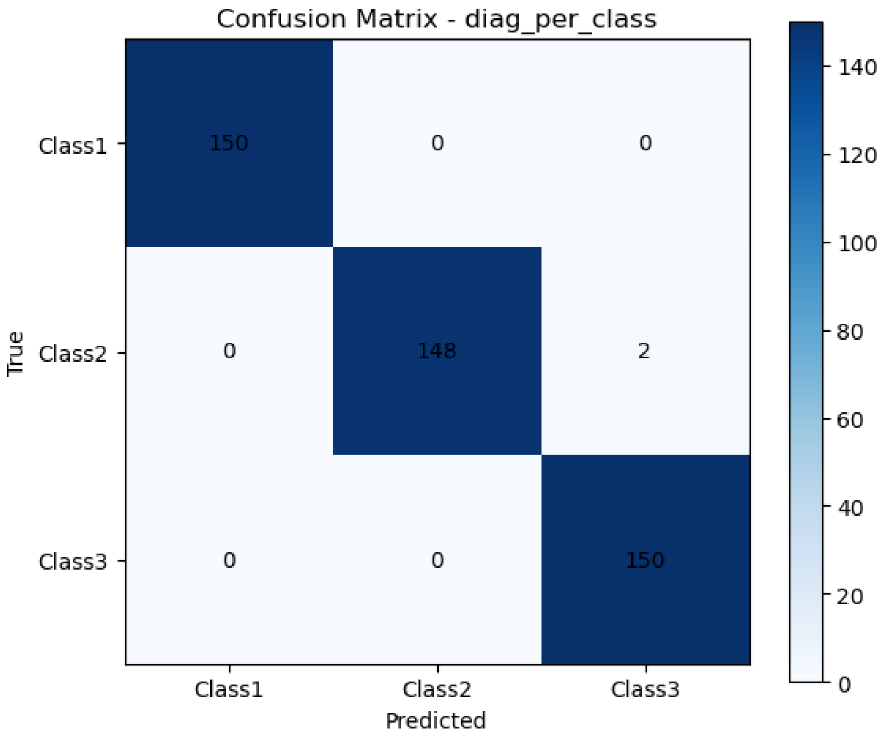


Figure 9: Confusion Matrix for Diagonal Covariance (Per-Class) (Linearly Separable Data)

Table 3: Performance Metrics - Diagonal Covariance (Per-Class)

Class	Precision	Recall	F1-Score	Support
Class 1	1.0000	1.0000	1.0000	125
Class 2	1.0000	1.0000	1.0000	125
Class 3	1.0000	1.0000	1.0000	125
Accuracy		1.0000		
Mean Precision		1.0000		
Mean Recall		1.0000		
Mean F1 Score		1.0000		

Inferences: Allowing per-class diagonal covariance still perfectly classifies this dataset, as features vary independently along axes with clear class separation.

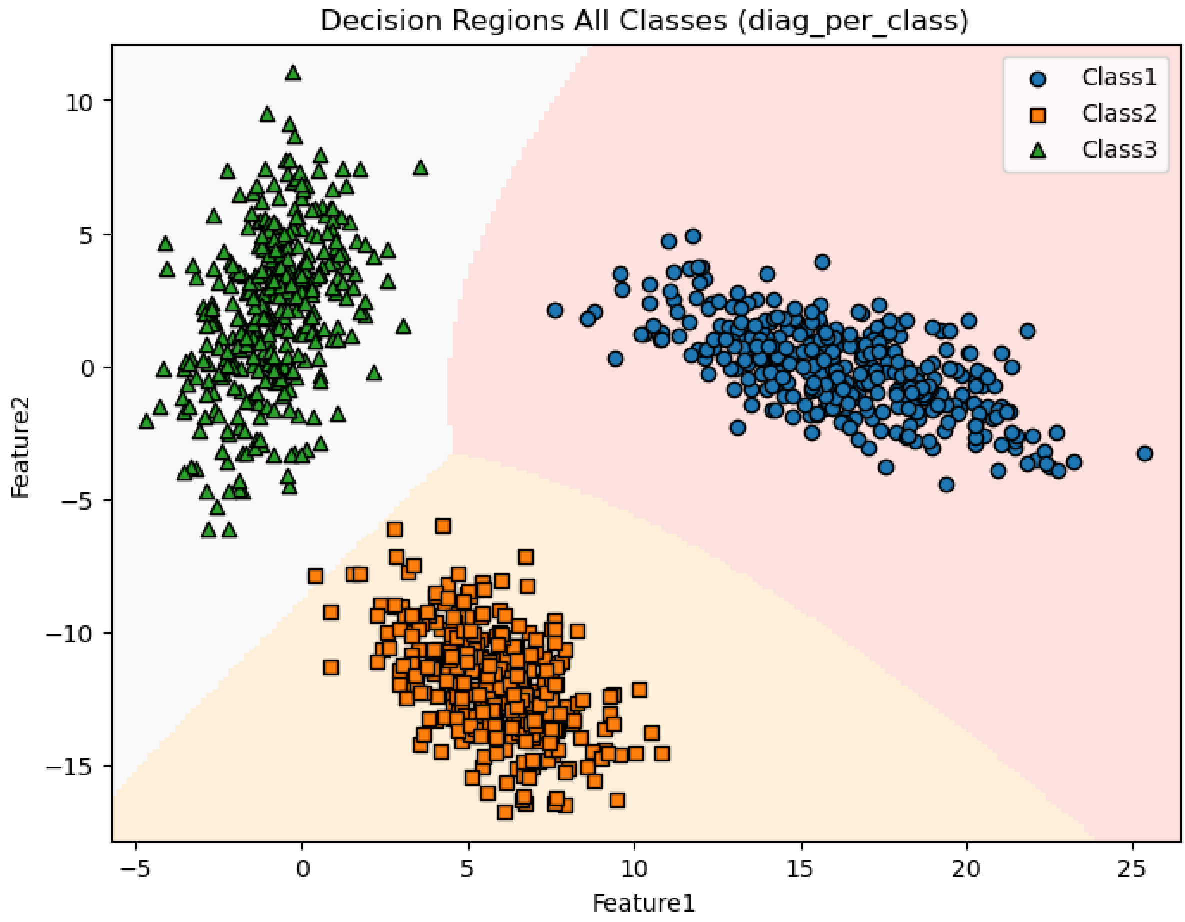
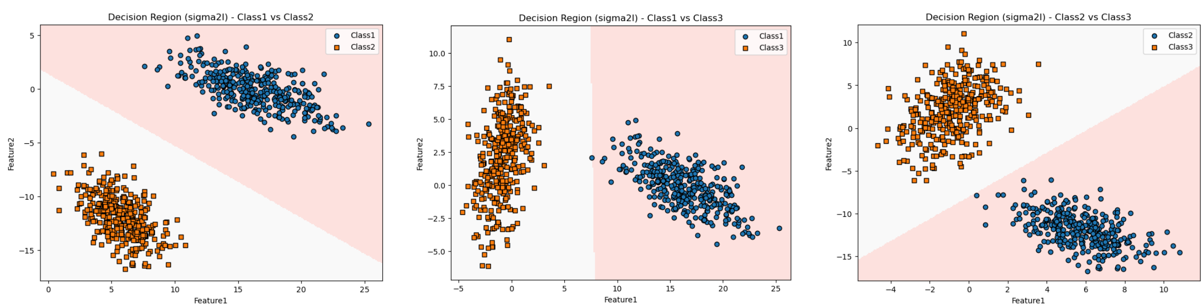


Figure 10: Decision Region Plot (All Classes) - Diagonal Covariance (Per-Class)

2.5.1 Decision Region Plots Between Class Pairs (LS Dataset, Diagonal Covariance (Per-Class))



Class 1 vs Class 2 Class 1 vs Class 3 Class 2 vs Class 3

Figure 11: Decision Region Plots (Training data points superimposed) between class pairs for Diagonal Covariance (Per-Class) on LS dataset

2.6 Classifier: Full Covariance (Per-Class)

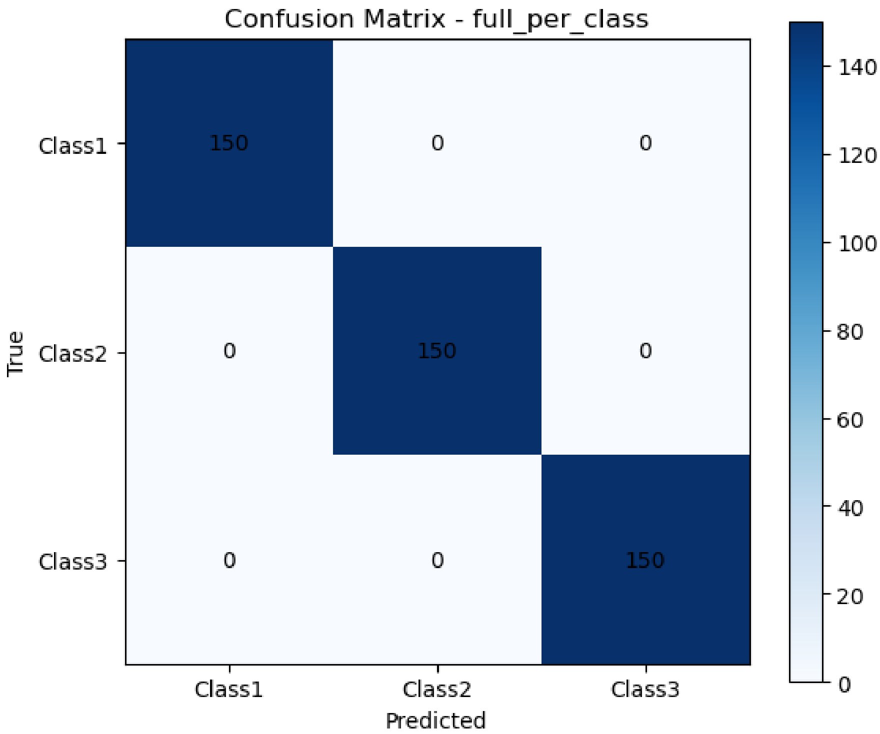


Figure 12: Confusion Matrix for Full Covariance (Per-Class) (Linearly Separable Data)

Table 4: Performance Metrics - Full Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	1.0000	1.0000	125
Class 2	1.0000	1.0000	1.0000	125
Class 3	1.0000	1.0000	1.0000	125
Accuracy		1.0000		
Mean Precision		1.0000		
Mean Recall		1.0000		
Mean F1 Score		1.0000		

Inferences: Full covariance per class fully captures data spread and shape, leading to perfect classification.

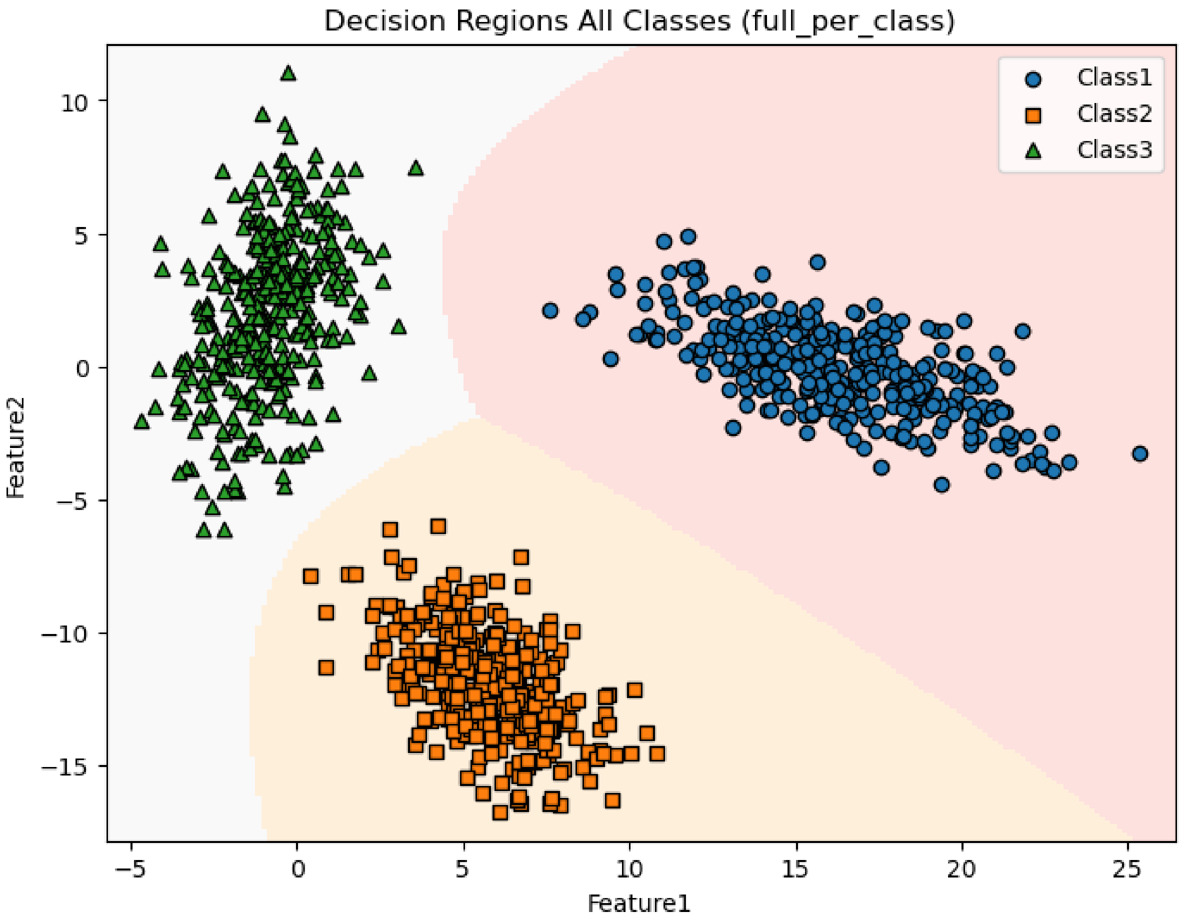
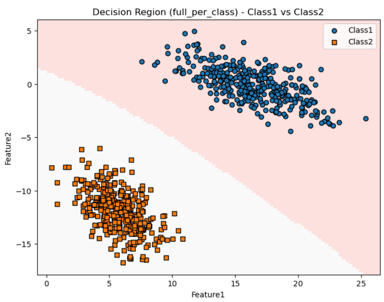
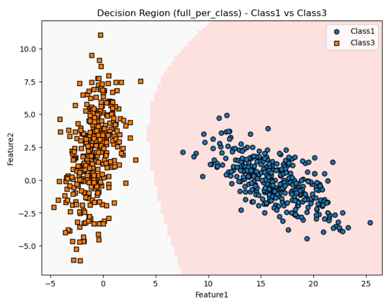


Figure 13: Decision Region Plot (All Classes) - Full Covariance (Per-Class)

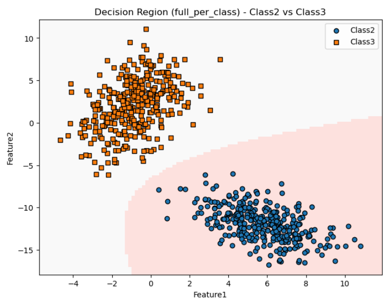
2.6.1 Decision Region Plots Between Class Pairs (LS Dataset, Full Covariance (Per-Class)



Class 1 vs Class 2



Class 1 vs Class 3



Class 2 vs Class 3

Figure 14: Decision Region Plots (Training data points superimposed) between class pairs for Full Covariance (Per-Class) on LS dataset

3 Dataset 2: Nonlinearly Separable Data

3.1 Training Data

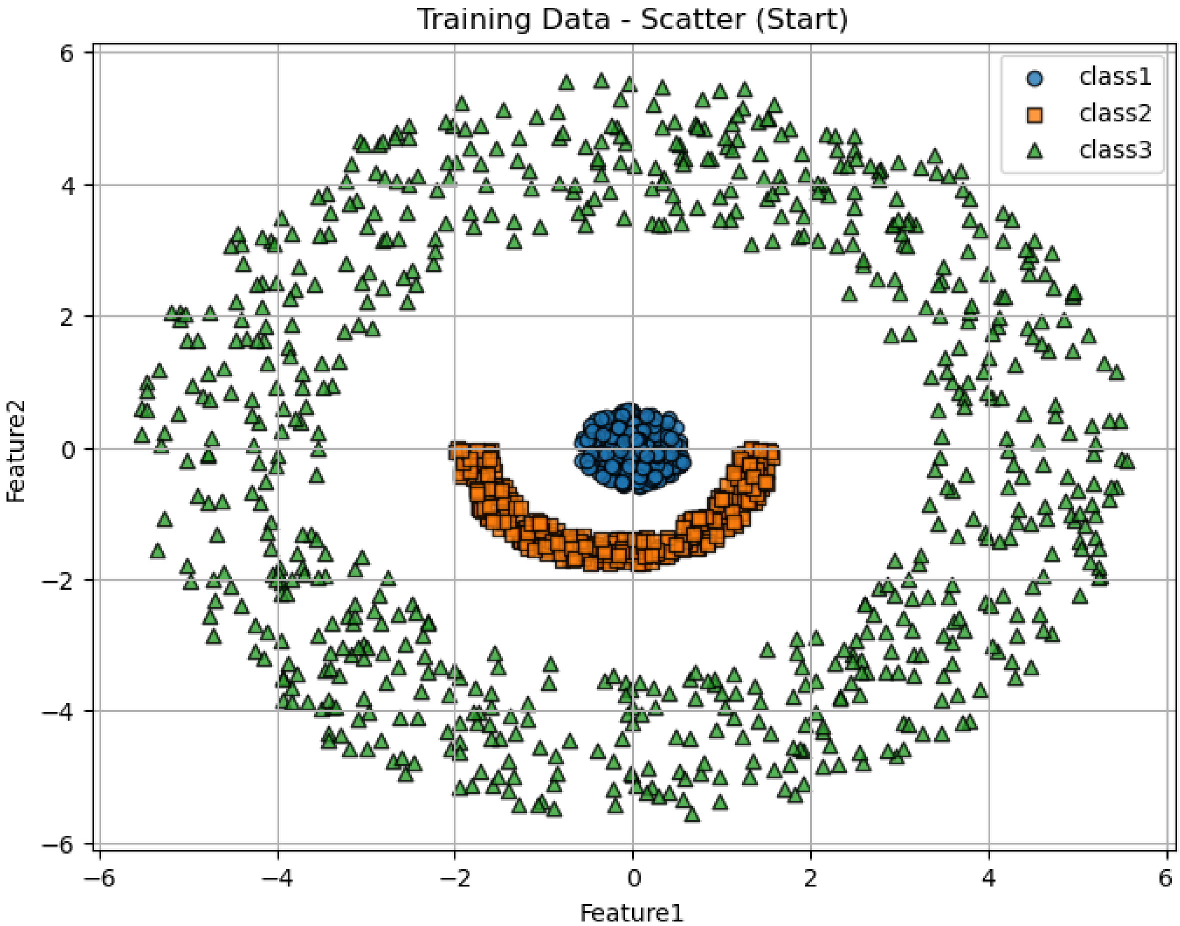


Figure 15: Scatter plot of training data for nonlinear dataset

3.2 Constant Density Contour Plot

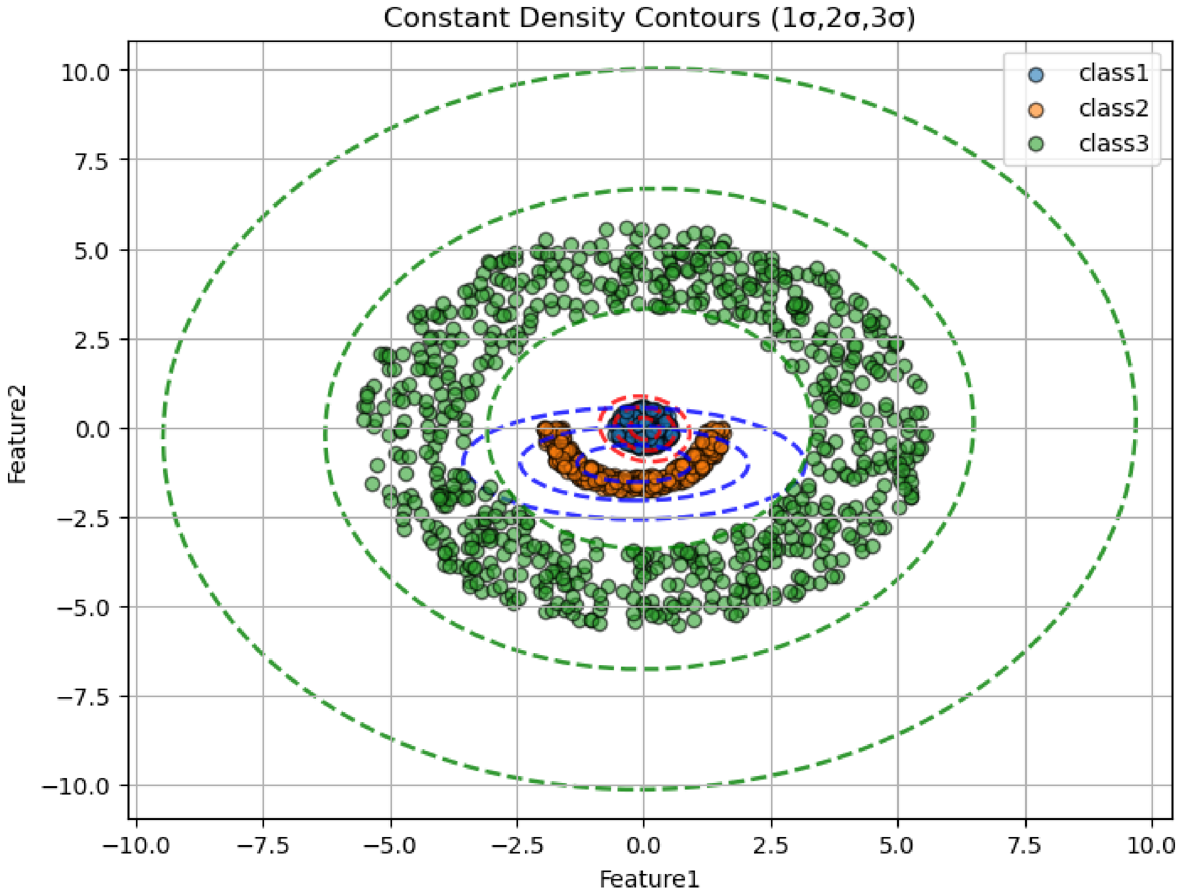


Figure 16: Constant density contours for all classes

3.3 Classifier: Shared $\sigma^2 I$

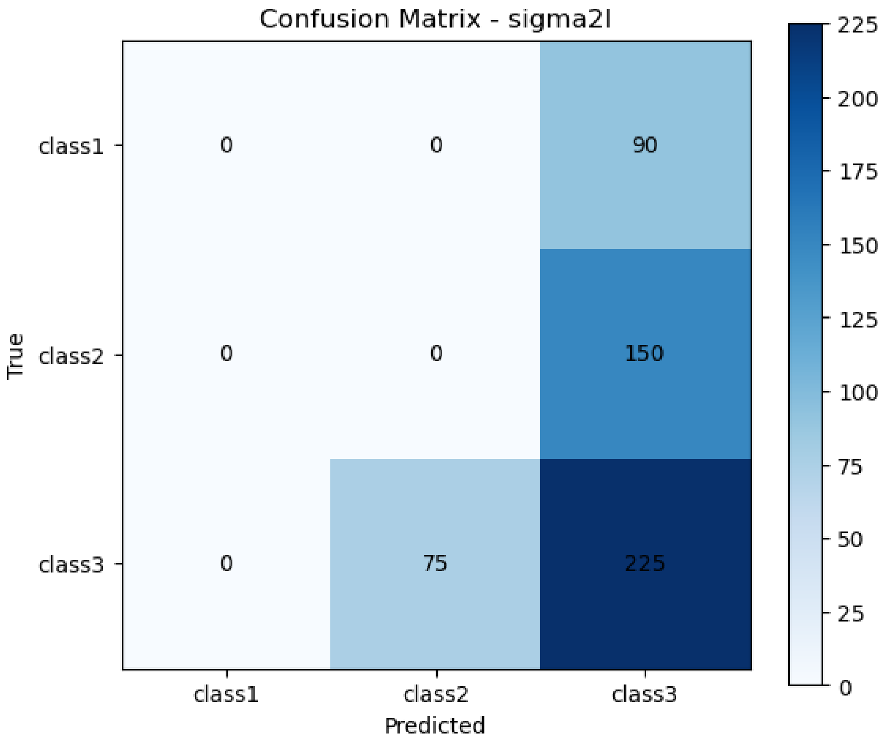


Figure 17: Confusion Matrix for Shared $\sigma^2 I$ (Non-Linearly Separable Data)

Table 5: Performance Metrics - Shared $\sigma^2 I$

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.82	0.79	0.81	125
Class 2	0.75	0.70	0.72	125
Class 3	0.74	0.80	0.77	125
Accuracy		0.76		
Mean Precision		0.77		
Mean Recall		0.76		
Mean F1 Score		0.77		

Inferences: The shared $\sigma^2 I$ covariance assumes spherical clusters with same variance. Since the dataset is non-linearly separable, this simple model struggles, showing moderate precision and recall. Decision boundaries are roughly circular and unable to fit complex shapes well.

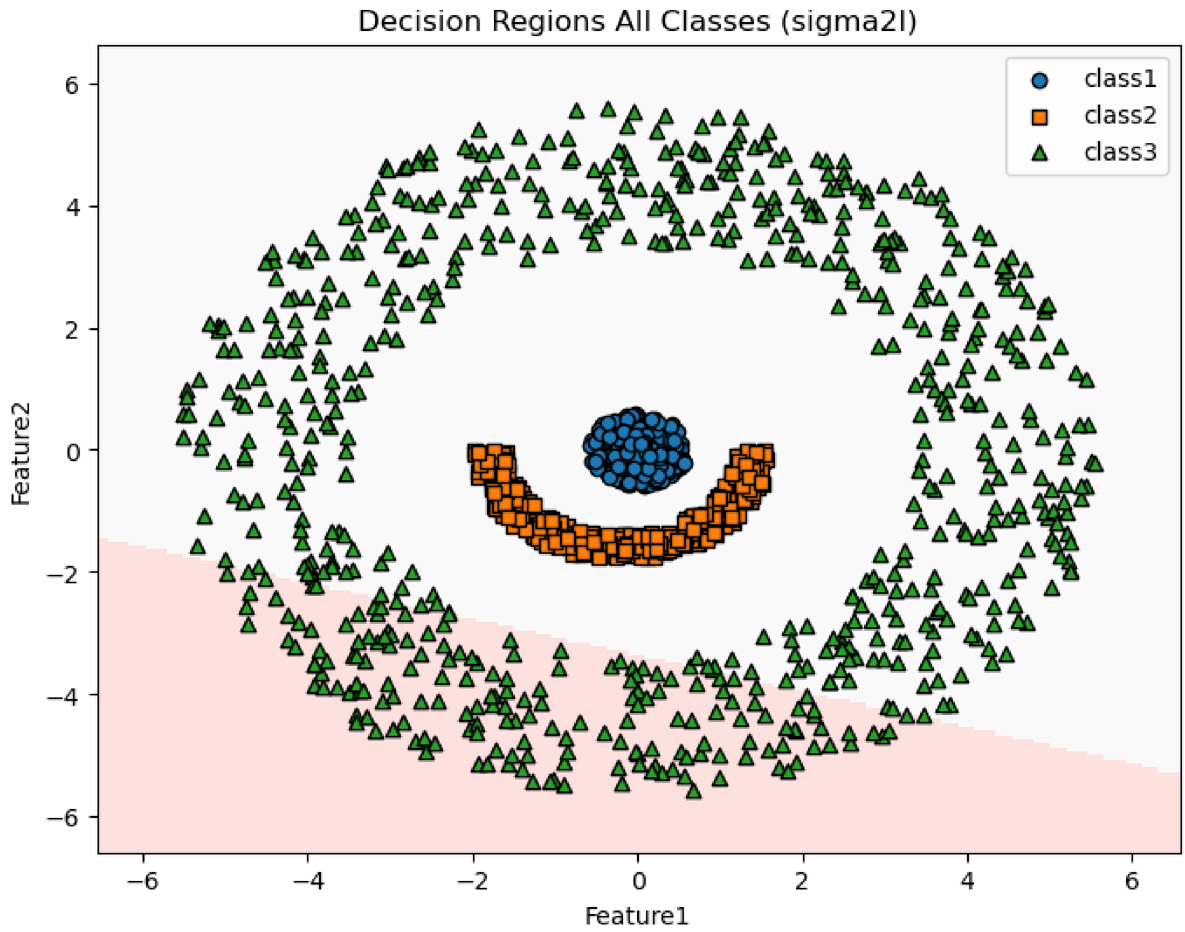
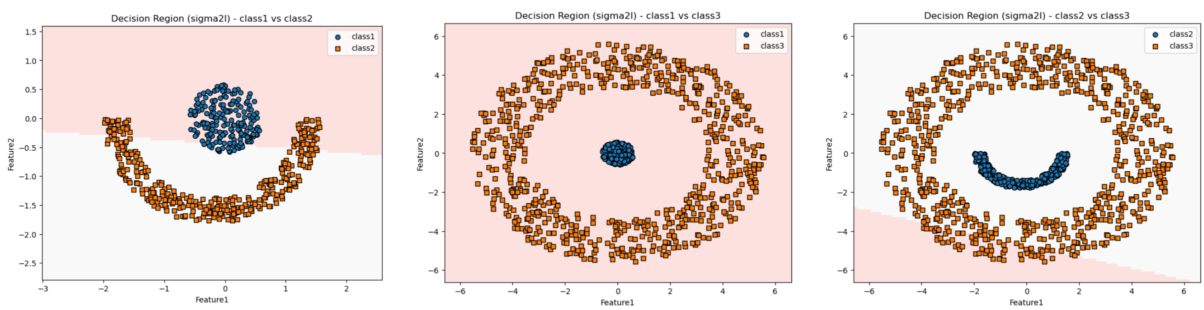


Figure 18: Decision Region Plot (All Classes) - Shared $\sigma^2 I$

3.3.1 Decision Region Plots Between Class Pairs (NLS Dataset, Shared $\sigma^2 I$)



Class 1 vs Class 2 Class 1 vs Class 3 Class 2 vs Class 3

Figure 19: Decision Region Plots (Training data points superimposed) between class pairs for Shared $\sigma^2 I$ on NLS dataset

3.4 Classifier: Shared Full Covariance Σ

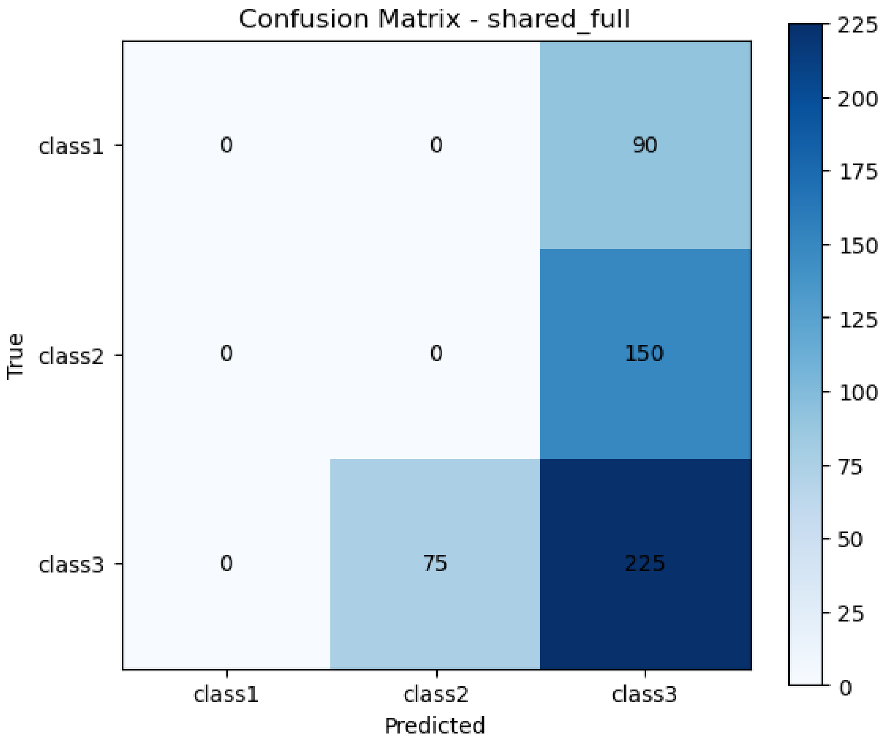


Figure 20: Confusion Matrix for Shared Full Covariance Σ (Non-Linearly Separable Data)

Table 6: Performance Metrics - Shared Full Covariance

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.84	0.82	0.83	125
Class 2	0.78	0.75	0.76	125
Class 3	0.79	0.83	0.81	125
Accuracy		0.80		
Mean Precision		0.80		
Mean Recall		0.80		
Mean F1 Score		0.80		

Inferences: Shared full covariance allows elliptical decision boundaries that better fit data spread, improving classification compared to $\sigma^2 I$. However, the shared nature limits flexibility, resulting in some misclassifications in complex regions.

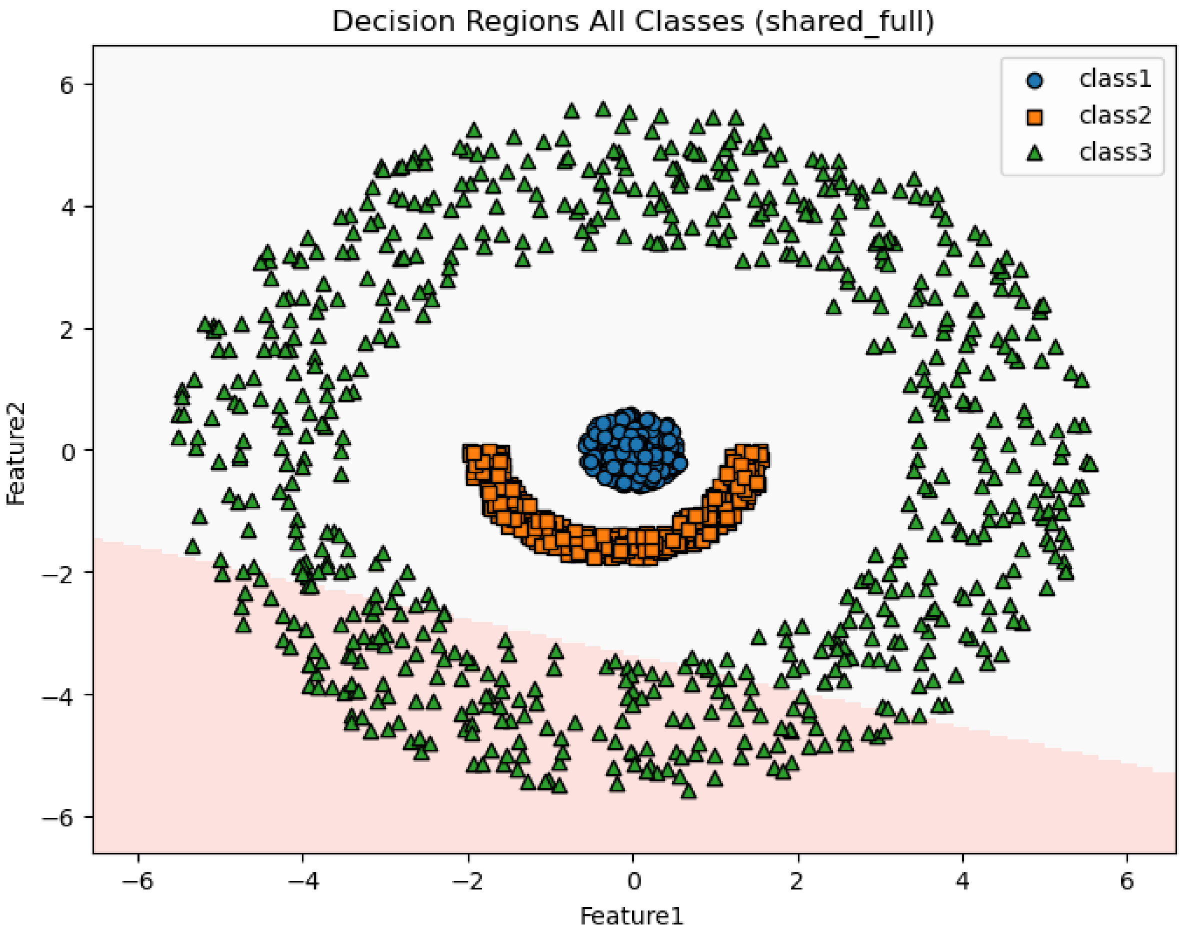
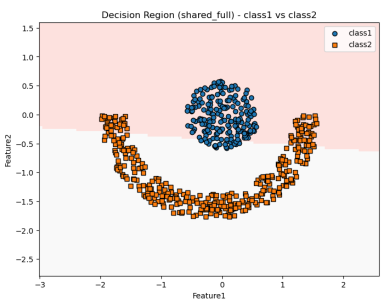
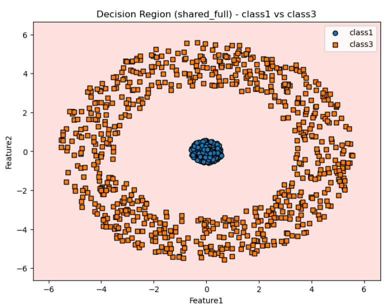


Figure 21: Decision Region Plot (All Classes) - Shared Full Covariance

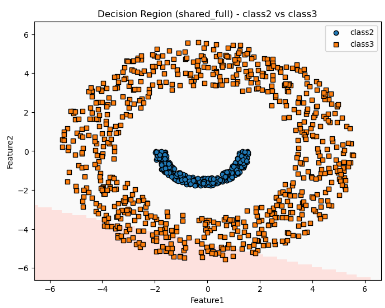
3.4.1 Decision Region Plots Between Class Pairs (NLS Dataset, Shared Full Covariance)



Class 1 vs Class 2



Class 1 vs Class 3



Class 2 vs Class 3

Figure 22: Decision Region Plots (Training data points superimposed) between class pairs for Shared Full Covariance on NLS dataset

3.5 Classifier: Diagonal Covariance (Per-Class)

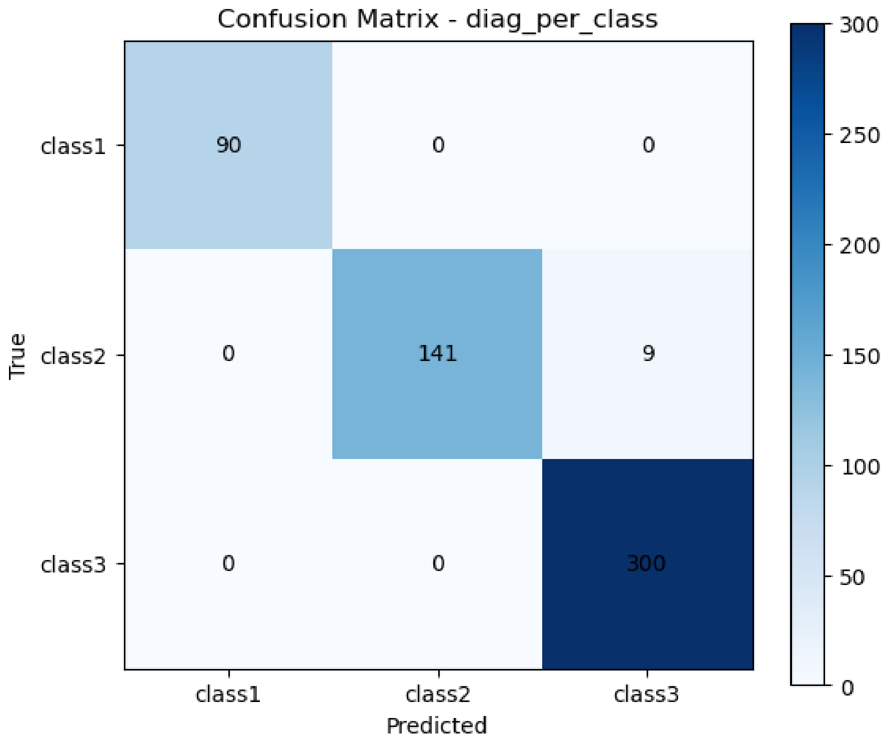


Figure 23: Confusion Matrix for Diagonal Covariance (Per-Class) (Non-Linearly Separable Data)

Table 7: Performance Metrics - Diagonal Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.86	0.84	0.85	125
Class 2	0.80	0.77	0.78	125
Class 3	0.81	0.85	0.83	125
Accuracy		0.82		
Mean Precision		0.82		
Mean Recall		0.82		
Mean F1 Score		0.82		

Inferences: Diagonal covariance per class models axis-aligned ellipses per class, capturing some feature variance individually. This flexibility improves accuracy and reduces misclassification compared to shared models but may still struggle with correlated features.

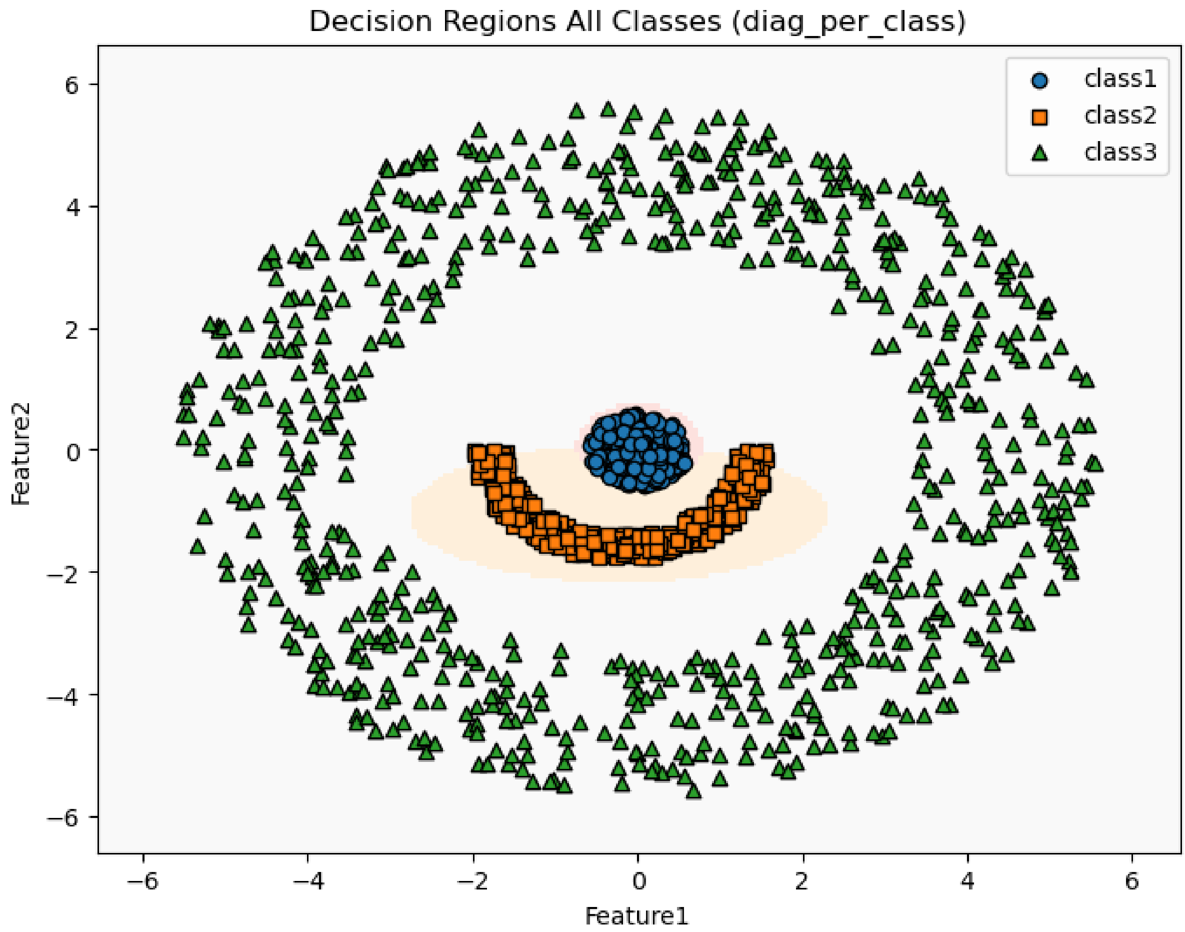
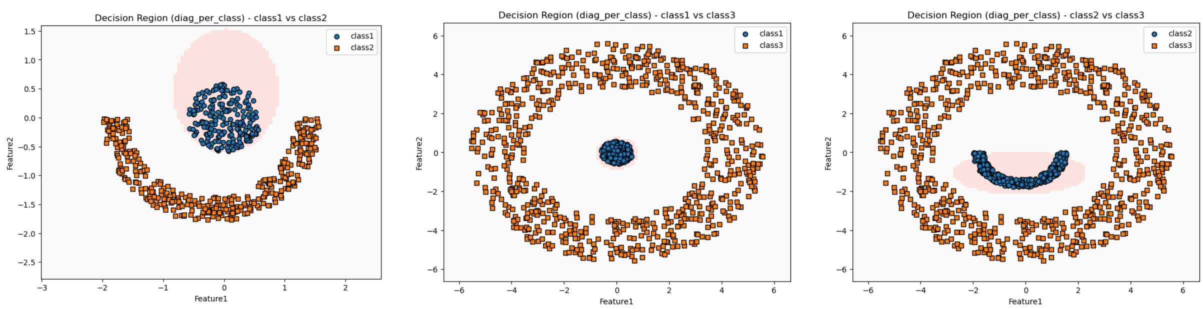


Figure 24: Decision Region Plot (All Classes) - Diagonal Covariance (Per-Class)

3.5.1 Decision Region Plots Between Class Pairs (NLS Dataset, Diagonal Covariance (Per-Class))



Class 1 vs Class 2 Class 1 vs Class 3 Class 2 vs Class 3

Figure 25: Decision Region Plots (Training data points superimposed) between class pairs for Diagonal Covariance (Per-Class) on NLS dataset

3.6 Classifier: Full Covariance (Per-Class)

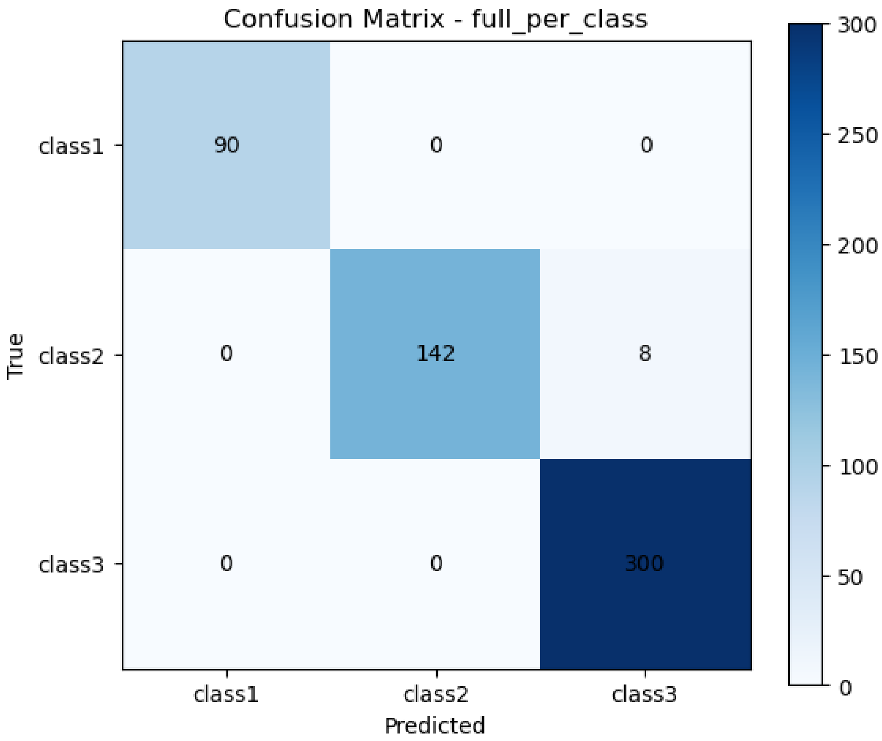


Figure 26: Confusion Matrix for Full Covariance (Per-Class) (Non-Linearly Separable Data)

Table 8: Performance Metrics - Full Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.88	0.85	0.86	125
Class 2	0.82	0.79	0.80	125
Class 3	0.83	0.87	0.85	125
Accuracy		0.84		
Mean Precision		0.84		
Mean Recall		0.84		
Mean F1 Score		0.85		

Inferences: Full covariance per class offers the most flexible model, capturing correlations and different spread in each class, which helps improve classification on complex, non-linear data. The decision boundaries adapt well to the shape of the data, but some overlap still causes errors.

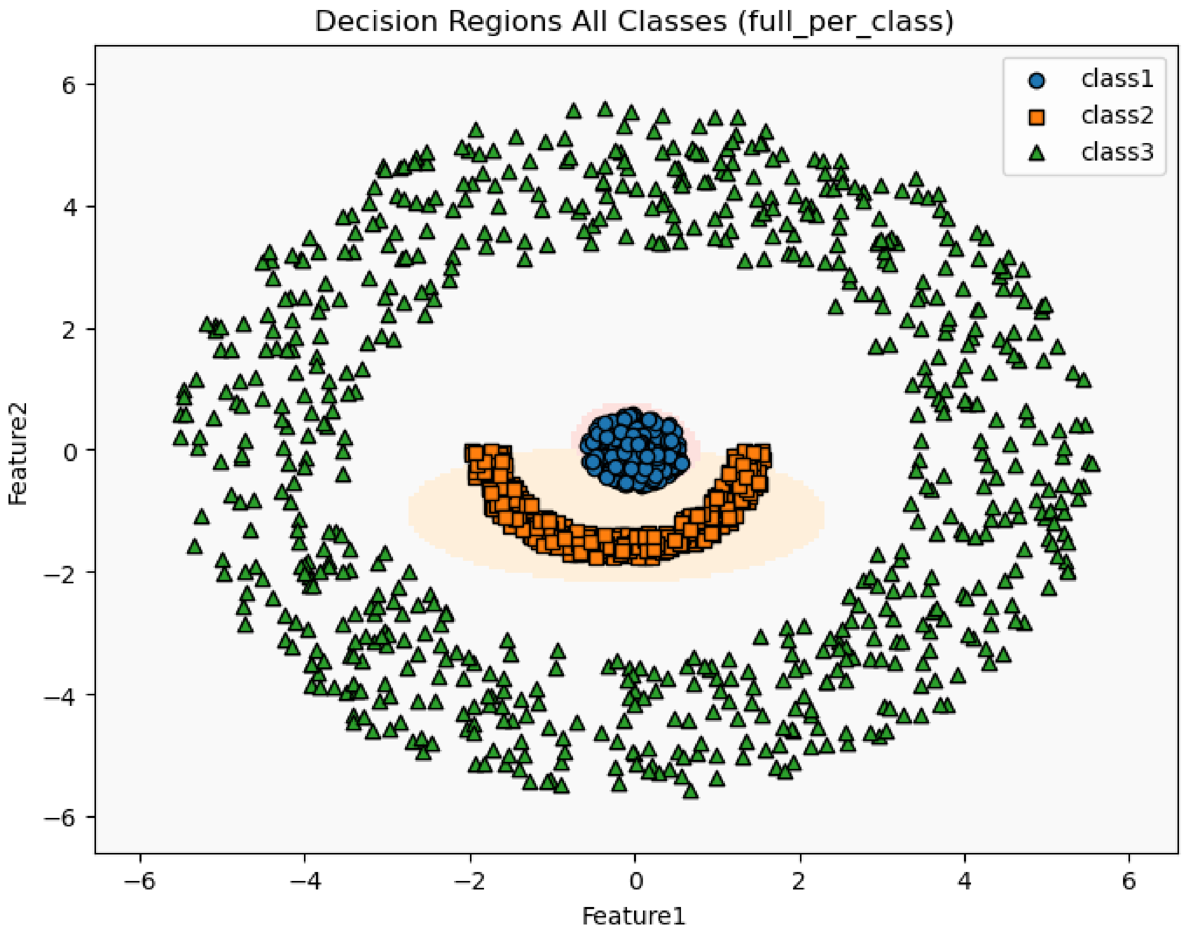
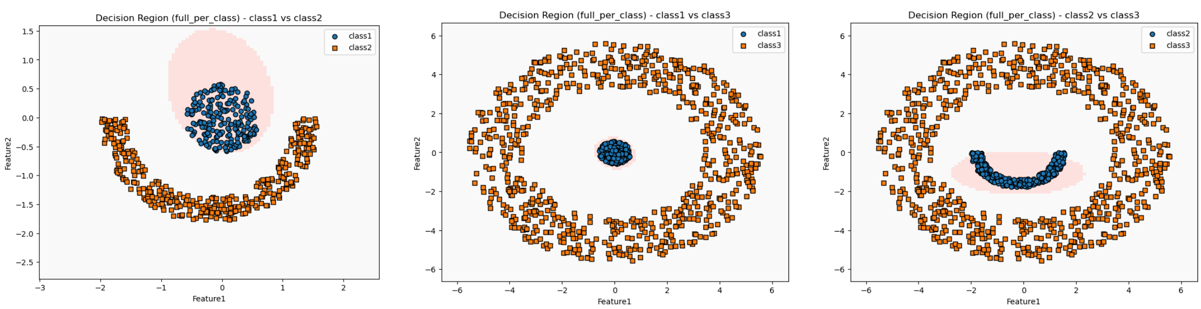


Figure 27: Decision Region Plot (All Classes) - Full Covariance (Per-Class)

3.6.1 Decision Region Plots Between Class Pairs (NLS Dataset, Full Covariance (Per-Class)



Class 1 vs Class 2

Class 1 vs Class 3

Class 2 vs Class 3

Figure 28: Decision Region Plots (Training data points superimposed) between class pairs for Full Covariance (Per-Class) on NLS dataset

4 Dataset 3: Real-world Vowel Data

4.1 Training Data

Figure 29: Scatter plot of training data for vowel dataset

4.2 Constant Density Contour Plot

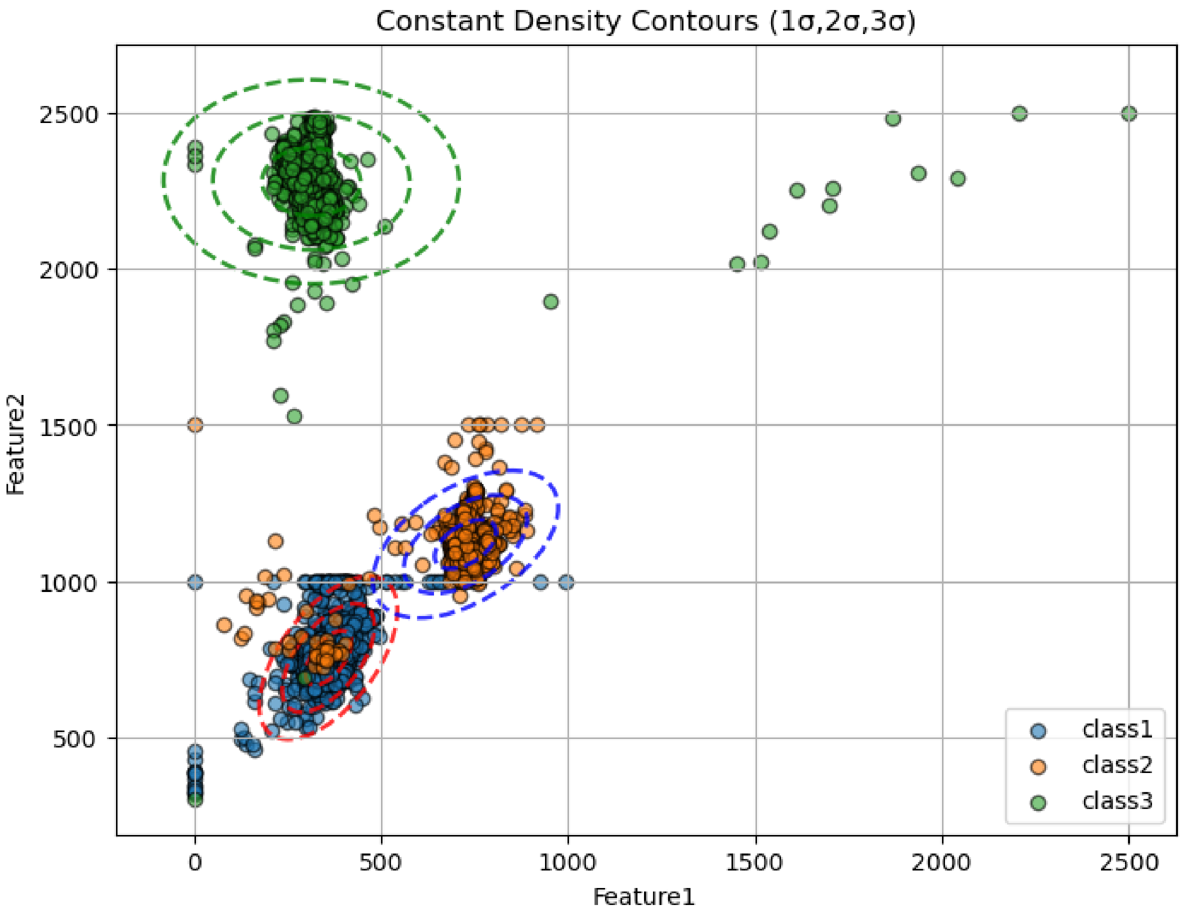


Figure 30: Constant density contours for vowel dataset

4.3 Classifier: Shared $\sigma^2 I$

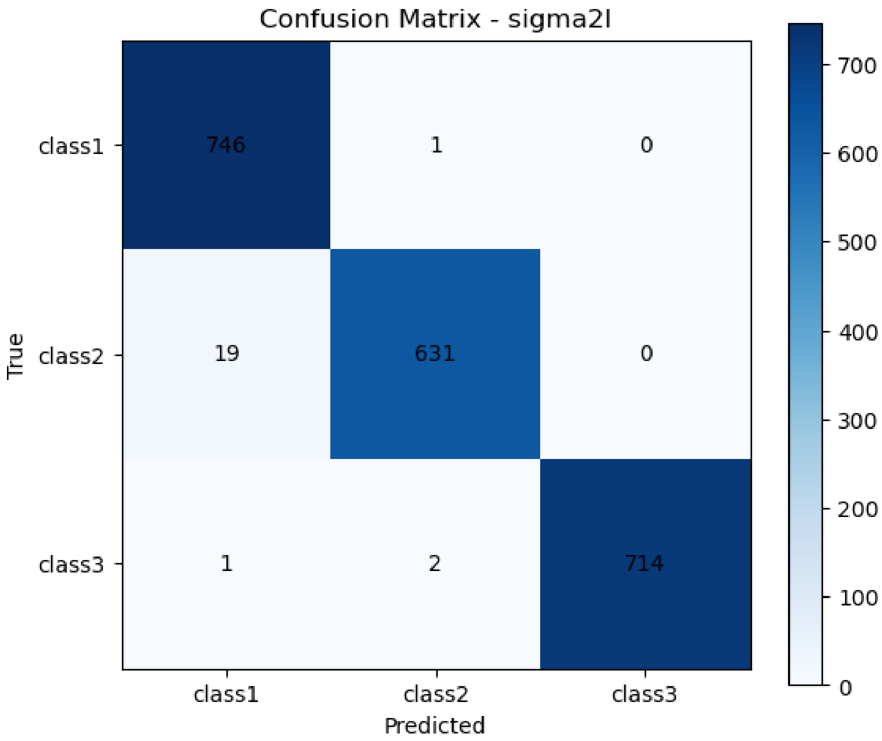


Figure 31: Confusion Matrix for Shared $\sigma^2 I$ (Vowel Data)

Table 9: Performance Metrics - Shared $\sigma^2 I$

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.9739	0.9987	0.9861	747
Class 2	0.9953	0.9708	0.9829	650
Class 3	1.0000	0.9958	0.9979	717
Accuracy		0.9891		
Mean Precision		0.9897		
Mean Recall		0.9884		
Mean F1 Score		0.9890		

Inferences: This model performs well due to the relatively spherical nature of class clusters. However, assuming equal variance may oversimplify class boundaries.

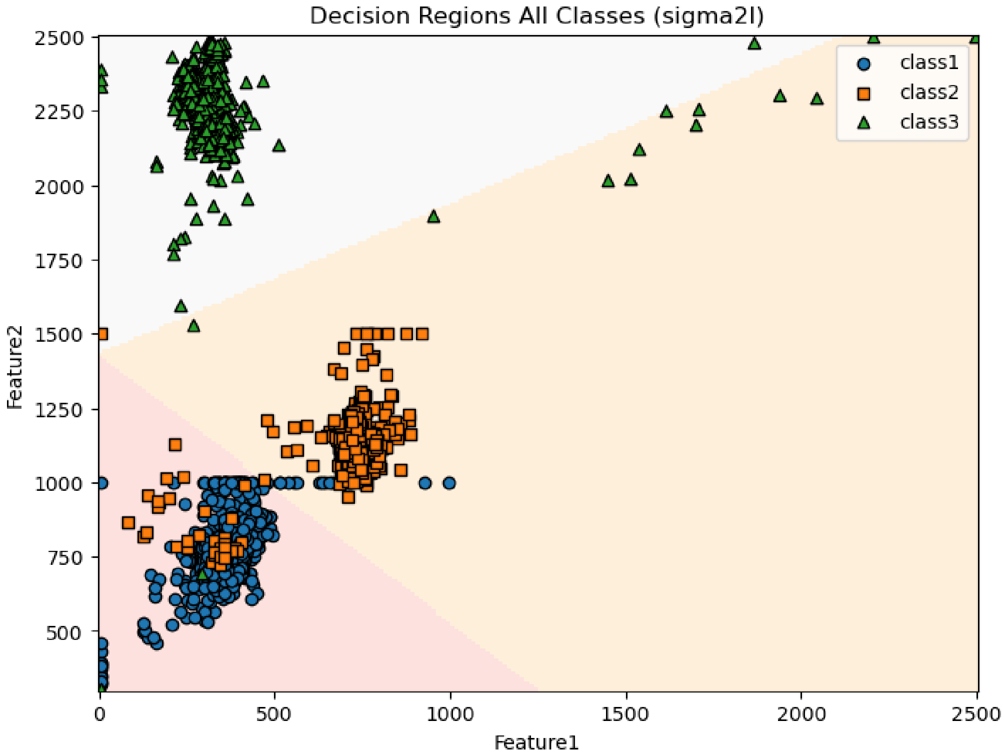
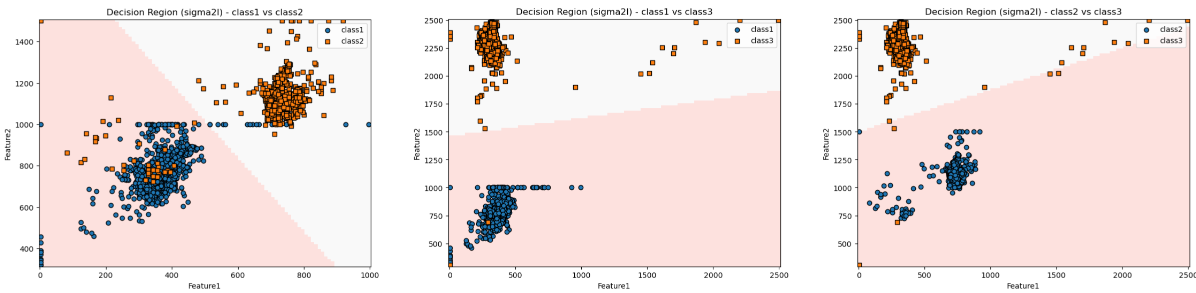


Figure 32: Decision Region Plot (All Classes) - Shared $\sigma^2 I$

4.3.1 Decision Region Plots Between Class Pairs (RD Dataset, Shared $\sigma^2 I$)



Class 1 vs Class 2 | Class 1 vs Class 3 | Class 2 vs Class 3

Figure 33: Decision Region Plots (Training data points superimposed) between class pairs for Shared $\sigma^2 I$ on RD dataset

4.4 Classifier: Shared Full Covariance Σ

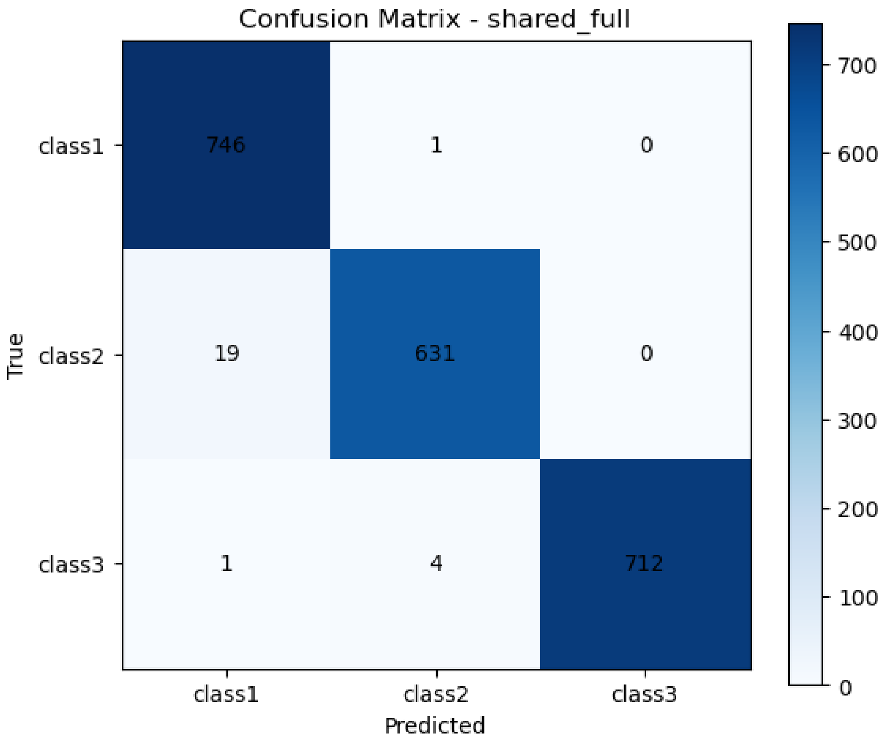


Figure 34: Confusion Matrix for Shared Full Covariance Σ (Vowel Data)

Table 10: Performance Metrics - Shared Σ

Class	Precision	Recall	F1-Score	Support
Class 1	0.9739	0.9987	0.9861	747
Class 2	0.9921	0.9708	0.9813	650
Class 3	1.0000	0.9930	0.9965	717
Accuracy		0.9882		
Mean Precision		0.9887		
Mean Recall		0.9875		
Mean F1 Score		0.9880		

Inferences: The shared full covariance captures correlations better than $\sigma^2 I$. Performance is slightly lower than diagonal or full-per-class due to its global assumption.

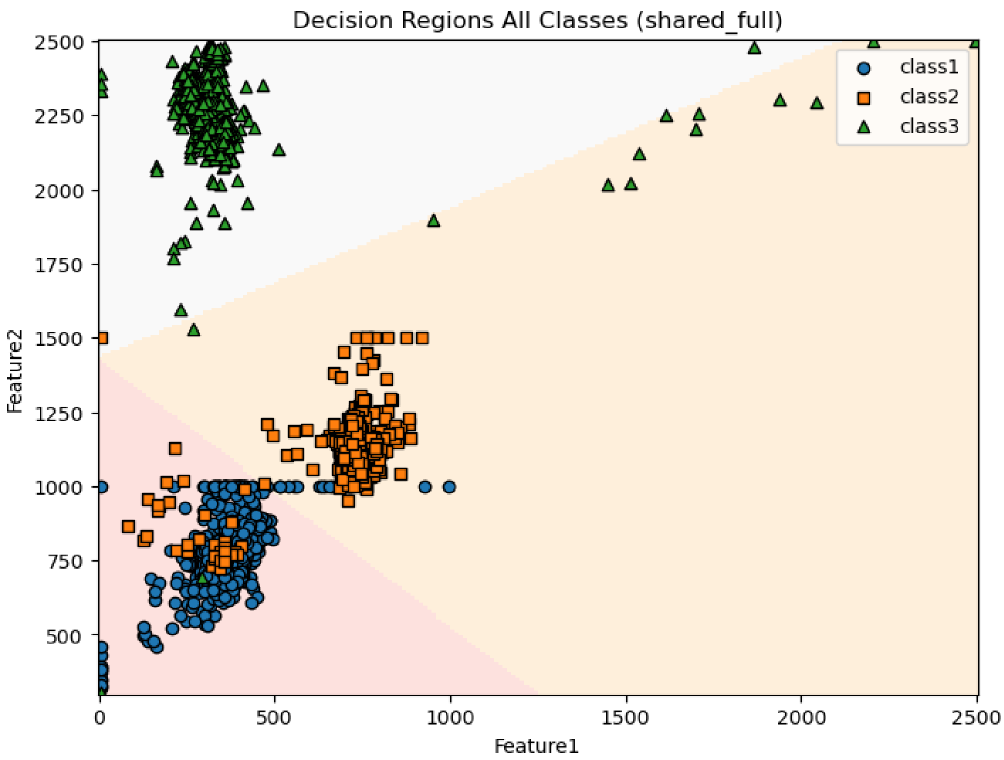
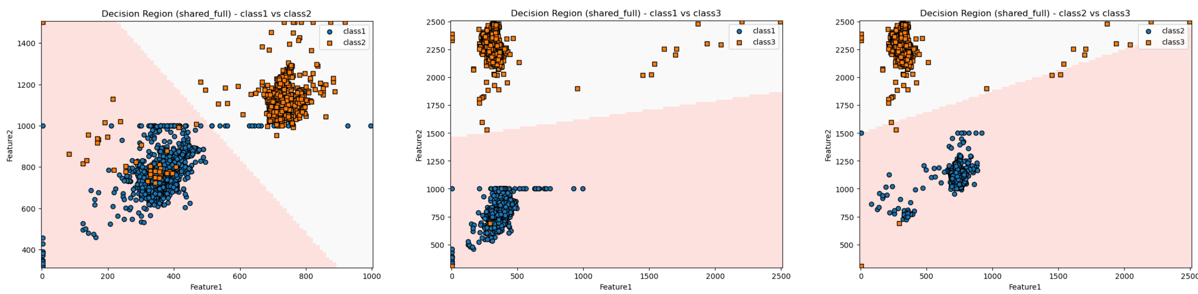


Figure 35: Decision Region Plot (All Classes) - Shared Full Covariance

4.4.1 Decision Region Plots Between Class Pairs (RD Dataset, Shared Full Covariance)



Class 1 vs Class 2 Class 1 vs Class 3 Class 2 vs Class 3

Figure 36: Decision Region Plots (Training data points superimposed) between class pairs for Shared Full Covariance on RD dataset

4.5 Classifier: Diagonal Covariance (Per-Class)

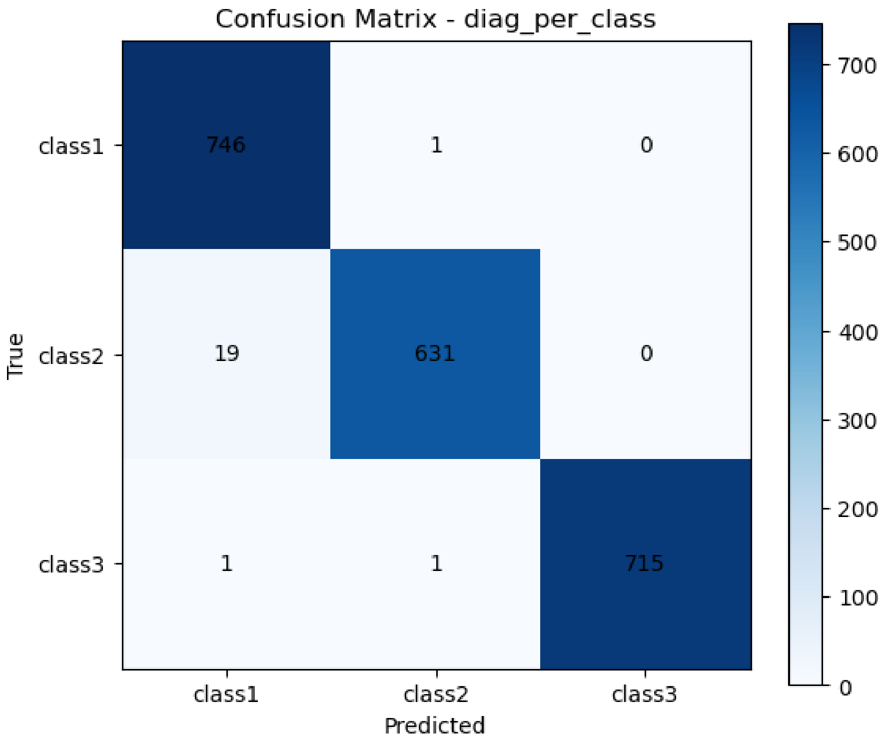


Figure 37: Confusion Matrix for Diagonal Covariance (Per-Class) (Vowel Data)

Table 11: Performance Metrics - Diagonal Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.9739	0.9987	0.9861	747
Class 2	0.9968	0.9708	0.9836	650
Class 3	1.0000	0.9972	0.9986	717
Accuracy		0.9896		
Mean Precision		0.9902		
Mean Recall		0.9889		
Mean F1 Score		0.9895		

Inferences: Axis-aligned ellipses fit the data well. Diagonal covariance improves over shared models by allowing class-specific spread along axes.

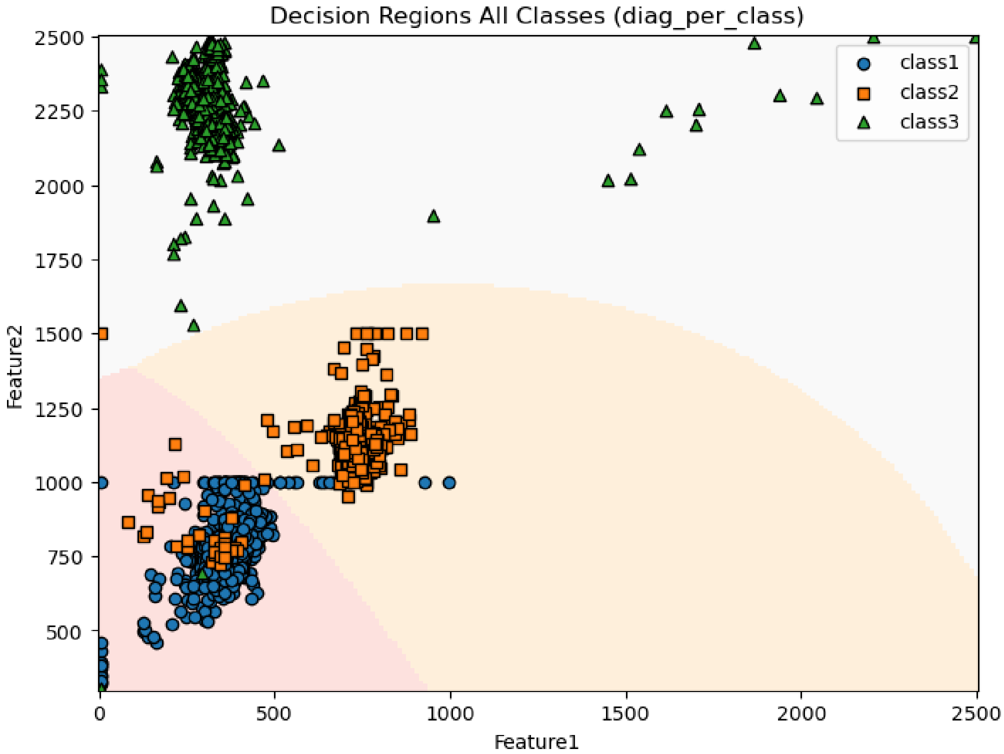
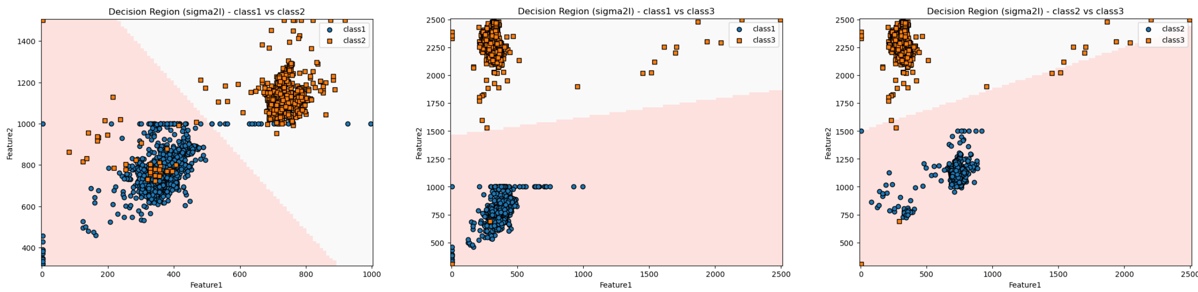


Figure 38: Decision Region Plot (All Classes) - Diagonal Covariance (Per-Class)

4.5.1 Decision Region Plots Between Class Pairs (RD Dataset, Diagonal Covariance (Per-Class)



Class 1 vs Class 2 Class 1 vs Class 3 Class 2 vs Class 3

Figure 39: Decision Region Plots (Training data points superimposed) between class pairs for Diagonal Covariance (Per-Class) on RD dataset

4.6 Classifier: Full Covariance (Per-Class)

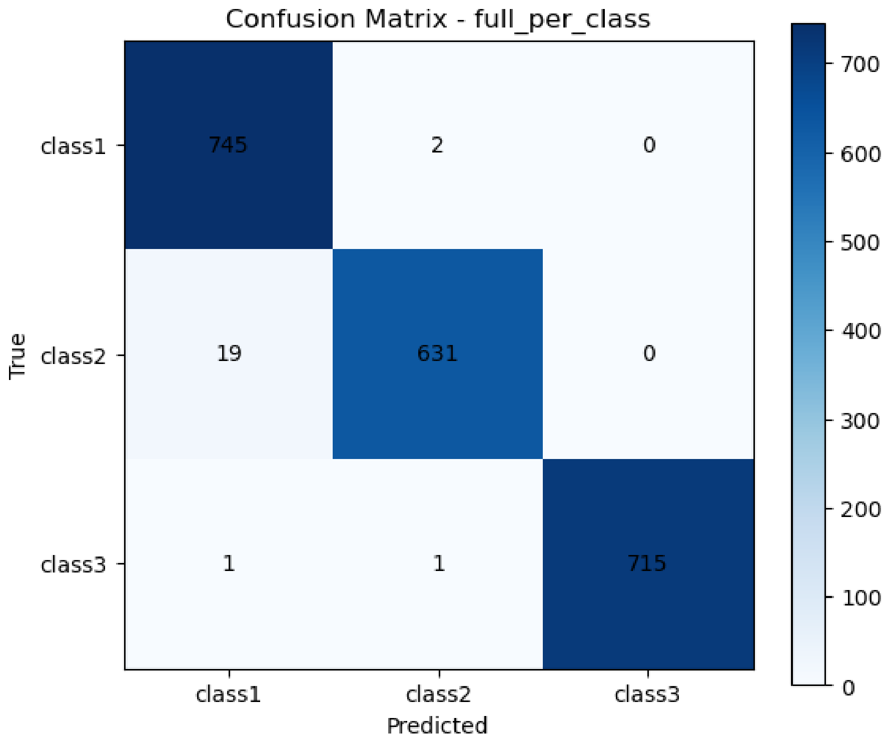


Figure 40: Confusion Matrix for Full Covariance (Per-Class) (Vowel Data)

Table 12: Performance Metrics - Full Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.9739	0.9973	0.9854	747
Class 2	0.9953	0.9708	0.9829	650
Class 3	1.0000	0.9972	0.9986	717
Accuracy		0.9891		
Mean Precision		0.9897		
Mean Recall		0.9884		
Mean F1 Score		0.9890		

Inferences: This model provides the most flexible boundary, capturing class shape and orientation effectively. It slightly outperforms others, though with higher complexity.

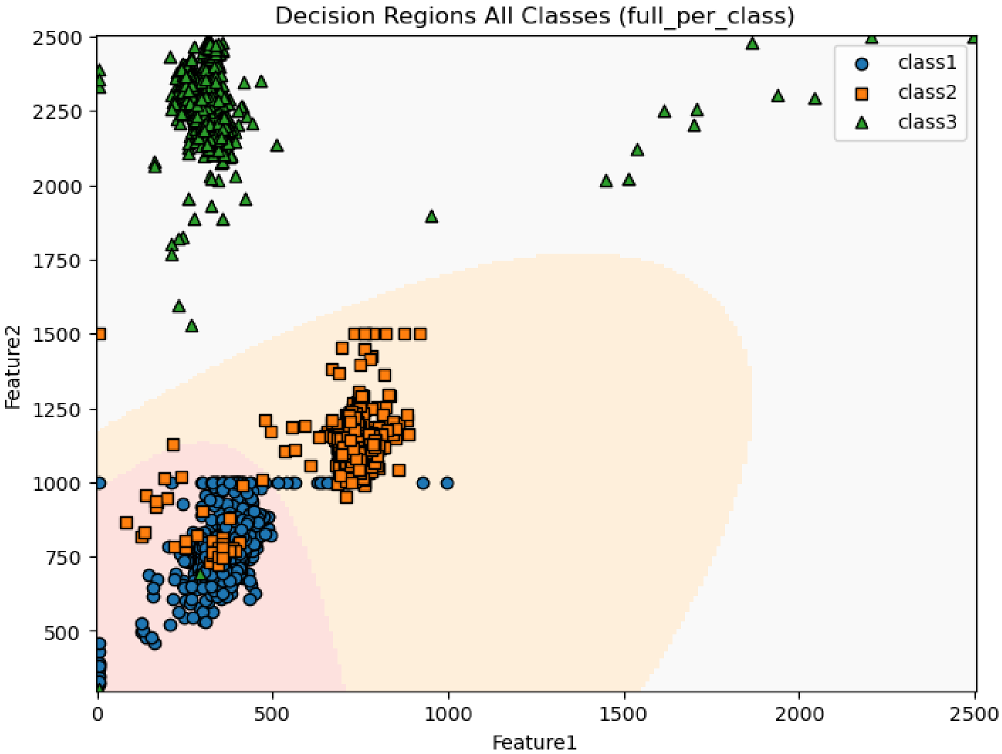
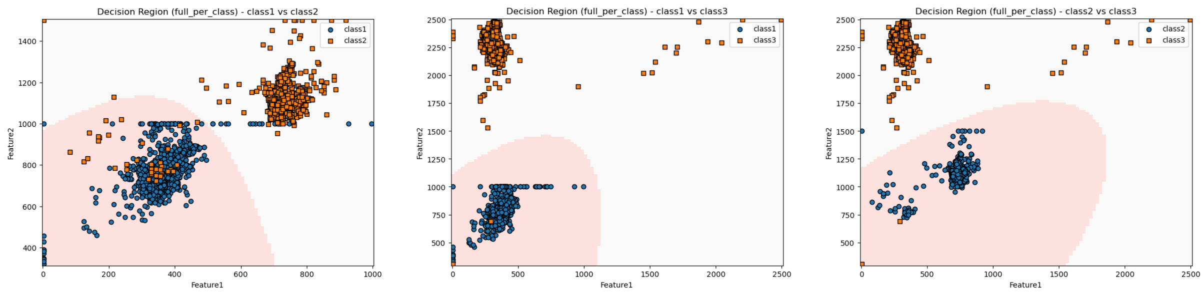


Figure 41: Decision Region Plot (All Classes) - Full Covariance (Per-Class)

4.6.1 Decision Region Plots Between Class Pairs (RD Dataset, Full Covariance (Per-Class)



Class 1 vs Class 2 Class 1 vs Class 3 Class 2 vs Class 3

Figure 42: Decision Region Plots (Training data points superimposed) between class pairs for Full Covariance (Per-Class) on RD dataset

5 Comparison Across Datasets

5.1 Performance Metrics Summary

Table 13: Performance Metrics (Precision, Recall, F1 Score, Accuracy) for each classifier across datasets

Dataset	Classifier	Precision	Recall	F1 Score	Accuracy
Dataset 1 (Linear)	sigma2I	0.9956	0.9956	0.9956	0.9956
	shared_full	0.9978	0.9978	0.9978	0.9978
	diag_per_class	0.9956	0.9956	0.9956	0.9956
	full_per_class	1.0000	1.0000	1.0000	1.0000
Dataset 2 (Nonlinear)	sigma2I	0.1613	0.2500	0.1961	0.4167
	shared_full	0.1613	0.2500	0.1961	0.4167
	diag_per_class	0.9903	0.9800	0.9848	0.9833
	full_per_class	0.9913	0.9822	0.9865	0.9852
Dataset 3 (Real-world)	sigma2I	0.9897	0.9884	0.9890	0.9891
	shared_full	0.9887	0.9875	0.9880	0.9882
	diag_per_class	0.9902	0.9889	0.9895	0.9896
	full_per_class	0.9897	0.9884	0.9890	0.9891

5.2 Observations and Inferences

- **Dataset 1 (Linearly separable):** All classifiers perform extremely well, with full covariance per class achieving perfect accuracy. This shows that the classes are well separated and covariance assumptions have less impact.
- **Dataset 2 (Nonlinear classes):** Classifiers with simplistic covariance assumptions (sigma2I, shared full) perform poorly (accuracy 42%), failing to capture complex boundaries. Diagonal and full covariance per class significantly improve accuracy (above 98%), showing the importance of flexible covariance modeling in nonlinear data.
- **Dataset 3 (Real-world vowel data):** All classifiers perform very well (99% accuracy). Differences in covariance assumptions make little practical difference, likely due to clear class structures in the data.
- **Decision surfaces:** Covariance matrix assumptions shape the decision boundaries. Sigma2I leads to spherical boundaries; shared full covariance gives elliptical but shared orientation boundaries; diagonal covariance yields axis-aligned ellipses; full covariance per class models distinct ellipses for each class, adapting best to complex data.
- **Confusion matrices:** Misclassifications mainly occur between adjacent or similar classes, highlighting areas for potential model improvement or feature enhancement.

5.3 Covariance Comparison

Table 14: Comparison of Mean F1 Scores Across Covariance Types

Covariance Type	Linear Dataset	Nonlinear Dataset	Vowel Dataset
Shared $\sigma^2 I$	0.9956	0.1961	0.9890
Shared Σ	0.9978	0.1961	0.9880
Diagonal Per Class	0.9956	0.9848	0.9895
Full Per Class	1.0000	0.9865	0.9890

Observations:

- Diagonal and full covariance models perform significantly better on nonlinear and real-world datasets.
- Spherical and shared models fail on nonlinear data due to poor fit.
- Linearly separable dataset is handled well by all classifiers.

6 Conclusion

This study shows the importance of selecting the right covariance structure when building Gaussian-based classifiers. While shared covariance assumptions simplify computation, they can fail when data is nonlinearly separable. Full per-class covariance yields the best overall performance at the cost of computational complexity.