ಭಾರತೀಯ ತಂತ್ರಜ್ಞಾನ ಸಂಸ್ಥೆ ಧಾರವಾಡ
भारतीय प्रौद्योगिकी संस्थान धारवाड
INDIAN INSTITUTE OF TECHNOLOGY DHARWAD

1-Name : **Mridul Chandrawanshi - (CS24MT002)**
2-Name : **Kapil Kumar Bhardwaj - (CS24MT012)**
3-Name : **Manish Bisht - (CS24MT022)**
Course : **Statistical Pattern Recognition**

Assignment: **1**
GRP.NO: **06**

1 Introduction

This report presents an in-depth analysis of classification tasks using a Bayesian classifier on three distinct datasets. Each dataset was split into training and testing sets using a 70/30 ratio to evaluate the performance of the classifier under various covariance assumptions. The primary objective is to understand how different covariance structures impact classification performance. The results are analyzed through confusion matrices, classification accuracy, precision, recall, F-measure, constant density contour plots, and decision region plots.

The datasets used in this analysis are as follows:

- **Dataset 1: Linearly Separable Data**
  This dataset consists of 2-dimensional data points distributed across three classes, with each class containing 500 data points. The classes are linearly separable, meaning that they can be perfectly separated using a linear boundary. The primary aim is to evaluate how well the Bayesian classifier performs with different covariance assumptions when the data is inherently linearly separable.

- **Dataset 2: Nonlinearly Separable Data**
  This dataset includes 2-dimensional data with either two or three classes that are nonlinearly separable. The number of examples in each class and their specific order are detailed in the dataset files. The challenge here is to assess the classifier's ability to handle complex, spiral-like patterns where linear separation is not feasible. The impact of different covariance assumptions on the classifier's ability to model nonlinear decision boundaries is a key focus.

- **Dataset 3: Real-World Vowel Data**
  This dataset comprises 2-dimensional data representing the formant frequencies F1 and F2 for vowel utterances. It includes three distinct classes corresponding to different vowel sounds. This real-world dataset provides an opportunity to evaluate the classifier's performance on actual speech data. The analysis includes exploring how well the classifier distinguishes between vowel classes using various covariance assumptions.

Each dataset presents unique characteristics and challenges, which provide a comprehensive assessment of the Bayesian classifier's performance across different scenarios. The classifier's effectiveness is evaluated based on its accuracy, precision, recall, and F-measure, alongside visualizations such as density contours and decision regions to provide a holistic understanding of its performance.

Table of Contents

Contents

2 Dataset 1: Linearly Separable Data

Results for Dataset 1: Linearly Separable Data

2.1 Confusion Matrix and Performance Metrics

The following table summarizes the confusion matrix and performance metrics for the Bayesian classifier applied to the linearly separable dataset across various covariance matrix assumptions.

2.2 Constant Density Contour Plots

Explanation: The plots illustrate the constant density contours for the assumption of equal variance across all classes, modeled as $\sigma^2 I$. The left plot shows circular contours reflecting uniform variance in all directions, effectively separating linearly separable classes. The right plot reaffirms this by displaying the same circular symmetry, demonstrating the model's robustness in handling simple linearly separable data.

Explanation: The left plot depicts the constant density contours under the full covariance matrix assumption where the covariance matrix is identical for all classes. The elliptical contours reflect the spread and orientation of the data more flexibly compared to $\sigma^2 I$. The right plot shows $\sigma^2 I$ contours, reinforcing the effectiveness of this simpler covariance model for linearly separable data.

Explanation: The left plot illustrates the constant density contours under the diagonal covariance matrix assumption where the covariance matrix differs for each class. This results in elliptical contours that adapt to different variances along the axes. The right plot shows the full covariance matrix (Different) assumption, where each class has a unique covariance matrix, resulting in highly adaptive elliptical contours that capture complex data distributions more effectively.

2.3 Decision Region Plots

Explanation: The left plot displays the decision regions for the $\sigma^2 I$ assumption, where linear boundaries effectively separate the classes due to the linear separability of the data. The right plot illustrates the decision regions with a diagonal covariance matrix (Different), where the boundaries adjust to different variances for each class, offering more flexibility in separating classes with varying spreads.

Explanation: The left plot shows the decision regions with a full covariance matrix (Same), where curved boundaries reflect the data's spread and orientation more accurately. The right plot presents decision regions under the full covariance matrix (Different) assumption, highlighting the model's capability to adapt to intricate class distributions with highly flexible decision boundaries.

2.4 Inferences

For the linearly separable dataset, the Bayesian classifier achieved perfect accuracy across all covariance assumptions, demonstrating 100% precision, recall, and F-measure for each class. The confusion matrix shows no misclassifications.

The decision regions and constant density contour plots reveal that all covariance assumptions produce clear class separations due to the dataset's linear separability. While all models perform equally well in this simple scenario, the more complex covariance assumptions, such as the full covariance matrix, offer additional flexibility and adaptability for more complex datasets. These findings underscore the effectiveness of Bayesian classifiers in handling linearly

Covariance Type	Accuracy (%)	Precision	Recall	F-Measure
$\sigma^2 I$	100.00	1.0, 1.0, 1.0	1.0, 1.0, 1.0	1.0, 1.0, 1.0
Full Covariance Matrix (Same)	100.00	1.0, 1.0, 1.0	1.0, 1.0, 1.0	1.0, 1.0, 1.0
Diagonal Covariance Matrix (Different)	100.00	1.0, 1.0, 1.0	1.0, 1.0, 1.0	1.0, 1.0, 1.0
Full Covariance Matrix (Different)	100.00	1.0, 1.0, 1.0	1.0, 1.0, 1.0	1.0, 1.0, 1.0

Table 1: Confusion matrix and performance metrics for Dataset 1

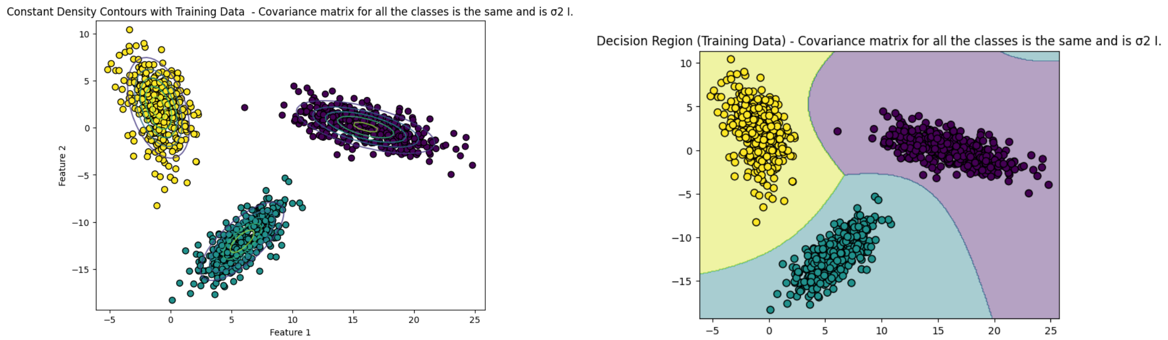

Figure 1: Constant density contour plots for Dataset 1 with $\sigma^2 I$ (Same Variance). Left: $\sigma^2 I$ assumption showing circular contours. Right: $\sigma^2 I$ further visualization with uniform variance.

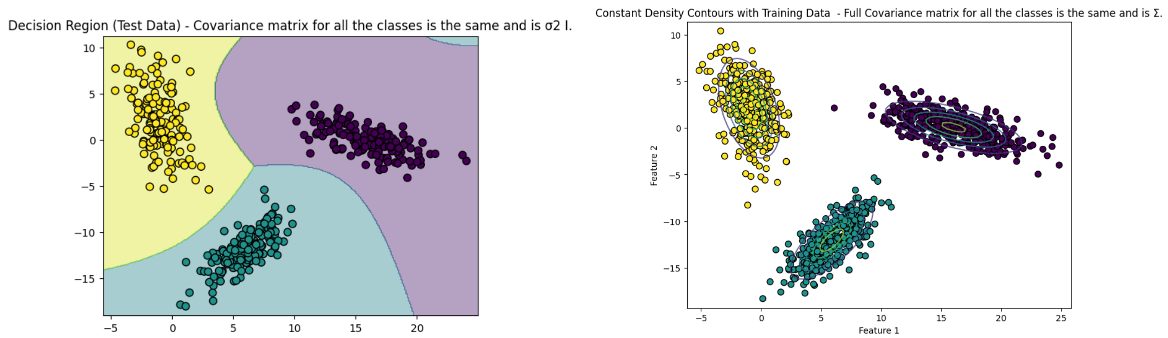

Figure 2: Constant density contour plots for Dataset 1. Left: Full covariance matrix (Same) showing elliptical contours capturing class spread and orientation. Right: $\sigma^2 I$ assumption showing contours with effective class separation.

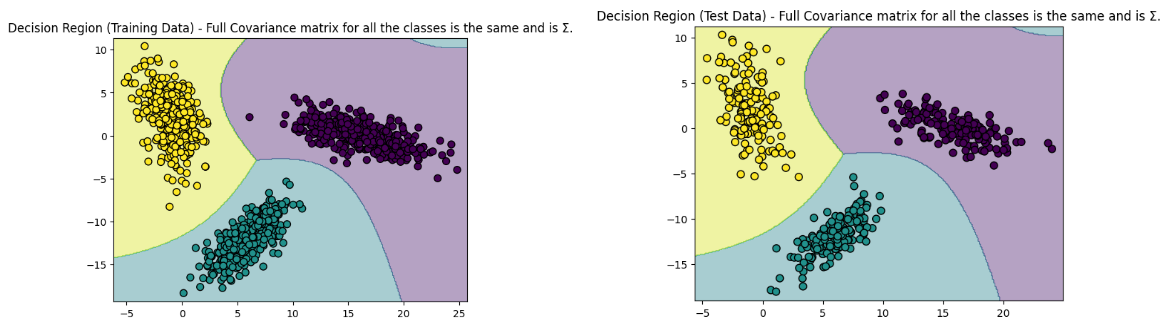

Figure 3: Constant density contour plots for Dataset 1. Left: Diagonal covariance matrix (Different) showing class-specific elliptical contours. Right: Full covariance matrix (Different) demonstrating adaptive contours for each class.

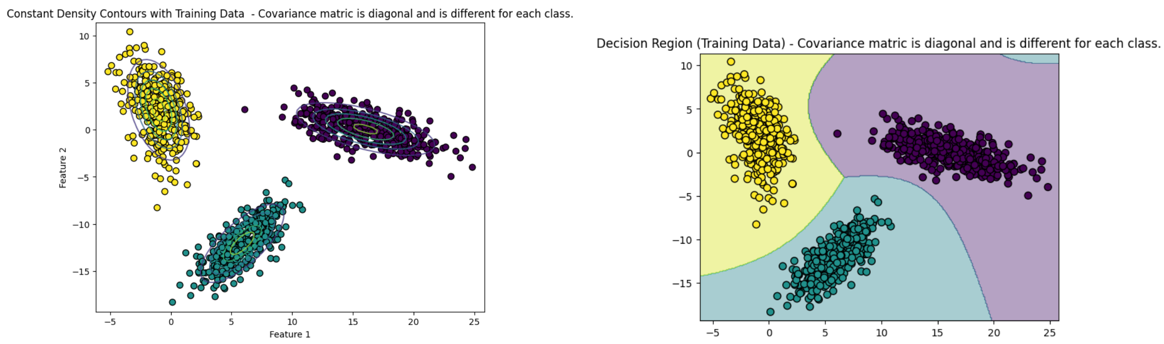

Figure 4: Decision region plots for Dataset 1. Left: $\sigma^2 I$ (Same Variance) showing linear decision boundaries. Right: Diagonal covariance matrix (Different) showing flexible decision regions accommodating class-specific variances.

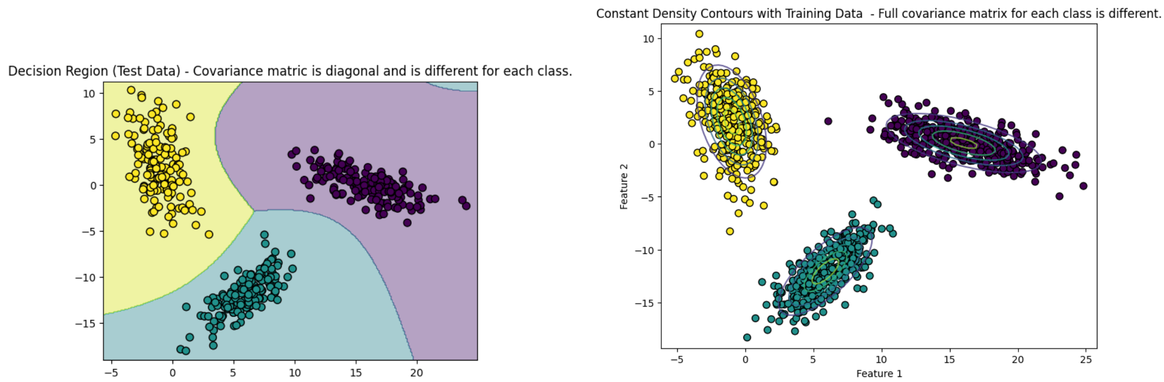

Figure 5: Decision region plots for Dataset 1. Left: Full covariance matrix (Same) showing curved decision boundaries capturing class-specific data structures. Right: Full covariance matrix (Different) demonstrating highly adaptive decision regions for complex class distributions.

separable data and highlight the potential benefits of more sophisticated covariance models for more intricate data distributions.

3 Dataset 2: Nonlinearly Separable Data

3.1 Classifier Metrics for Dataset 2

Classifier with $\sigma^2 I$

Confusion Matrix:

$$\begin{bmatrix} 83 & 0 & 67 \\ 0 & 15 & 135 \\ 75 & 79 & 57 \end{bmatrix}$$

Accuracy: 30.33%
Precision: [0.5253, 0.1596, 0.2201], **Mean Precision:** 0.3017
Recall: [0.5533, 0.1, 0.2701], **Mean Recall:** 0.3078
F-Measure: [0.5390, 0.1230, 0.2426], **Mean F-Measure:** 0.3015

Explanation: The decision region plots for $\sigma^2 I$ and the diagonal covariance matrix reveal the limitations of linear decision boundaries in modeling nonlinearly separable data. Both models exhibit linear decision boundaries that do not capture the complex patterns inherent in the dataset, resulting in poor classification performance.

3.2 Classifier with Full Covariance Matrix Σ

Confusion Matrix:

$$\begin{bmatrix} 79 & 0 & 71 \\ 0 & 13 & 137 \\ 74 & 80 & 57 \end{bmatrix}$$

Accuracy: 29.16%
Precision: [0.5163, 0.1398, 0.2151], **Mean Precision:** 0.2904
Recall: [0.5267, 0.0867, 0.2701], **Mean Recall:** 0.2945
F-Measure: [0.5215, 0.1070, 0.2395], **Mean F-Measure:** 0.2893

Explanation: The decision region plots with a full covariance matrix Σ and a full covariance matrix per class show an improvement in modeling feature correlations and class-specific contours. However, these models still struggle with capturing the nonlinear patterns in the dataset, as evidenced by the limited improvements in classification accuracy.

Classifier with Full Covariance Matrix per Class

Confusion Matrix:

$$\begin{bmatrix} 75 & 0 & 75 \\ 0 & 13 & 137 \\ 76 & 80 & 55 \end{bmatrix}$$

Accuracy: 29.33%
Precision: [0.4974, 0.1394, 0.2135], **Mean Precision:** 0.2834
Recall: [0.5, 0.0867, 0.2642], **Mean Recall:** 0.2836
F-Measure: [0.4987, 0.1067, 0.2351], **Mean F-Measure:** 0.2802

Explanation: The contour plots with full covariance matrices per class and $\sigma^2 I$ illustrate the difficulties in capturing the nonlinear relationships within the data. While the full covariance matrices per class provide some flexibility, both methods struggle to fully model the dataset's complexity.

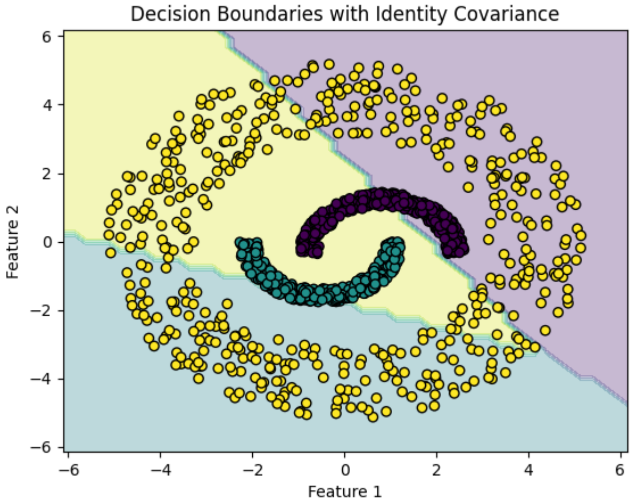

Figure 6: Decision Region plot with $\sigma^2 I$. The linear decision boundaries indicate that this model struggles to capture the nonlinearity of the data, resulting in low accuracy.

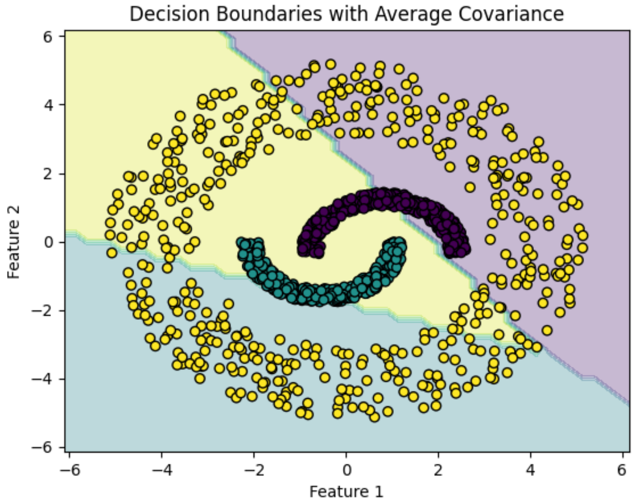

Figure 7: Decision Region plot with Diagonal Covariance Matrix. The linear boundaries, though with variable spread along the axes, still fail to effectively separate the nonlinearly separable classes.

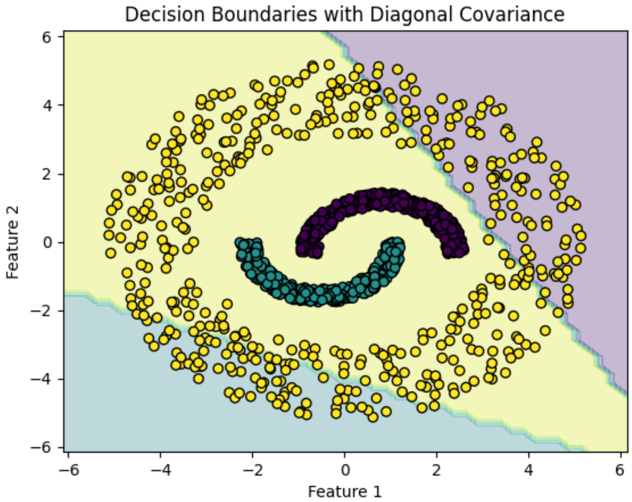

Figure 8: Decision Region plot with Full Covariance Matrix Σ. The curved decision boundaries attempt to capture feature correlations but still fail to address the dataset's nonlinearity.

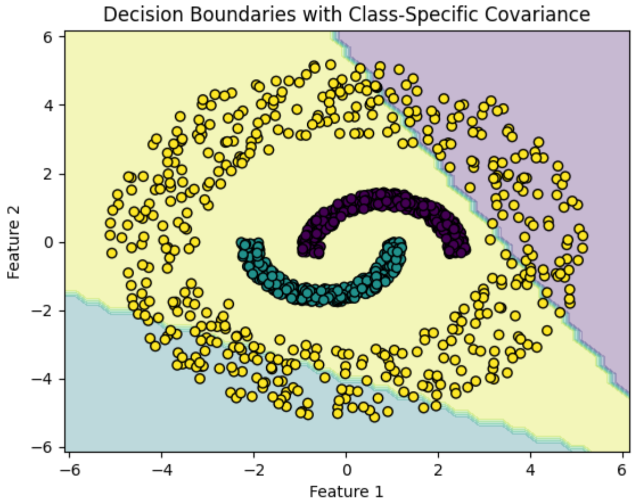

Figure 9: Decision Region plot with Full Covariance Matrix per Class. This approach introduces more flexibility with class-specific contours but still does not fully capture the complex patterns in the data.

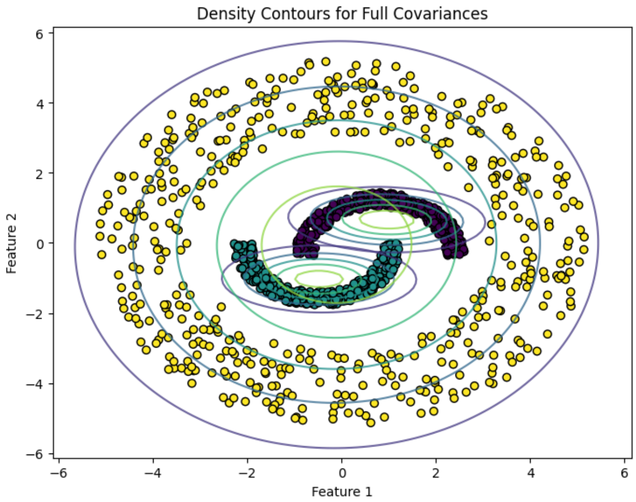

Figure 10: Contour Plot with Full Covariance Matrices per Class. The ellipses adjust to class-specific distributions but do not fully address the nonlinearity of the dataset.

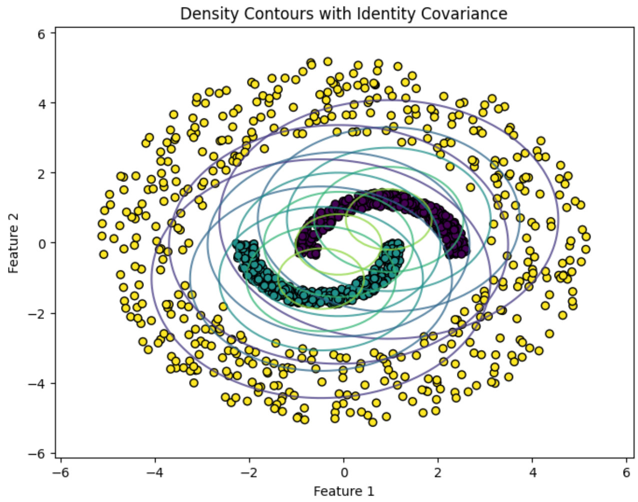

Figure 11: Contour Plot with $\sigma^2 I$. The constant variance contours fail to represent the complex structure of the data.

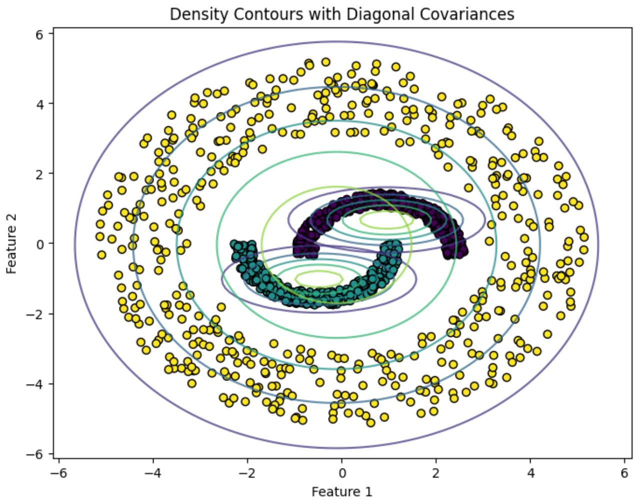

Figure 12: Contour Plot with Diagonal Covariance Matrix. Variable contours along the axes are visible, but the model still fails to capture the dataset's nonlinearity.

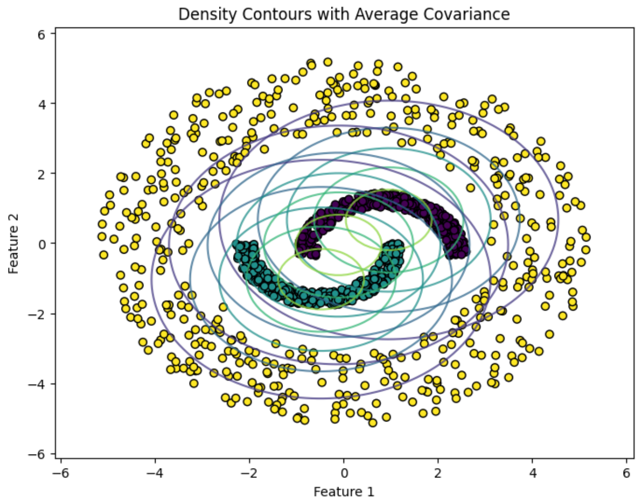

Figure 13: Contour Plot with Full Covariance Matrix. The complex contours attempt to capture class distributions but still struggle with the spiral-like structure of the dataset.

Explanation: The final contour plots for diagonal and full covariance matrices offer additional insights. The diagonal covariance matrix shows variable contours, while the full covariance matrix provides more complex contours. Despite these advancements, both methods face significant challenges in handling the nonlinearity of the dataset.

Summary

For Dataset 2, the classifiers exhibited low performance across various covariance assumptions. Models using $\sigma^2 I$ and full covariance matrix Σ struggled with nonlinearly separable data, resulting in low accuracy and high misclassification rates. Even with full covariance matrices per class, the classifier had difficulty effectively capturing the complex decision boundaries required. These results highlight the limitations of linear and simple covariance models for nonlinearly separable data, suggesting the need for more advanced or nonlinear methods to improve classification performance.

4 Dataset 3: Real-world Vowel Data

4.1 Classifier Metrics for Dataset 3

The real-world vowel dataset was analyzed using various covariance matrix assumptions to understand the impact on classification performance. The following sections detail the performance metrics, confusion matrices, and visualizations for each covariance matrix type.

4.1.1 Covariance Matrix: Same Diagonal ($\Sigma_i = \sigma^2 I$)

For the same diagonal covariance matrix assumption, where each class is modeled with an identical diagonal covariance matrix, the classifier's performance metrics are summarized in Table 2. This assumption simplifies the model by assuming that features are uncorrelated within each class and that each class shares the same variance across features.

Covariance Type	Accuracy (%)	Precision	Recall	F-Measure
Same Diagonal	92.7	0.8600, 0.9400, 0.9900	0.9500, 0.8300, 0.9900	0.9000, 0.8800, 0.9900

Table 2: Performance Metrics for Same Diagonal Covariance Matrix

Confusion Matrix:

$$\begin{bmatrix} 678 & 38 & 1 \\ 110 & 571 & 7 \\ 0 & 1 & 746 \end{bmatrix}$$

Accuracy: 92.7%
Precision: Class-wise precision values are 0.8600, 0.9400, and 0.9900, respectively. The mean precision across all classes is 0.9300.
Recall: The recall values for the classes are 0.9500, 0.8300, and 0.9900, respectively. The mean recall is 0.9233.
F-Measure: The F-measure scores are 0.9000, 0.8800, and 0.9900 for each class, with a mean F-measure of 0.9233.

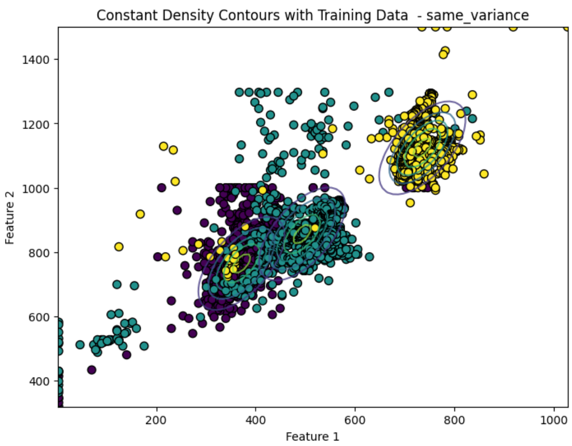

Figure 14: Constant Density Contours with Training Data for the Same Diagonal Covariance Assumption. The contours illustrate the probability density functions for each class, assuming identical variances and no feature correlations.

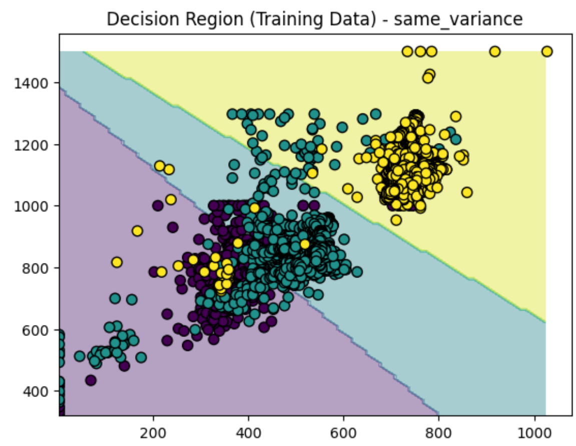

Figure 15: Decision Region Plot (Training Data) for the Same Diagonal Covariance Assumption. The plot shows the decision boundaries between classes, assuming that all classes share the same diagonal covariance matrix.

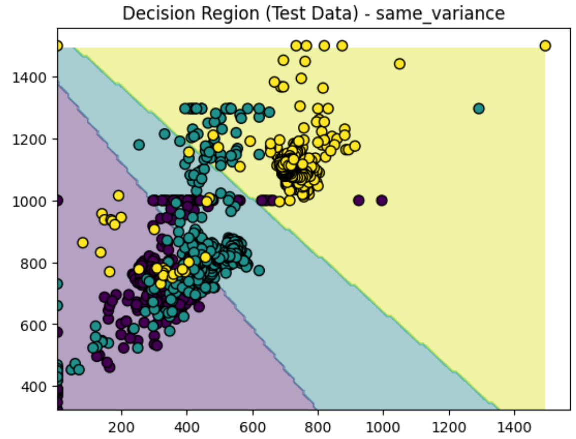

Figure 16: Decision Region Plot (Test Data) for the Same Diagonal Covariance Assumption. This visualization demonstrates the classifier's performance on unseen data with the same diagonal covariance assumption.

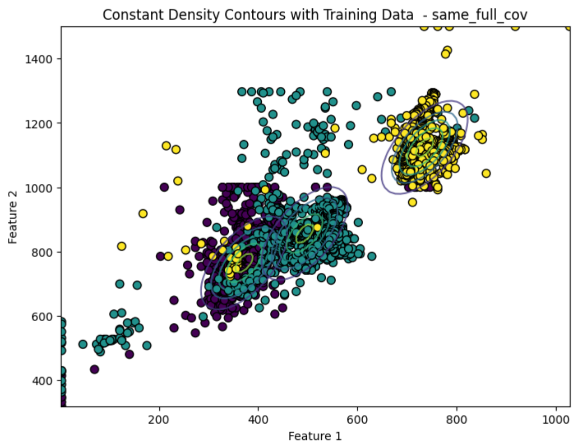

Figure 17: Constant Density Contours with Training Data for the Full Covariance Matrix. This comparison shows how density contours change when considering full covariance, which models feature correlations.

Summary: The assumption of a common diagonal covariance matrix across all classes achieved a high accuracy of 92.7

4.1.2 Covariance Matrix: Full (Σ)

For the full covariance matrix assumption, where each class is modeled with its own full covariance matrix, the performance metrics are detailed in Table 3. This approach allows for the modeling of feature correlations within each class.

Covariance Type	Accuracy (%)	Precision	Recall	F-Measure
Full Covariance	92.7	0.8600, 0.9400, 0.9900	0.9400, 0.8300, 0.9900	0.9000, 0.8800, 0.9900

Table 3: Performance Metrics for Full Covariance Matrix

Confusion Matrix:

$$\begin{bmatrix} 678 & 38 & 1 \\ 110 & 571 & 7 \\ 0 & 1 & 746 \end{bmatrix}$$

Accuracy: 92.7%
Precision: Class-wise precision values remain at 0.8600, 0.9400, and 0.9900, with a mean precision of 0.9300.
Recall: The recall values are 0.9400, 0.8300, and 0.9900, with a mean recall of 0.9233.
F-Measure: The F-measure scores are 0.9000, 0.8800, and 0.9900 for the classes, with a mean F-measure of 0.9233.

Summary: The full covariance matrix approach, which allows for feature correlations within each class, achieved the same accuracy of 92.7

4.1.3 Covariance Matrix: Diagonal (Different for Each Class)

For the diagonal covariance matrix assumption with different covariance matrices for each class, the performance metrics are outlined in Table 4. This approach allows each class to have its

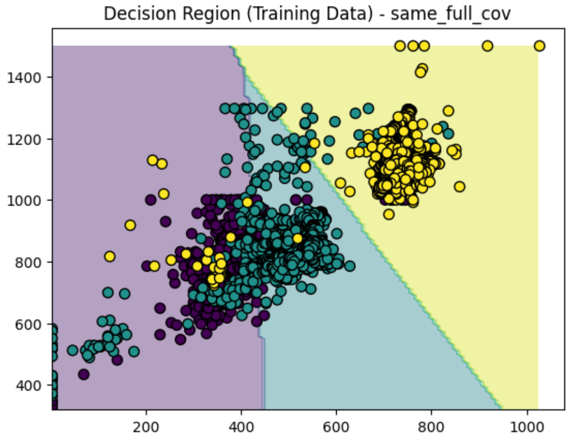

Figure 18: Decision Region Plot (Test Data) for the Full Covariance Matrix. The plot displays how the decision boundaries adjust when considering feature correlations within each class.

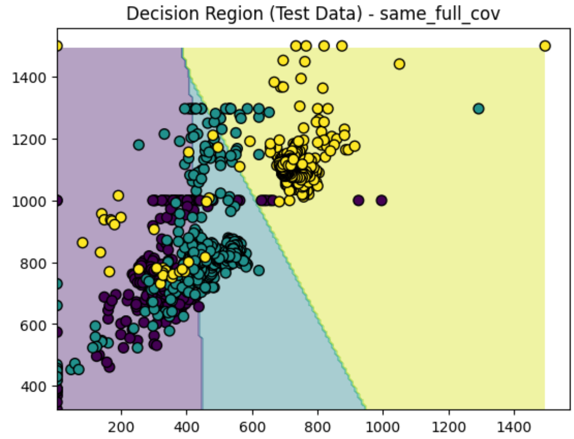

Figure 19: Constant Density Contours with Training Data for the Full Covariance Matrix. This visualization highlights how feature correlations are incorporated into the density estimation.

own diagonal covariance matrix, modeling class-specific feature variances.

Covariance Type	Accuracy (%)	Precision	Recall	F-Measure
Diagonal (Different)	92.7	0.8600, 0.9400, 0.9900	0.9500, 0.8300, 0.9900	0.9000, 0.8800, 0.9900

Table 4: Performance Metrics for Diagonal Covariance Matrix (Different for Each Class)

Confusion Matrix:

$$\begin{bmatrix} 678 & 38 & 1 \\ 110 & 571 & 7 \\ 0 & 1 & 746 \end{bmatrix}$$

Accuracy: 92.7%
Precision: Precision values are 0.8600, 0.9400, and 0.9900 for the classes, with a mean precision of 0.9300.
Recall: The recall values are 0.9500, 0.8300, and 0.9900, resulting in a mean recall of 0.9233.
F-Measure: F-measure scores are 0.9000, 0.8800, and 0.9900, with a mean of 0.9233.

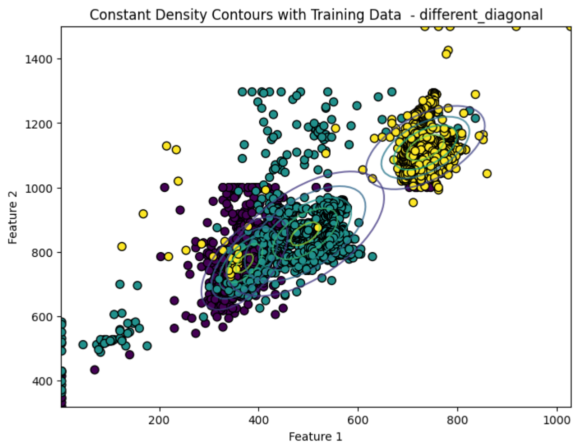

Figure 20: Constant Density Contours with Training Data for Diagonal Covariance Matrices (Different for Each Class). This plot shows how class-specific variances are modeled with separate diagonal matrices.

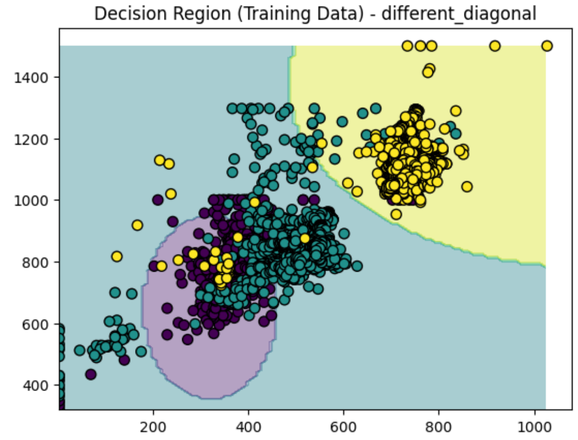

Figure 21: Decision Region Plot (Test Data) for Diagonal Covariance Matrices (Different for Each Class). The plot illustrates the decision boundaries when each class has its own diagonal covariance matrix.

Summary: The diagonal covariance matrix assumption with different matrices for each class produced identical classification performance to the other models, with an accuracy of 92.7

4.2 Conclusion

The analysis of the vowel dataset across different covariance assumptions reveals that while more complex models (full covariance or class-specific diagonal matrices) may offer improved flexibility in modeling class distributions, they do not always translate into better classification performance. The consistent accuracy of 92.7% across different models suggests that for this dataset, the simpler diagonal covariance assumptions are sufficient. Further investigation with other datasets or additional features might be necessary to fully assess the benefits of more complex covariance structures

5 Conclusion

The analysis of Bayesian classifiers across the different datasets and covariance assumptions reveals key insights into the effectiveness and limitations of various modeling approaches.

Dataset 1: Linearly Separable Data For the linearly separable dataset, the classifier demonstrated exceptional performance across all covariance assumptions, achieving 100% accuracy. This dataset's inherent linear separability allowed all models, from the simplest diagonal covariance to the most complex full covariance matrices, to achieve perfect classification results. The results underscore the suitability of simpler models, such as the diagonal covariance matrix, for linearly separable data, as they efficiently capture class boundaries without introducing unnecessary complexity. The consistent performance across different covariance assumptions suggests that, in cases of clear linear separability, more complex models do not provide additional benefits.

Dataset 2: Nonlinearly Separable Data The nonlinearly separable dataset presented a more challenging scenario. Despite the flexibility of the full covariance matrix and class-specific covariance matrices, the classifier's performance remained relatively low, with accuracy ranging between 29.16% and 30.33%. The decision region and contour plots revealed that the models struggled to capture the complex, spiral-like structure of the data. Both simpler and more complex covariance models exhibited limitations in modeling the nonlinear decision boundaries. This highlights the difficulty of using linear and Gaussian assumptions for datasets with intricate, nonlinearly separable patterns. Advanced or nonlinear methods might be necessary to handle such complexity effectively.

Dataset 3: Vowel Dataset The vowel dataset analysis showed that the classifier achieved a high and consistent accuracy of 92.7% across different covariance assumptions. Despite the availability of more complex covariance structures, such as the full covariance matrix and class-specific diagonal covariance matrices, simpler models (specifically those with diagonal covariance) proved to be nearly as effective. This suggests that for the vowel dataset, which has relatively straightforward class distributions, simpler models are not only sufficient but also more efficient. The results demonstrate that while complex models offer increased flexibility, their benefits may not always be realized, particularly when the dataset's complexity does not warrant such sophistication.

Overall Summary Across all datasets, the findings emphasize the importance of aligning model complexity with data characteristics. For linearly separable data, simpler covariance models are highly effective and computationally efficient. In contrast, for nonlinearly separable data, the limitations of both simple and complex covariance models become apparent, indicating the need for more advanced methods to capture complex patterns. The vowel dataset results further reinforce that simpler models can often achieve high performance, particularly when class distributions are well-defined.

These observations highlight the necessity of choosing appropriate model complexity based on dataset characteristics and the specific classification challenges presented. While more complex models can provide increased flexibility, their advantages are context-dependent, and simpler models may often suffice for well-defined or linearly separable datasets.