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

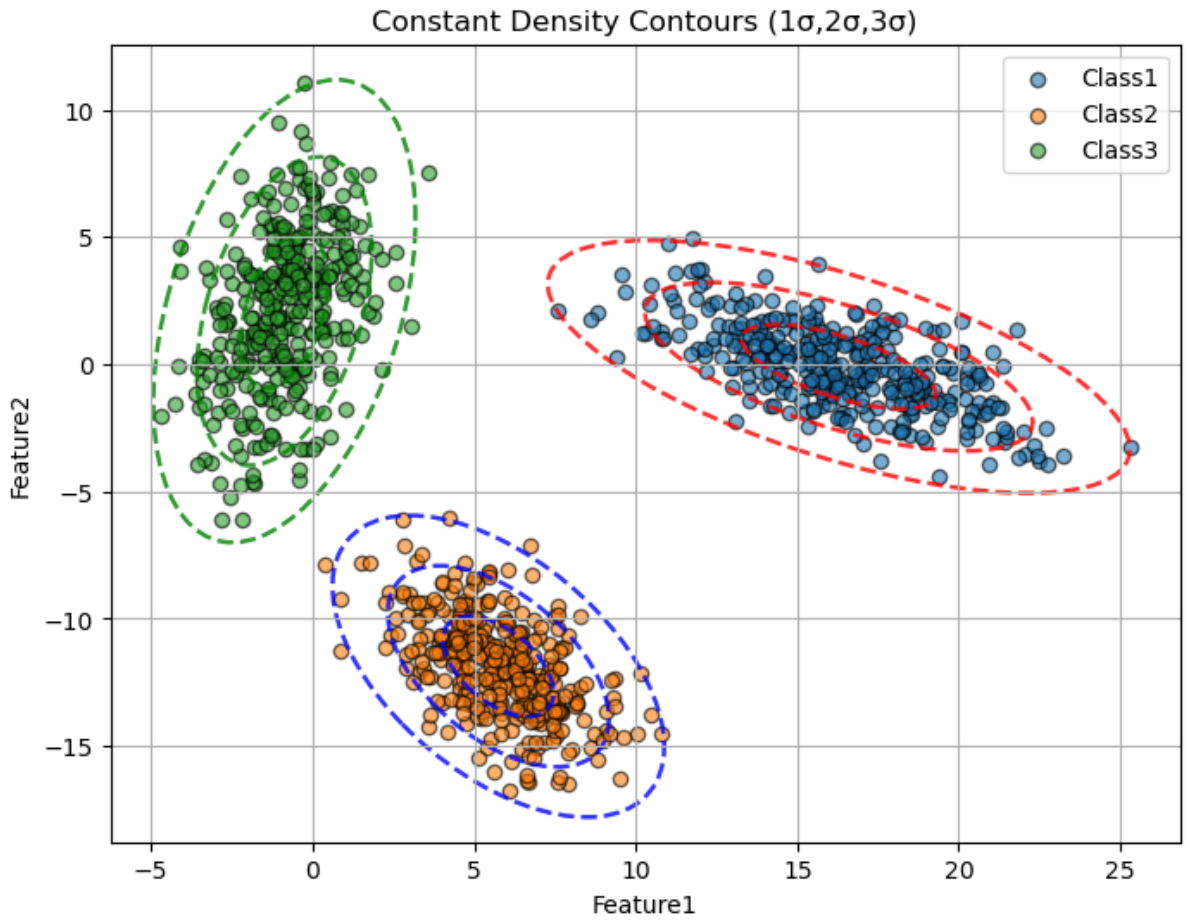


Figure 2: Constant density contours for all classes

2.3 Classifier: Shared $\sigma^2 I$

Table 1: Confusion Matrix for Shared $\sigma^2 I$ (Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	148	0	2
Class 2	0	150	0
Class 3	0	0	150

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

Class	Precision	Recall	F1-Score	Support
Class 1	1.0000	0.9867	0.9933	150
Class 2	1.0000	1.0000	1.0000	150
Class 3	0.9868	1.0000	0.9934	150
Accuracy		0.9956		
Mean Precision		0.9956		
Mean Recall		0.9956		
Mean F1 Score		0.9956		

Inference: The classifier with shared $\sigma^2 I$ achieved very high accuracy ($\approx 99.6\%$) on the linearly separable dataset. Only a few misclassifications occurred in Class 1, showing that simple covariance is sufficient for this dataset.

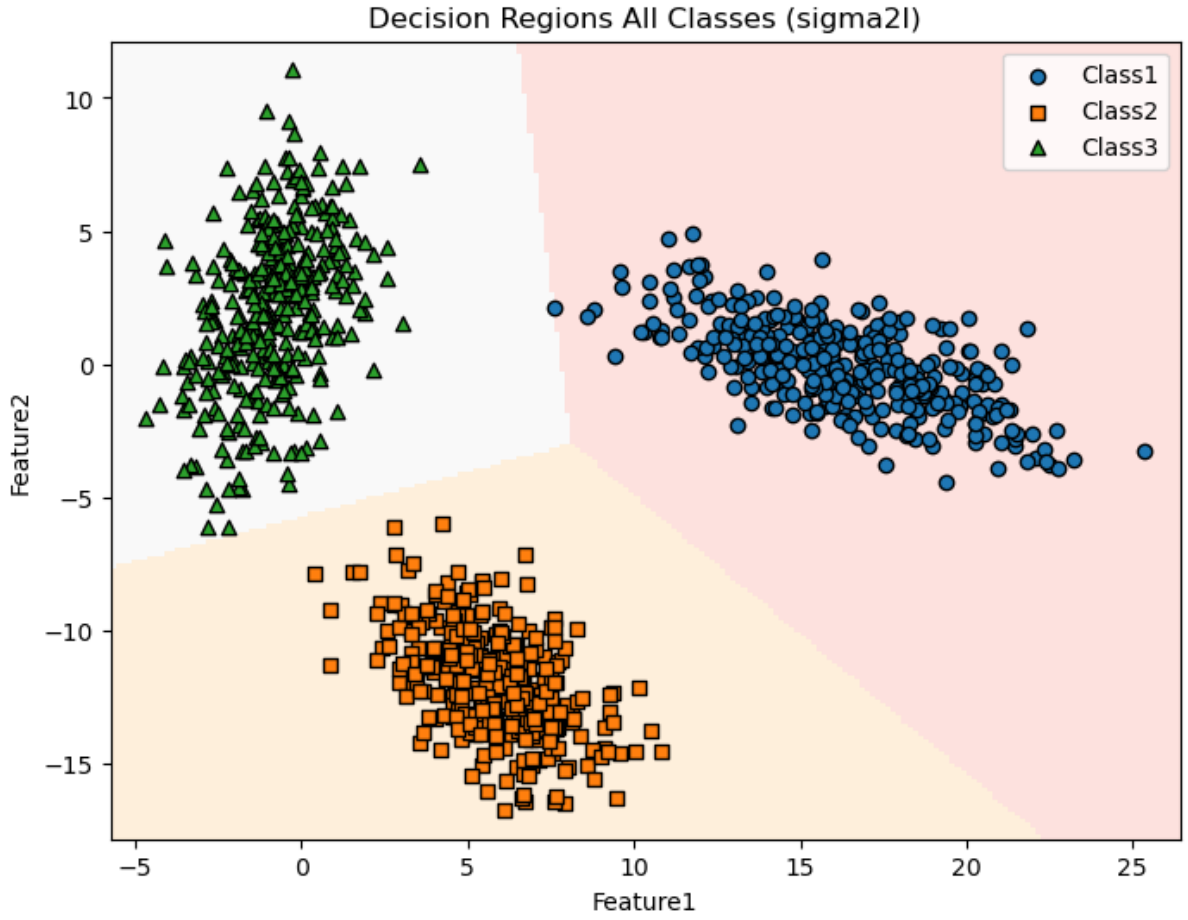


Figure 3: Decision Region Plot (All Classes) - Shared $\sigma^2 I$

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 4: 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

Table 3: Confusion Matrix for Shared Full Covariance Σ (Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	149	0	1
Class 2	0	150	0
Class 3	0	0	150

Table 4: Performance Metrics - Shared Full Covariance

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	0.9933	0.9967	150
Class 2	1.0000	1.0000	1.0000	150
Class 3	0.9934	1.0000	0.9967	150
Accuracy		0.9978		
Mean Precision		0.9978		
Mean Recall		0.9978		
Mean F1 Score		0.9978		

Inference: The shared full covariance Σ model achieved very high accuracy ($\approx 99.8\%$) with only one misclassification, showing excellent performance on linearly separable data.

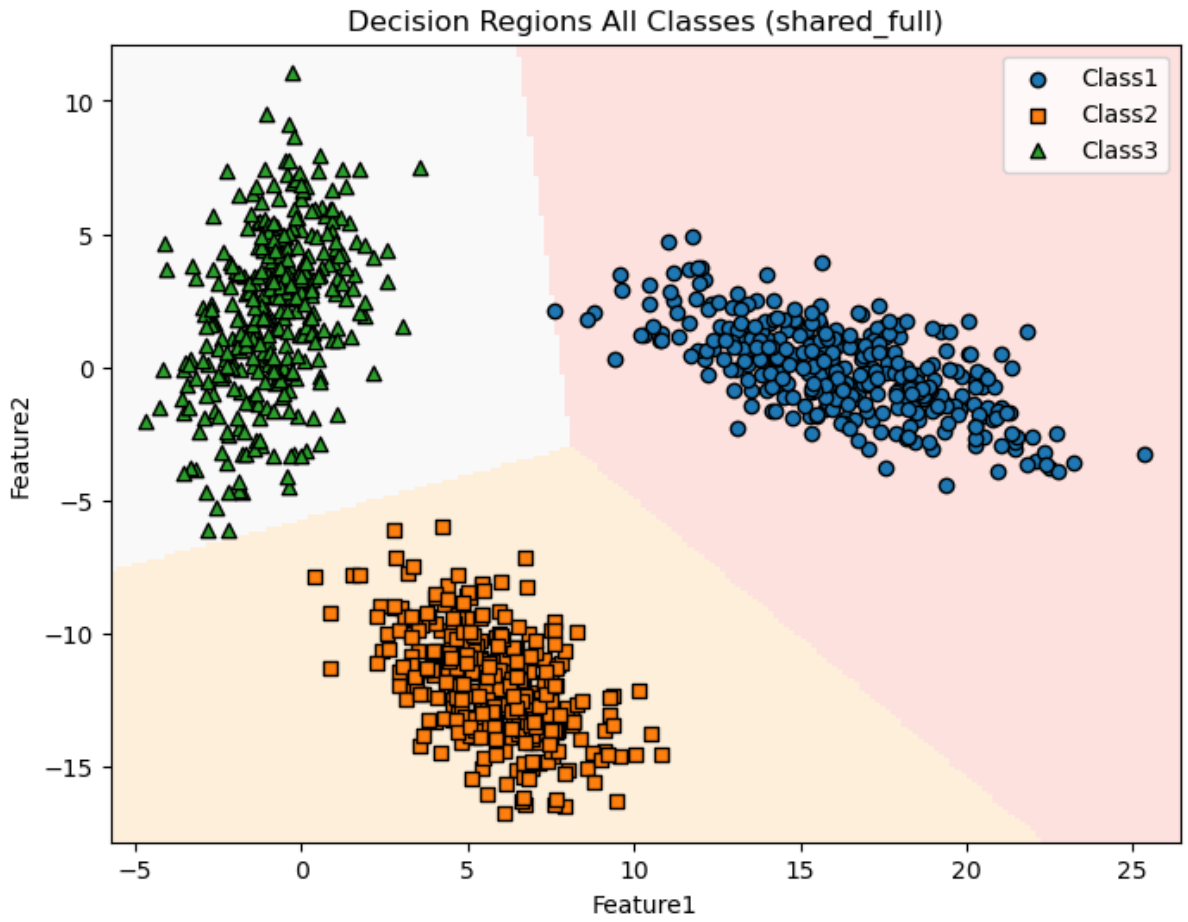
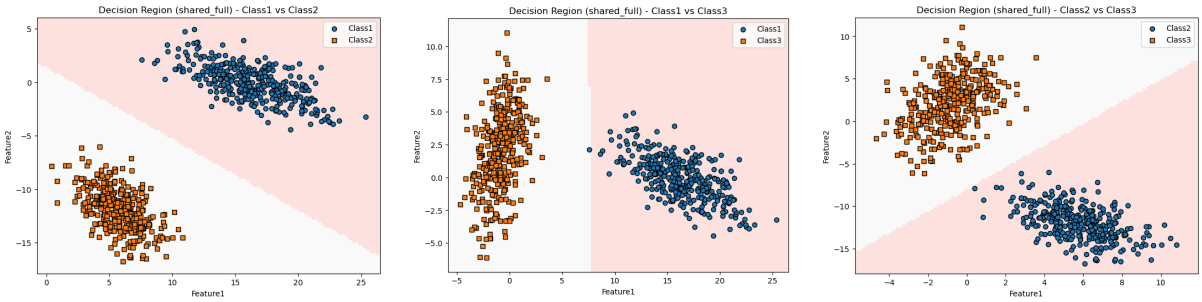


Figure 5: Decision Region Plot (All Classes) - Shared Full Covariance

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

2.5 Classifier: Diagonal Covariance (Per-Class)

Table 5: Confusion Matrix for Diagonal Covariance (Per-Class) (Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	150	0	0
Class 2	0	148	2
Class 3	0	0	150

Table 6: Performance Metrics - Diagonal Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	1.0000	1.0000	150
Class 2	1.0000	0.9867	0.9933	150
Class 3	0.9868	1.0000	0.9934	150
Accuracy		0.9956		
Mean Precision		0.9956		
Mean Recall		0.9956		
Mean F1 Score		0.9956		

Inference: The diagonal per-class covariance model gave very high accuracy ($\approx 99.6\%$), with only a couple of errors in Class 2, proving it is also effective for linearly separable data.

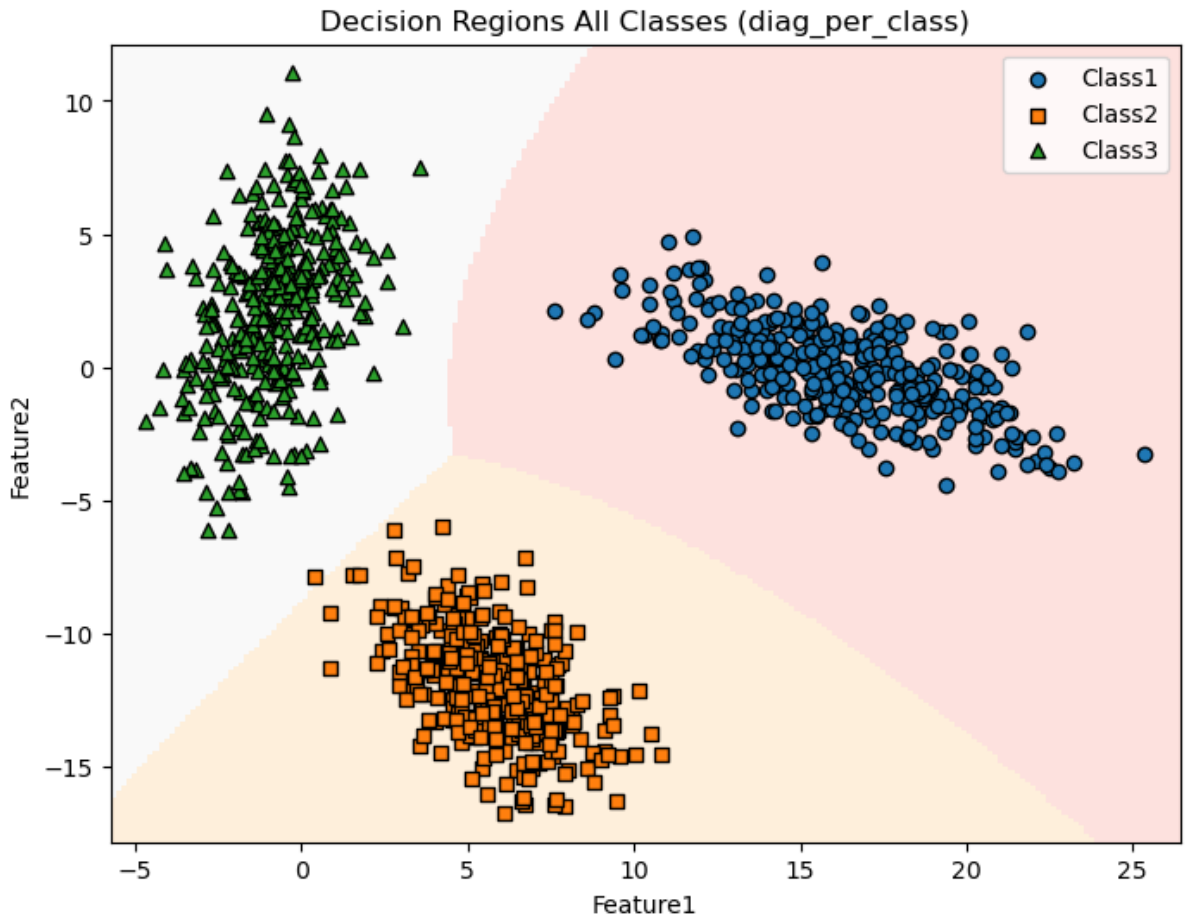
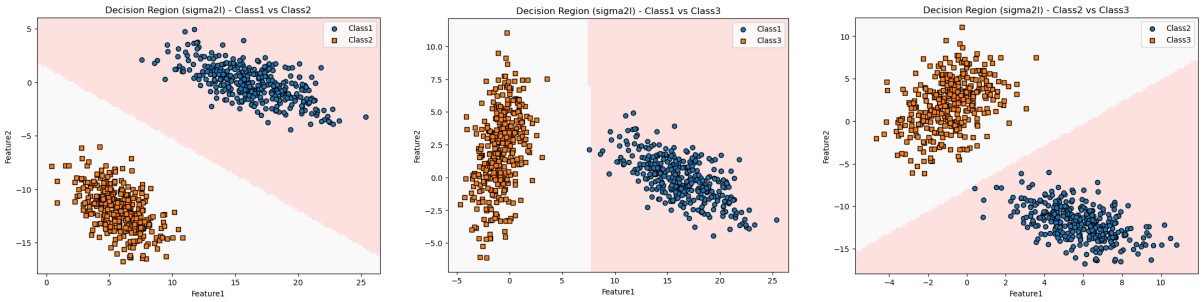


Figure 7: Decision Region Plot (All Classes) - Diagonal Covariance (Per-Class)

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 8: 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)

Table 7: Confusion Matrix for Full Covariance (Per-Class) (Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	150	0	0
Class 2	0	150	0
Class 3	0	0	150

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

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	1.0000	1.0000	150
Class 2	1.0000	1.0000	1.0000	150
Class 3	1.0000	1.0000	1.0000	150
Accuracy		1.0000		
Mean Precision		1.0000		
Mean Recall		1.0000		
Mean F1 Score		1.0000		

Inference: The full per-class covariance model achieved perfect classification (100% accuracy) with no errors, making it the best performer on the linearly separable dataset.

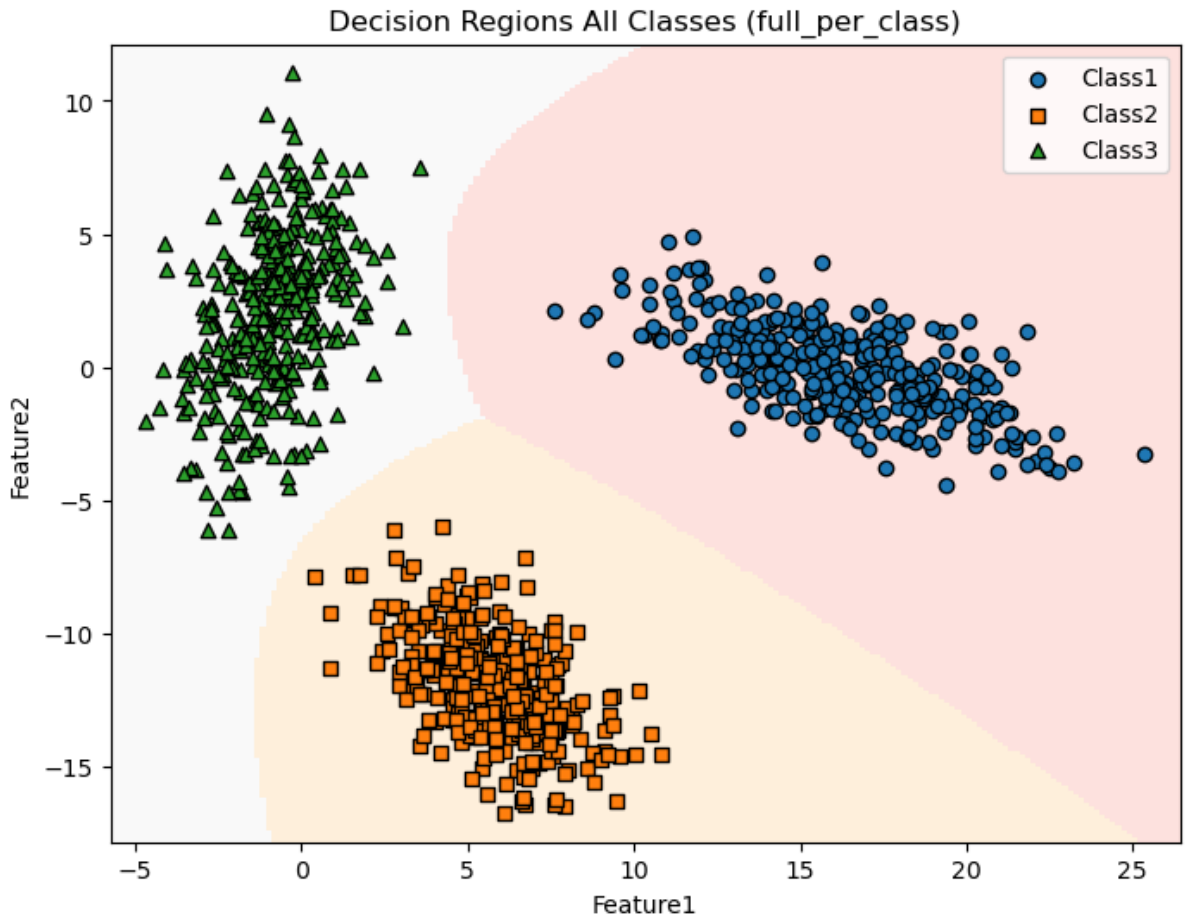
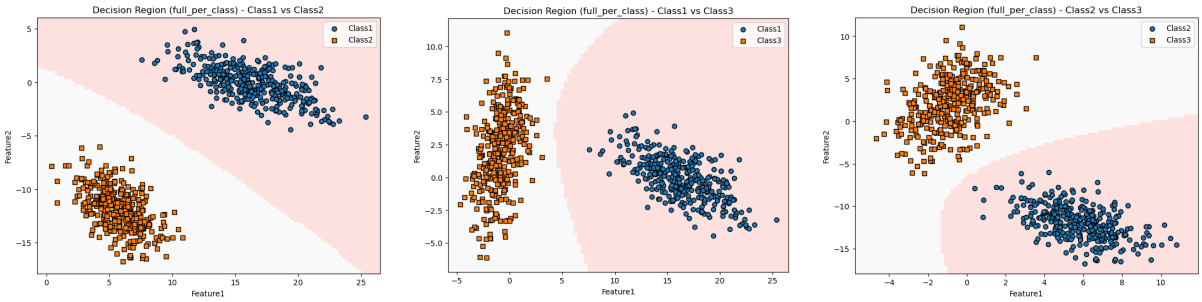


Figure 9: Decision Region Plot (All Classes) - Full Covariance (Per-Class)

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 10: 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

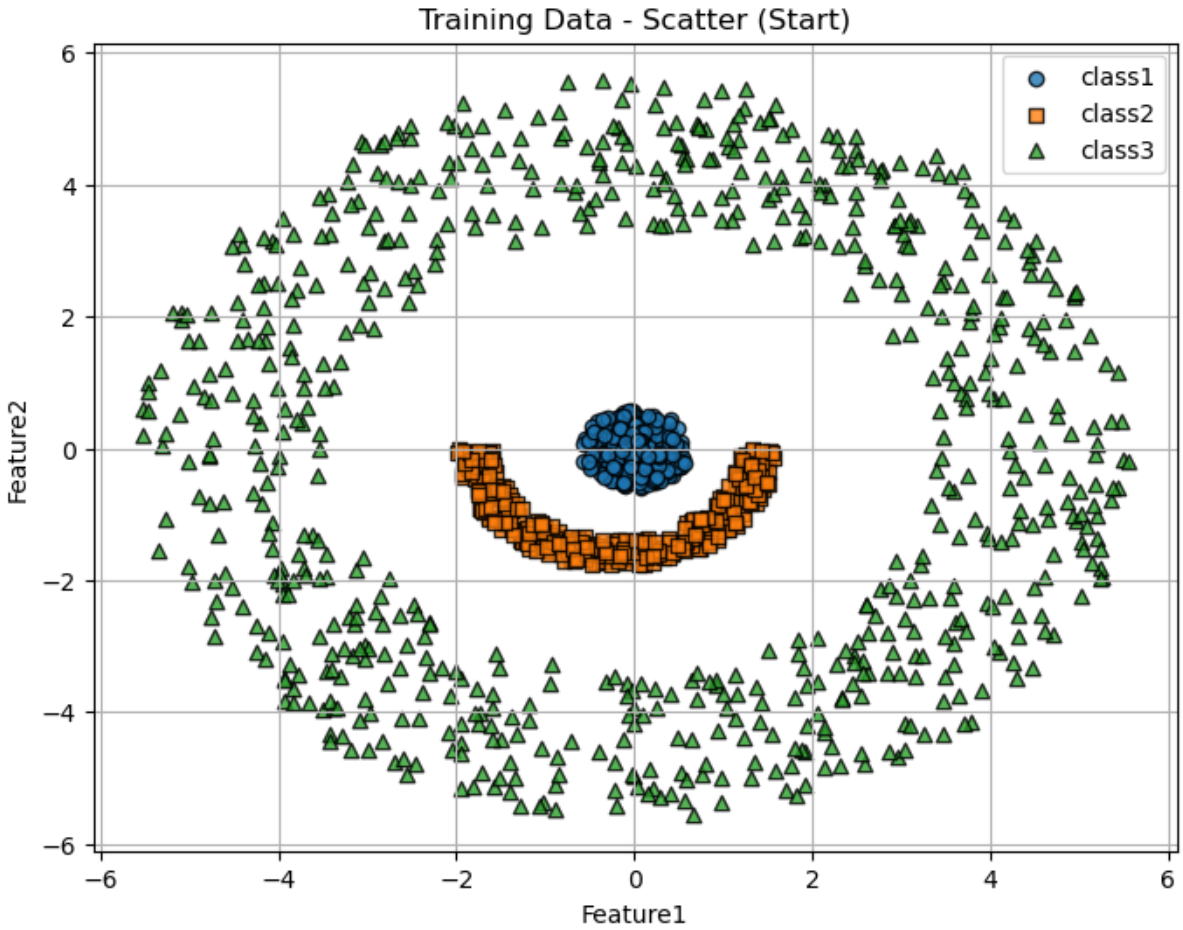


Figure 11: Scatter plot of training data for nonlinear dataset

3.2 Constant Density Contour Plot

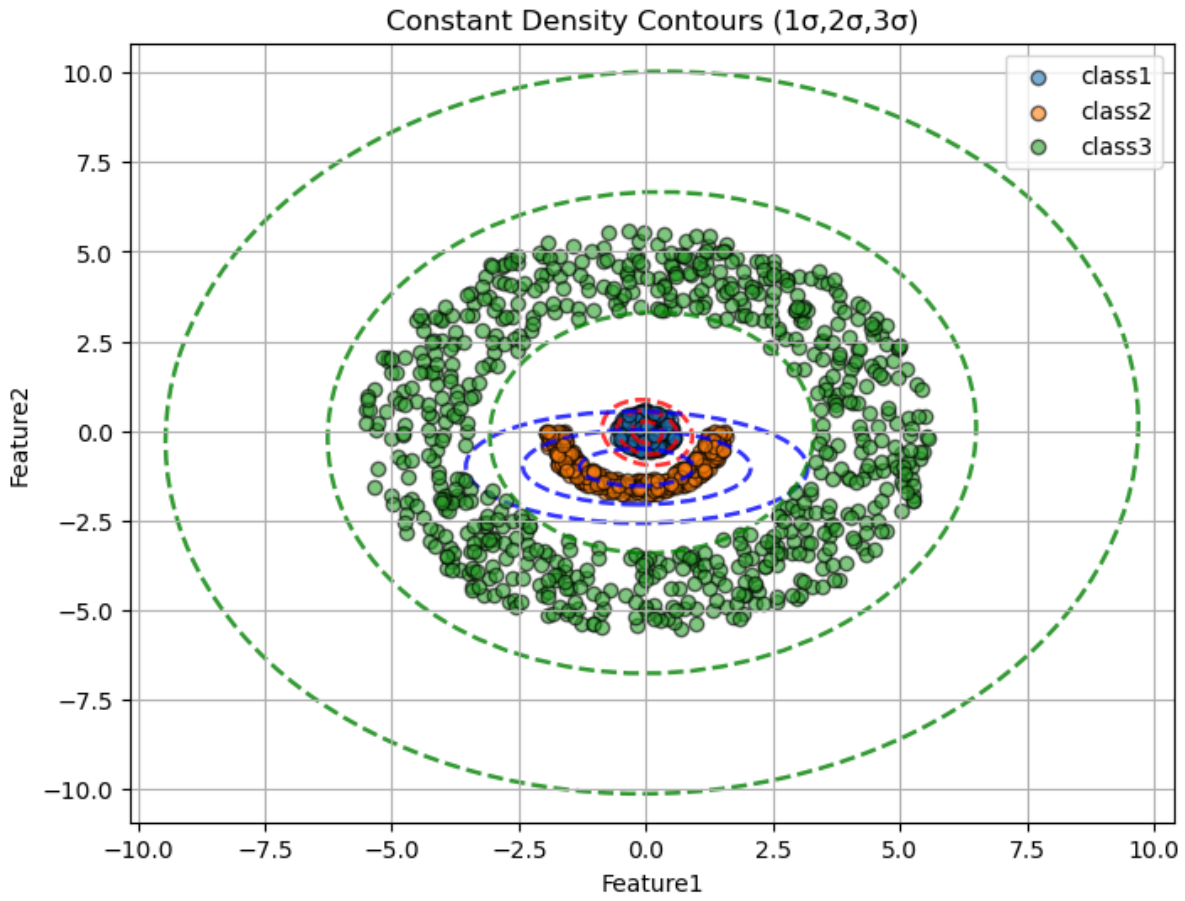


Figure 12: Constant density contours for all classes

3.3 Classifier: Shared $\sigma^2 I$

Table 9: Confusion Matrix for Shared $\sigma^2 I$ (Non-Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	0	0	90
Class 2	0	0	150
Class 3	0	77	223

Table 10: Performance Metrics - Shared $\sigma^2 I$

Class	Precision	Recall	F1-Score	Support
Class 1	0.0000	0.0000	0.0000	90
Class 2	0.0000	0.0000	0.0000	150
Class 3	0.4816	0.7433	0.5845	300
Accuracy		0.4130		
Mean Precision		0.1605		
Mean Recall		0.2478		
Mean F1 Score		0.1948		

Inference: The shared $\sigma^2 I$ model performed very poorly on nonlinear data (accuracy $\approx 41.3\%$), failing to classify Classes 1 and 2 correctly and showing that simple covariance cannot handle complex boundaries.

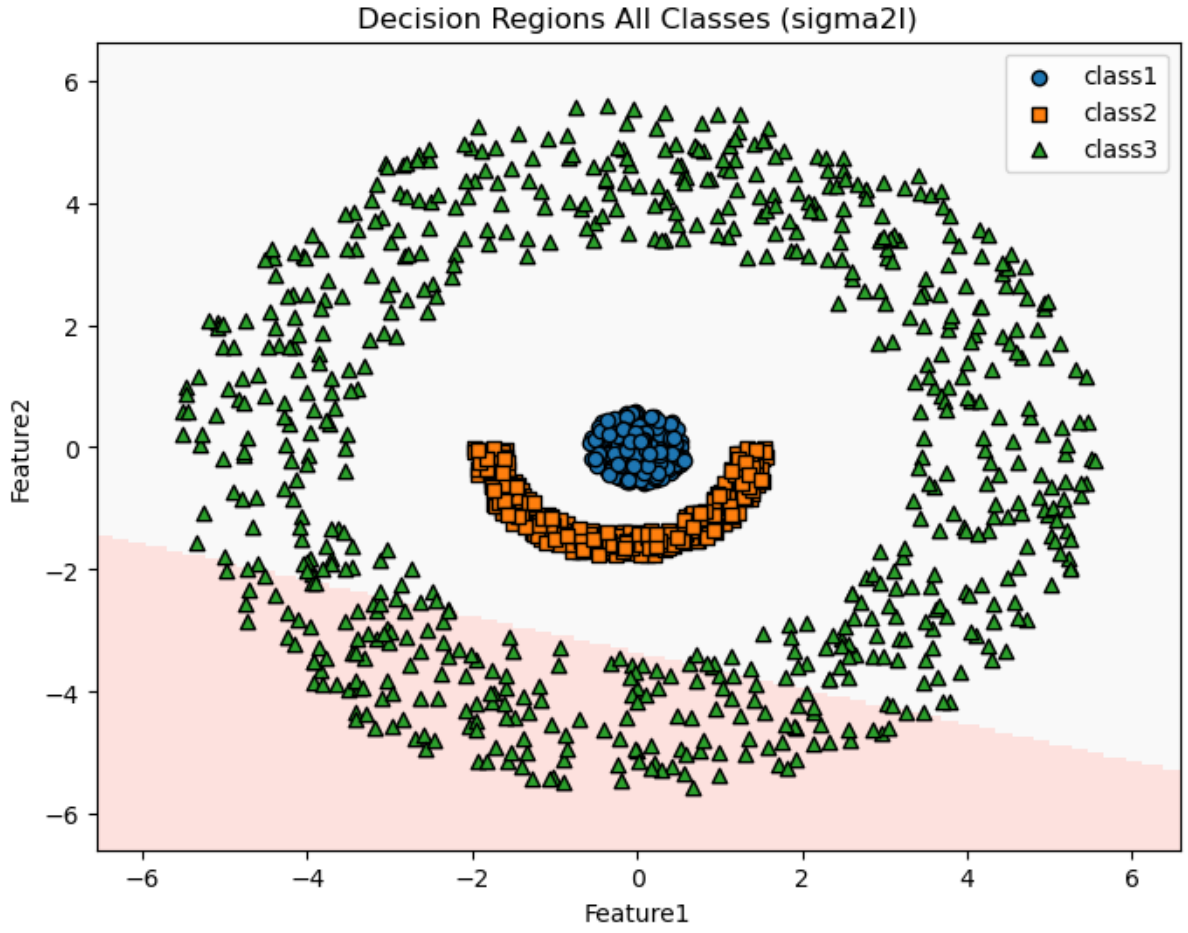
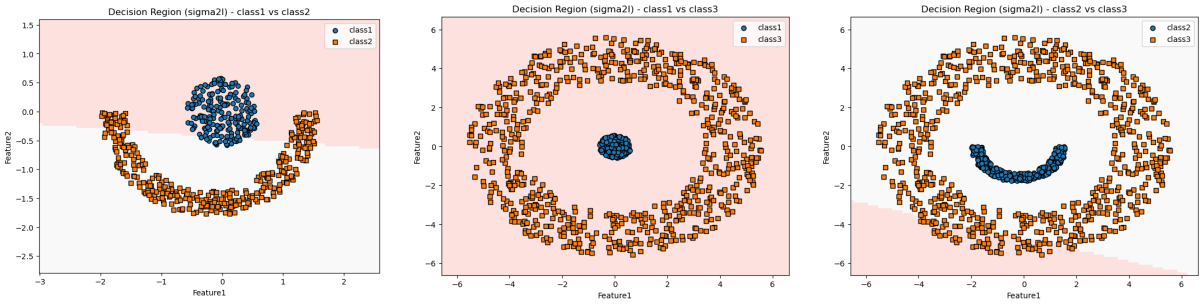


Figure 13: Decision Region Plot (All Classes) - Shared $\sigma^2 I$

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 14: 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

Table 11: Confusion Matrix for Shared Full Covariance Σ (Non-Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	0	0	90
Class 2	0	0	150
Class 3	0	75	225

Table 12: Performance Metrics - Shared Full Covariance

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	0.0000	0.0000	0.0000	90
Class 2	0.0000	0.0000	0.0000	150
Class 3	0.4839	0.7500	0.5882	300
Accuracy		0.4167		
Mean Precision		0.1613		
Mean Recall		0.2500		
Mean F1 Score		0.1961		

Inference: The shared full covariance Σ also failed on nonlinear data (accuracy $\approx$ 41.7%), misclassifying all samples of Classes 1 and 2 while performing only moderately on Class 3.

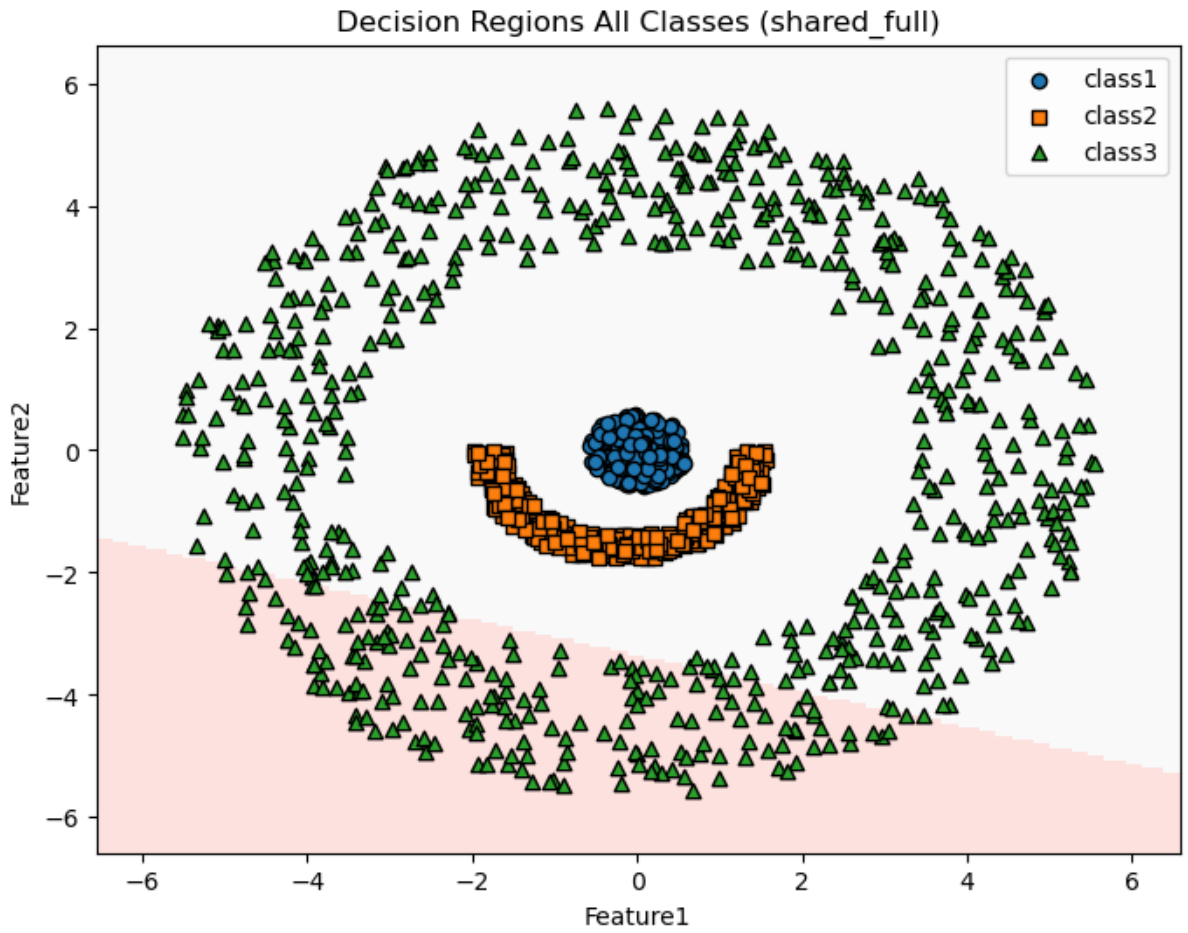
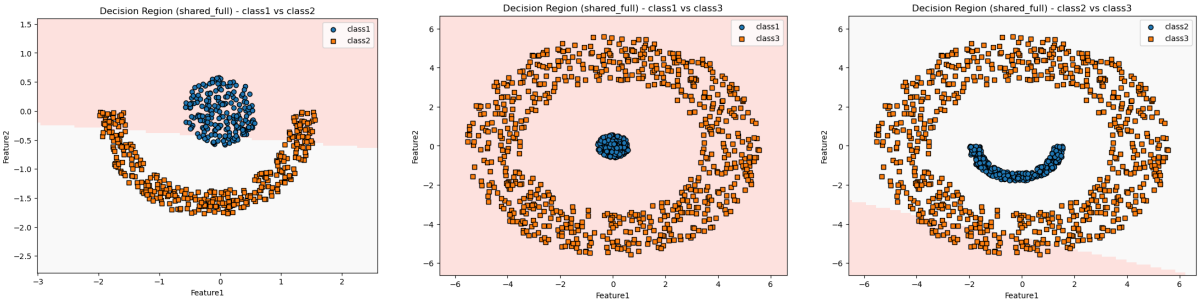


Figure 15: Decision Region Plot (All Classes) - Shared Full Covariance

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

3.5 Classifier: Diagonal Covariance (Per-Class)

Table 13: Confusion Matrix for Diagonal Covariance (Per-Class) (Non-Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	90	0	0
Class 2	0	140	10
Class 3	0	0	300

Table 14: Performance Metrics - Diagonal Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	1.0000	1.0000	90
Class 2	1.0000	0.9333	0.9655	150
Class 3	0.9677	1.0000	0.9836	300
Accuracy		0.9815		
Mean Precision		0.9892		
Mean Recall		0.9778		
Mean F1 Score		0.9830		

Inference: The diagonal per-class covariance model performed very well on nonlinear data with an accuracy of 98.15%, showing high precision and recall across all classes and only minor misclassifications in Class 2.

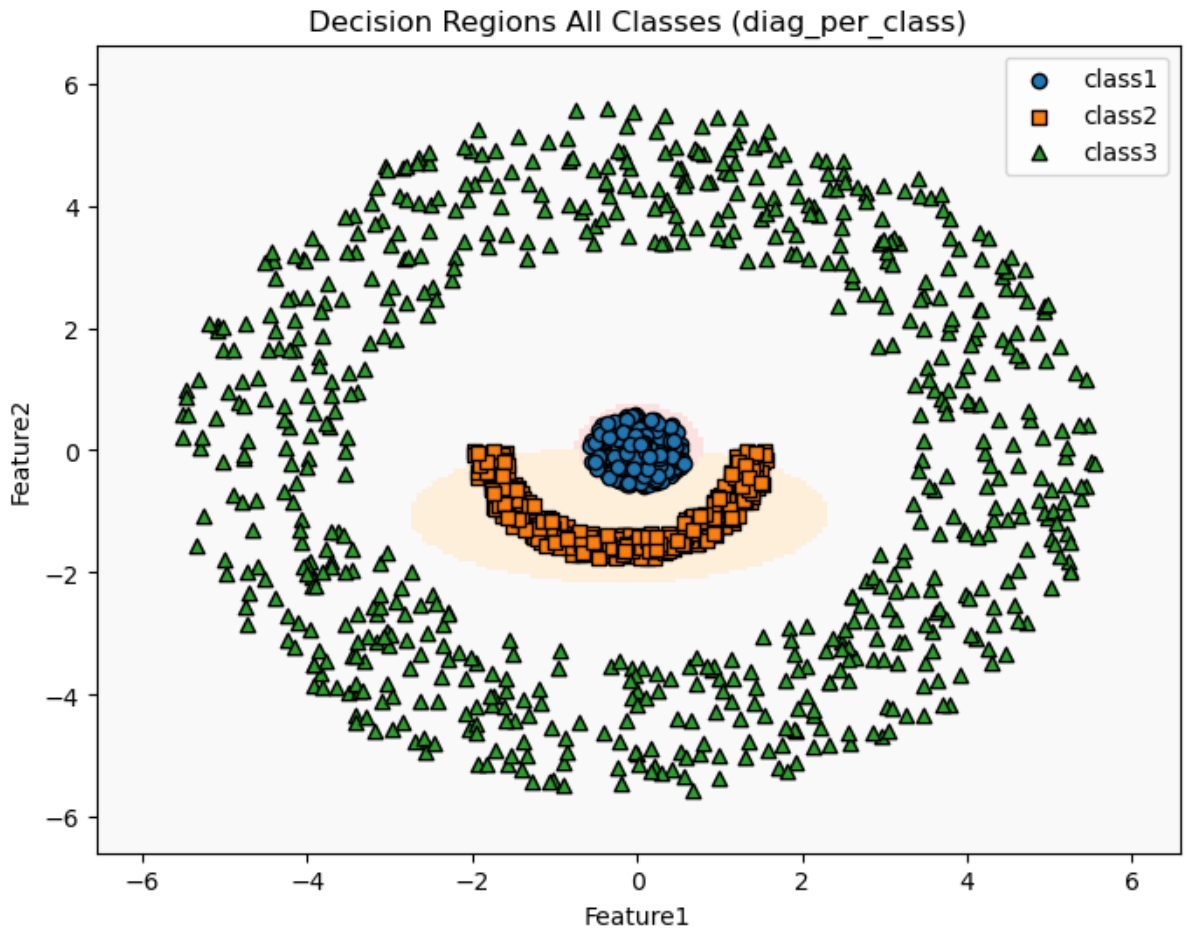
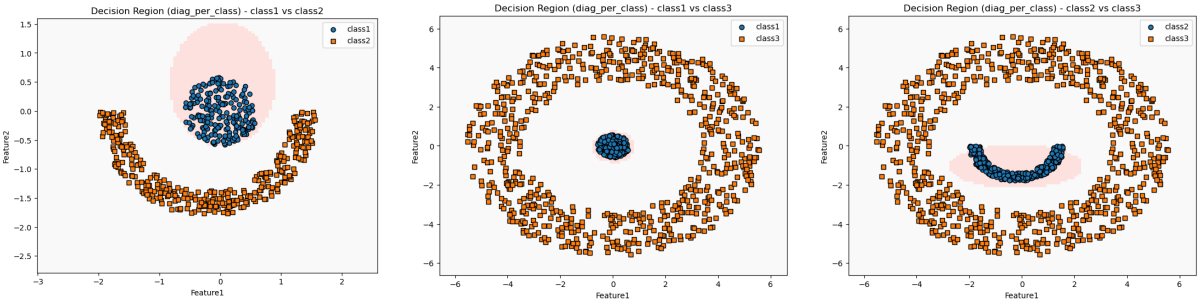


Figure 17: Decision Region Plot (All Classes) - Diagonal Covariance (Per-Class)

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 18: 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)

Table 15: Confusion Matrix for Full Covariance (Per-Class) (Non-Linearly Separable Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	90	0	0
Class 2	0	142	8
Class 3	0	0	300

Table 16: Performance Metrics - Full Covariance (Per-Class)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 1	1.0000	1.0000	1.0000	90
Class 2	1.0000	0.9467	0.9726	150
Class 3	0.9740	1.0000	0.9868	300
Accuracy		0.9852		
Mean Precision		0.9913		
Mean Recall		0.9822		
Mean F1 Score		0.9865		

Inference: The full per-class covariance model achieved the highest accuracy of 98.52% on nonlinear data, with excellent precision and recall, making it the most effective among all covariance approaches.

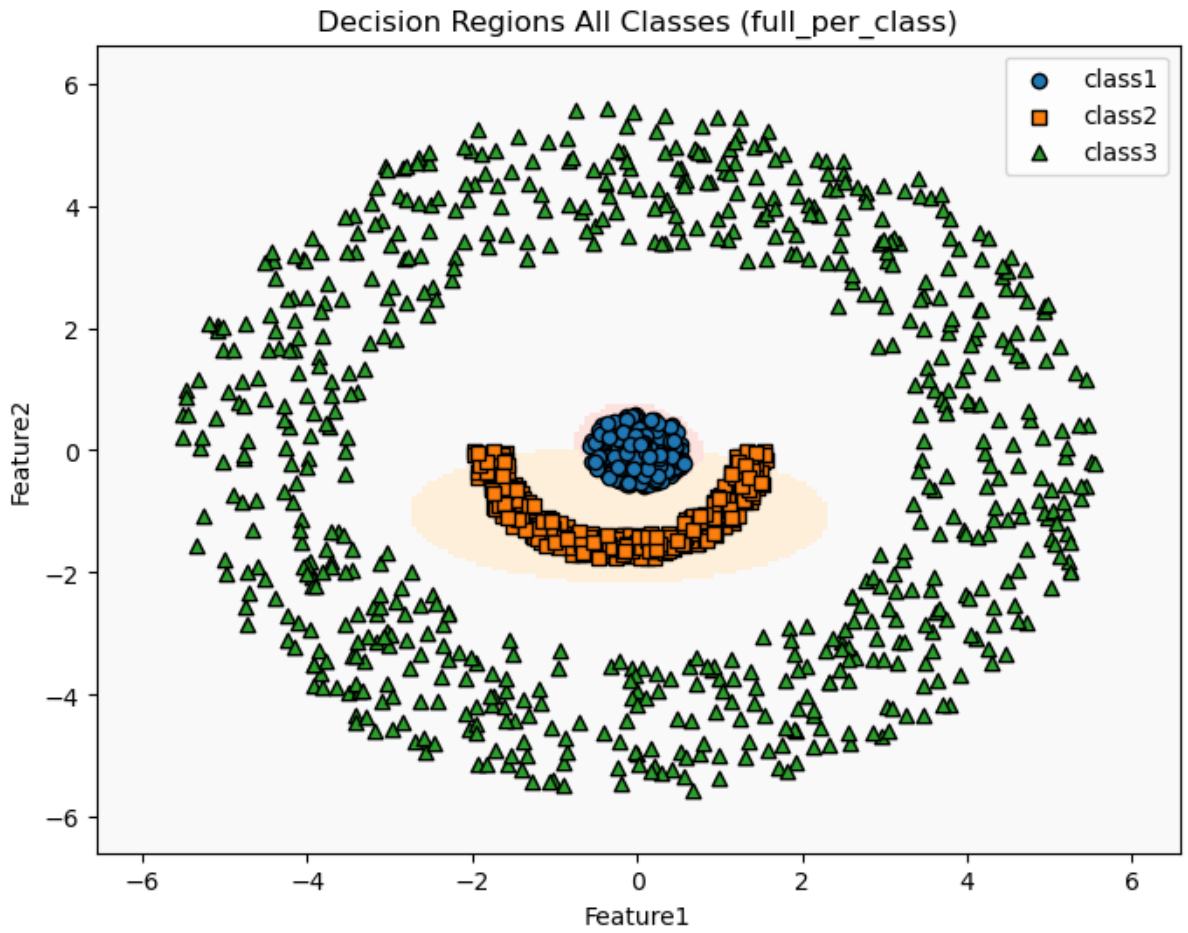
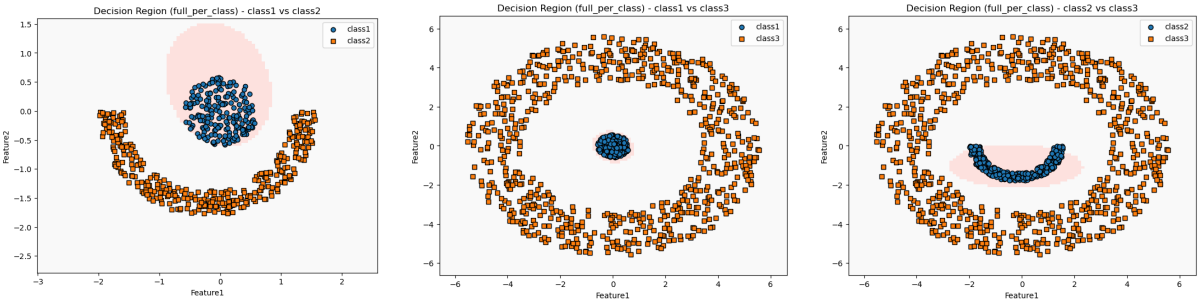


Figure 19: Decision Region Plot (All Classes) - Full Covariance (Per-Class)

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 20: 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 21: Scatter plot of training data for vowel dataset

4.2 Constant Density Contour Plot

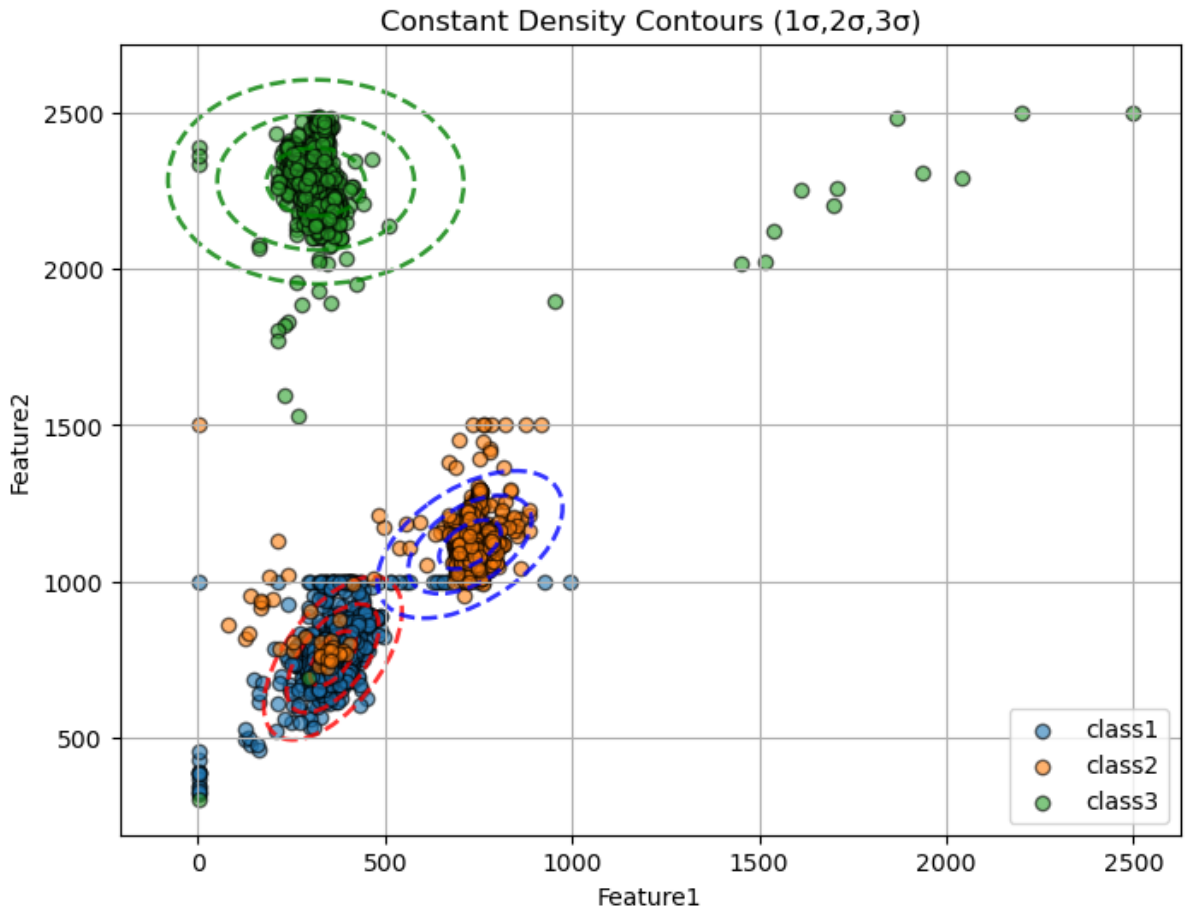


Figure 22: Constant density contours for vowel dataset

4.3 Classifier: Shared $\sigma^2 I$

Table 17: Confusion Matrix for Shared $\sigma^2 I$ (Vowel Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	746	1	0
Class 2	19	631	0
Class 3	1	2	714

Table 18: 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		

Inference: The shared $\sigma^2 I$ classifier performed very well on vowel data with an accuracy of 98.91%, showing strong precision and recall across all classes.

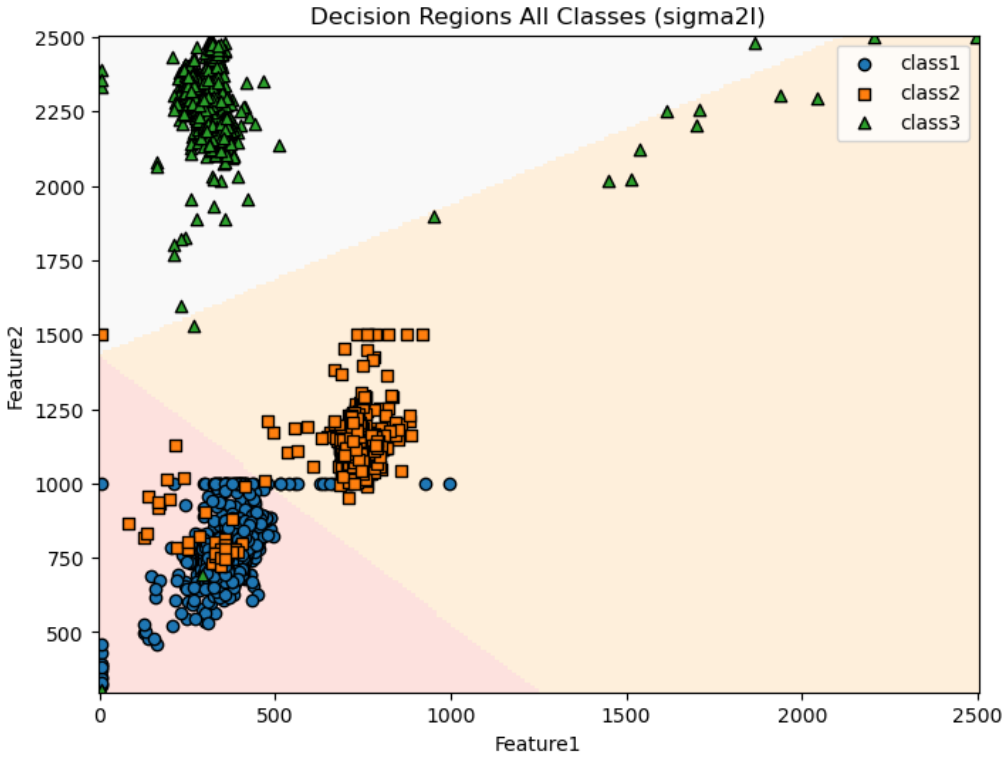
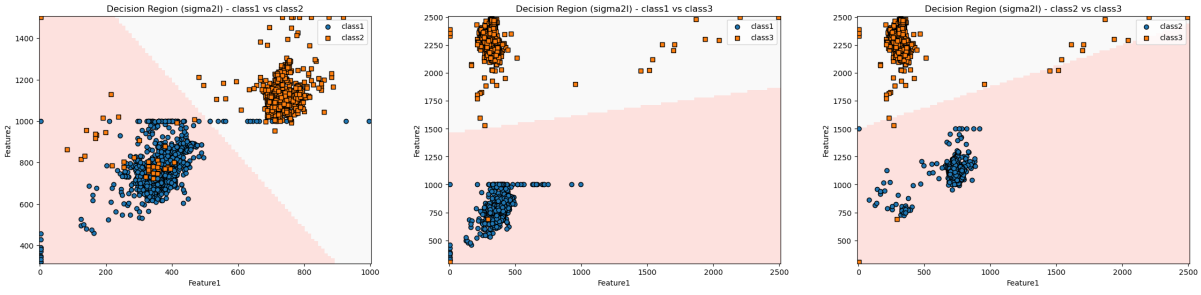


Figure 23: Decision Region Plot (All Classes) - Shared $\sigma^2 I$

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 24: 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

Table 19: Confusion Matrix for Shared Full Covariance Σ (Vowel Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	746	1	0
Class 2	19	631	0
Class 3	1	4	712

Table 20: 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		

Inference: The shared Σ classifier performed very well on vowel data with an accuracy of 98.82%, showing strong precision and recall across all classes.

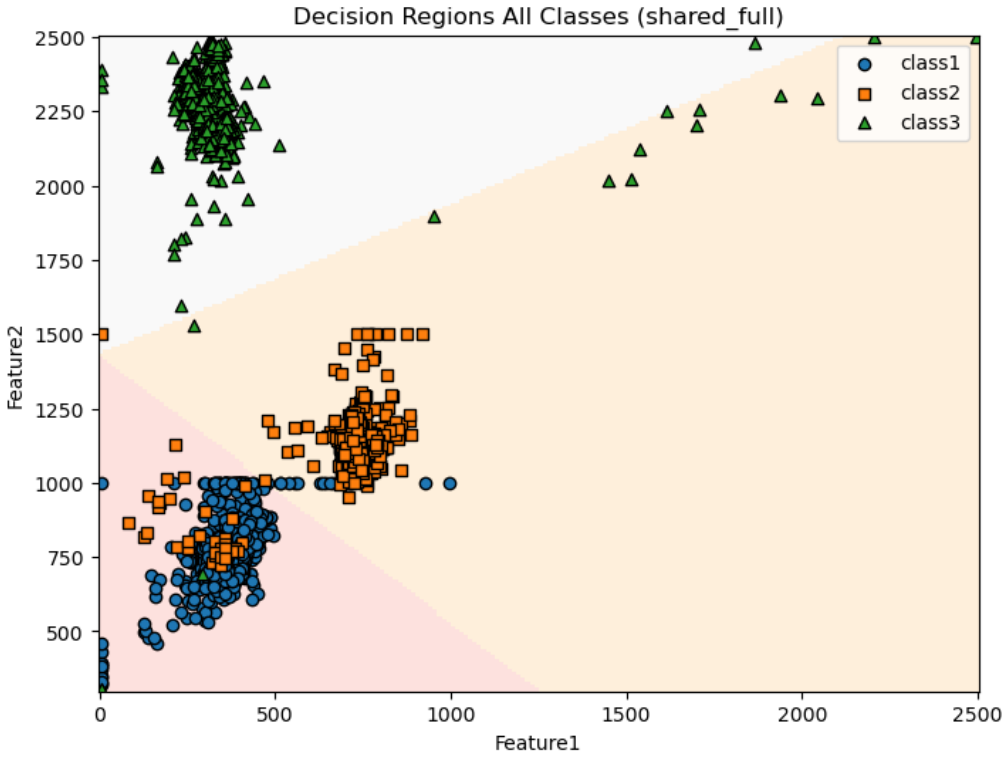
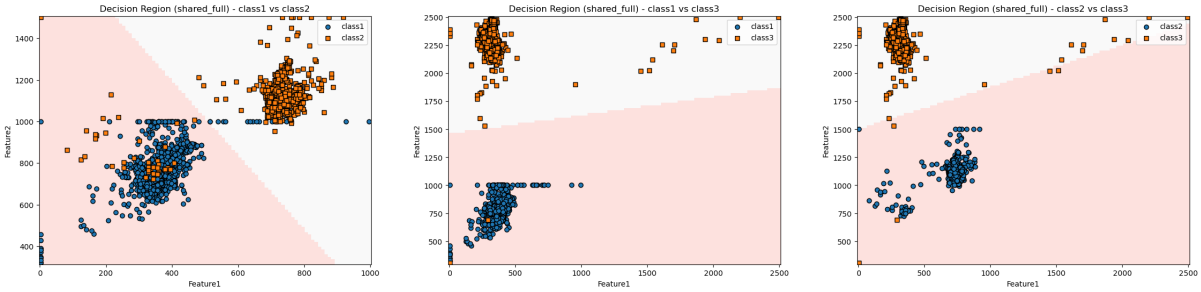


Figure 25: Decision Region Plot (All Classes) - Shared Full Covariance

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

4.5 Classifier: Diagonal Covariance (Per-Class)

Table 21: Confusion Matrix for Diagonal Covariance (Per-Class) (Vowel Data)

Actual \ Predicted	Class 1	Class 2	Class 3
Class 1	746	1	0
Class 2	19	631	0
Class 3	1	1	715

Table 22: 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		

Inference: The diagonal covariance (per-class) classifier performed excellently on vowel data with an accuracy of 98.96%, achieving high precision and recall across all classes.

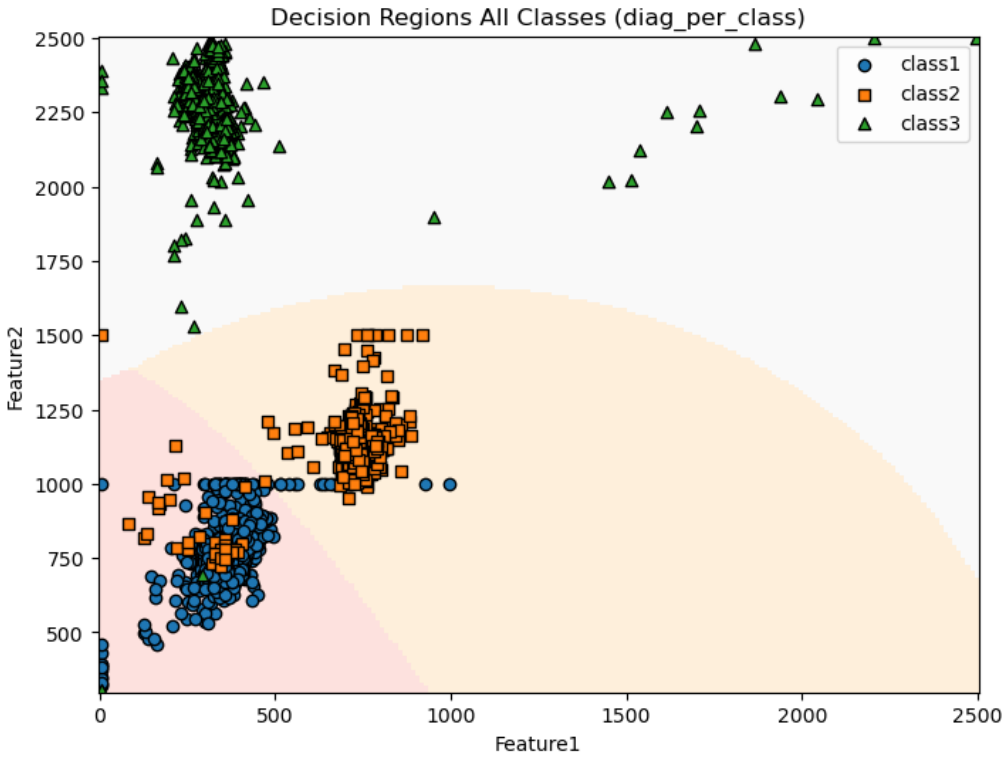
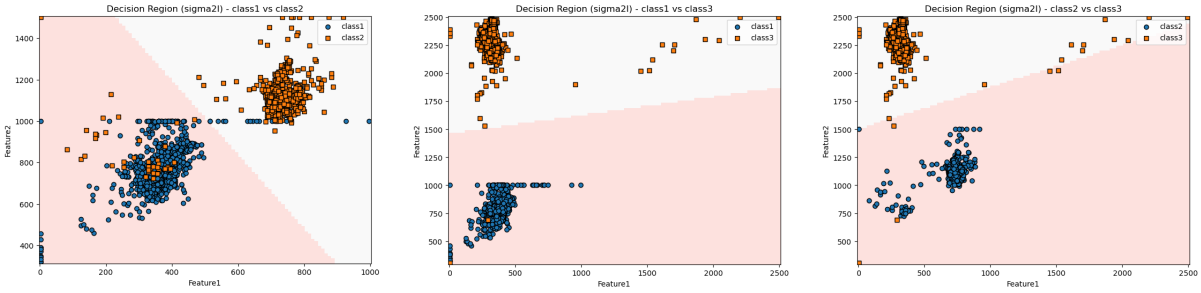


Figure 27: Decision Region Plot (All Classes) - Diagonal Covariance (Per-Class)

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 28: 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)

Table 23: Confusion Matrix for Full Covariance (Per-Class) (Vowel Data)

Actual \ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	745	2	0
Class 2	19	631	0
Class 3	1	1	715

Table 24: 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		

Inference: The full covariance (per-class) classifier performed very well on vowel data with an accuracy of 98.91%, showing consistently high precision and recall across all classes.

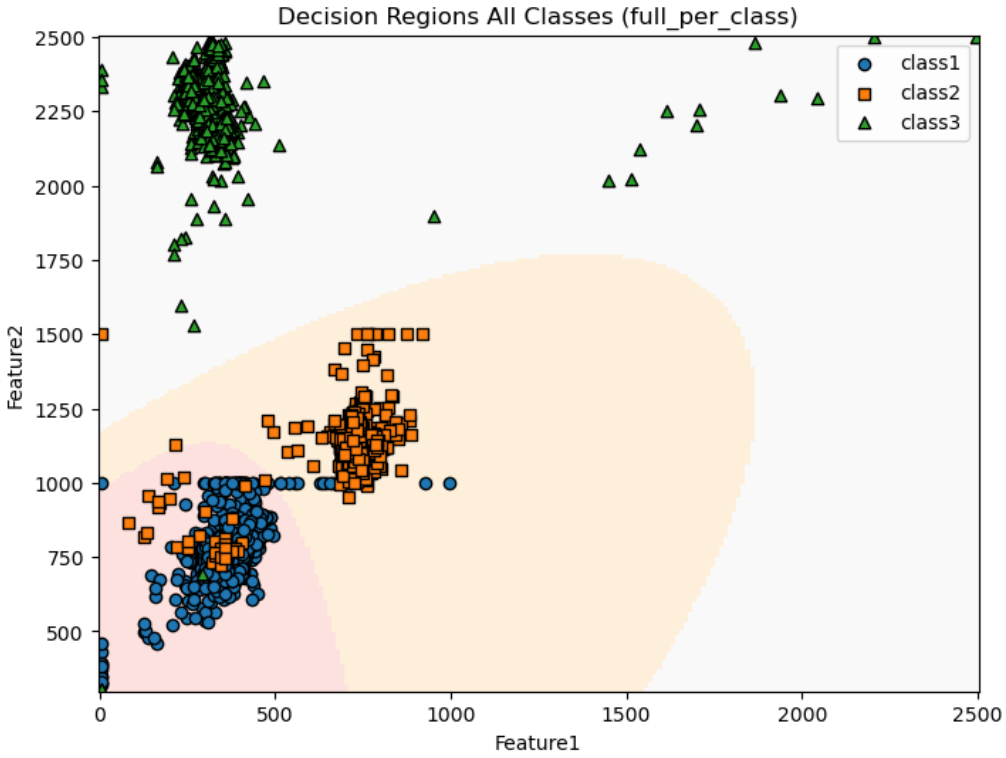
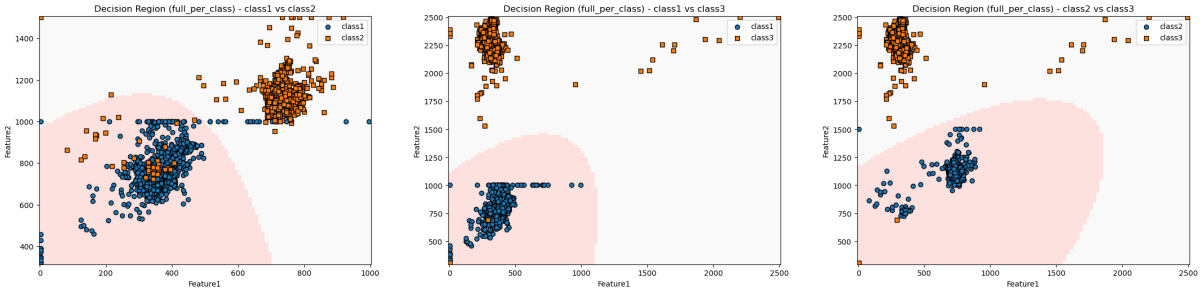


Figure 29: Decision Region Plot (All Classes) - Full Covariance (Per-Class)

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 30: 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 25: 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	0.4130	0.1605	0.2478	0.1948
	shared_full	0.4167	0.1613	0.2500	0.1961
	diag_per_class	0.9815	0.9892	0.9778	0.9830
	full_per_class	0.9852	0.9913	0.9822	0.9865
Dataset 3 (Real-world)	sigma2I	0.9891	0.9897	0.9884	0.9890
	shared_full	0.9882	0.9887	0.9875	0.9880
	diag_per_class	0.9896	0.9902	0.9889	0.9895
	full_per_class	0.9891	0.9897	0.9884	0.9890

6 Observations and Inferences

6.1 Dataset 1 (Linearly separable)

- All covariance types gave very high accuracy and F1-scores (close to 1.0).
- The data is almost perfectly separable.
- Even simple covariance assumptions work well.
- The decision surfaces are smooth and nearly linear.

6.2 Dataset 2 (Nonlinear classes)

- Shared covariance models performed very poorly (F1 $\approx$ 0.19).
- Accuracy improved a lot (F1 $\approx$ 0.98) with per-class diagonal or full covariance.
- The dataset has complex class boundaries.
- Simple shared covariance is not sufficient.

6.3 Dataset 3 (Real-world vowel data)

- All models gave high performance (F1 $\approx$ 0.99).
- Differences between covariance types were very small.
- Vowel features are already well separated across classes.

6.4 Decision surfaces

- For the linear dataset, boundaries are straight and clean.
- For the nonlinear dataset, per-class covariance gave curved and accurate regions.
- For the vowel dataset, decision regions are stable and robust.

6.5 Confusion matrices

- Linear and vowel datasets show almost perfect classification with very few misclassifications.
- Nonlinear dataset shows many misclassifications under shared covariance.
- Performance improves significantly with per-class covariance.

6.6 Covariance Comparison

Table 26: Comparison of Mean F1 Scores Across Covariance Types

Covariance Type	**Linear Dataset**	**Nonlinear Dataset**	**Vowel Dataset**
Shared $\sigma^2 I$	0.9956	0.2478	0.9884
Shared Σ	0.9978	0.2500	0.9875
Diagonal Per Class	0.9956	0.9778	0.9889
Full Per Class	1.0000	0.9822	0.9884

6.6.1 Observations

- For the **linear dataset**, all covariance types give almost perfect performance (F1 close to 1.0).
- For the **nonlinear dataset**, shared covariance performs very poorly (F1 $\approx$ 0.25), while per-class covariance (diagonal and full) gives excellent results (F1 $\approx$ 0.98).
- For the **vowel dataset**, performance is consistently high (F1 $\approx$ 0.99) across all covariance types.

6.6.2 Inference

- Shared covariance ($\sigma^2 I$ or Σ) is the **odd one out**, as it fails on nonlinear data.
- Per-class covariance (diagonal or full) is more flexible and reliable across all datasets.
- In general, when class boundaries are complex, per-class covariance is necessary for good classification.

7 Conclusion

- For the **linear dataset**, all covariance assumptions work well, as the classes are almost perfectly separable.
- For the **nonlinear dataset**, shared covariance models fail, but per-class covariance (diagonal or full) gives strong performance.
- For the **vowel dataset**, all covariance types perform consistently well since the classes are naturally well separated.
- Overall, **per-class covariance** is the most reliable approach, while shared covariance is the weak option and the odd one out.