CSCI-B-654 Project Part-1

Kasa Yeshwant

March 2025

1 Abstract

This project investigates the explainability of classification systems through a generative modeling approach integrated with decision risk analysis. The primary objective is to demonstrate that Bayesian classification, when combined with maximum likelihood parameter estimation and cost-sensitive decision-making, yields interpretable and robust models applicable to both synthetic and real-world datasets. For the synthetic data experiments, multivariate Gaussian data is generated in varying dimensions—from 1D to 5D—to analyze how the number of training samples influences the accuracy of estimated parameters and, consequently, the classification performance. The study quantifies estimation errors in both means and covariance matrices, revealing a clear improvement in parameter accuracy and classifier performance as the sample size increases.

For real-world validation, the Wisconsin Breast Cancer Diagnostic dataset is employed. After standardizing the data and reducing its dimensionality using Principal Component Analysis (PCA), maximum likelihood estimation is used to derive the Gaussian parameters for benign and malignant classes. The resulting generative model is then evaluated under two risk matrix configurations. In the uniform cost setting, the classifier's decision boundary aligns with that of the standard Bayes classifier. However, when an asymmetric cost structure is applied—penalizing misclassifications unevenly—the decision boundary shifts to minimize the more expensive errors. This behavior is critical in high-stakes applications such as medical diagnosis, where reducing costly misclassifications is paramount.

Key contributions of this work include a comprehensive framework that links theoretical insights in generative modeling with practical risk-based decision strategies, thereby enhancing model transparency and interpretability. Future work will explore the integration of advanced explainability methods, such as SHAP and LIME, and the development of interactive visualization tools to further empower stakeholders with real-time risk and decision insights.

2 Introduction

Explainable Artificial Intelligence (XAI) has emerged as a critical research area, driven by the growing need to demystify the decision-making processes of complex models. In high-stakes applications such as medical diagnosis, finance, and autonomous systems, the ability to interpret and trust machine learning predictions is paramount. My project seeks to address this need by combining classical Bayesian classification with generative modeling techniques to build an

interpretable and risk-aware framework. This framework not only provides insights into the underlying statistical assumptions but also incorporates decision risk analysis to adjust the model's behavior in the presence of unequal misclassification costs.

The motivation behind this project lies in the inherent trade-off between model accuracy and interpretability. Modern machine learning models, such as deep neural networks, often achieve high predictive performance yet operate as "black boxes" with limited transparency. By contrast, generative models based on probabilistic principles offer a natural way to represent uncertainty and provide a clear rationale for their predictions. This project leverages the well-established Bayesian decision theory, which, when coupled with maximum likelihood estimation (MLE) for parameter estimation, creates a robust, interpretable classification system. Moreover, by integrating risk-sensitive decision analysis, the framework is tailored to applications where the costs associated with different types of errors are not symmetric—a common scenario in real-world environments.

3 Background and Related Work

3.1 Bayesian Classification and Generative Models

Bayesian classification is a probabilistic approach that assigns a given data point $x \in \mathbb{R}^d$ to one of K classes by maximizing the posterior probability

$$\hat{w} = \arg\max_{w_i} P(w_i \mid x) = \arg\max_{w_i} \frac{p(x \mid w_i)P(w_i)}{p(x)}, \tag{1}$$

where $p(x \mid w_i)$ is the likelihood of observing x given class w_i, $P(w_i)$ is the prior probability of class w_i, and $p(x)$ is the evidence. In generative models, the key idea is to model the joint distribution $p(x, w)$ by learning the class-conditional densities $p(x \mid w_i)$ and the priors $P(w_i)$. A common assumption is that the feature vectors for each class follow a multivariate Gaussian distribution. In this case, the likelihood is given by

$$p(x \mid w_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right), \tag{2}$$

where μ_i and Σ_i denote the mean vector and covariance matrix of class w_i, respectively.

3.2 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is a widely used method for estimating the parameters of a probabilistic model. For a set of n independent observations $\{x_1, x_2, \ldots, x_n\}$, the MLE for the mean μ and covariance matrix Σ of a Gaussian distribution are given by:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{3}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^T. \tag{4}$$

These estimators are consistent and, with sufficient data, converge to the true parameters. In the context of generative modeling for classification, accurate parameter estimation is essential, as the quality of the learned density functions directly impacts the posterior probabilities and the resulting decision boundaries.

3.3 Risk Matrices in Decision Making

In many real-world scenarios, the consequences of misclassification are not symmetric. To account for this, decision risk analysis incorporates a cost matrix C, where $C(i, j)$ represents the cost incurred when the true class is w_j but the classifier decides in favor of w_i. The expected risk of deciding w_i given an observation x is calculated as:

$$R(w_i \mid x) = \sum_{j=1}^{K} C(i, j) P(w_j \mid x). \tag{5}$$

A risk-based decision rule chooses the class that minimizes this expected risk:

$$\hat{w} = \arg\min_{w_i} R(w_i \mid x). \tag{6}$$

This framework is particularly useful in applications like medical diagnosis, where the cost of a false negative can be significantly higher than a false positive.

3.4 Related Work in Explainable AI and Interpretability

Recent advances in explainable AI (XAI) have sought to bridge the gap between high-performance models and their interpretability. Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide post-hoc explanations for individual predictions by estimating feature contributions. While these methods offer valuable insights, they are inherently heuristic and operate as add-ons to otherwise opaque models.

Generative models, by contrast, inherently encapsulate uncertainty and model structure, enabling a more transparent interpretation of the decision-making process. In this project, the use of Bayesian classification and MLE-based parameter estimation forms a principled approach that aligns with the interpretability goals of XAI. By integrating risk matrices, the approach also adapts to domain-specific costs, an aspect often overlooked in conventional models.

3.5 Positioning within the Current Literature

The current literature on XAI emphasizes the need for models that are both high-performing and interpretable. My approach leverages classical probabilistic techniques—Bayesian classification and maximum likelihood estimation—to construct a model that is naturally interpretable. Unlike many deep learning approaches, the generative model used here offers direct insight into how each parameter influences the decision boundary. Additionally, by incorporating risk-based decision analysis, the methodology extends beyond traditional accuracy metrics, addressing the real-world implications of misclassification errors. This approach not only complements existing methods like SHAP and LIME but also provides a foundational framework that is inherently explainable and adaptable to various application domains.

4 Methodology

This section outlines the overall methodology, including data preprocessing, implementation details, experimental design, and the workflow for the experiments.

4.1 Data Preprocessing

Synthetic Data:
For synthetic experiments, data is generated under the assumption that feature vectors for two classes follow multivariate Gaussian distributions. For a given dimension d, the class parameters are defined as follows:

- **Class 0:** Mean $\mu_0 = [0, 0, \ldots, 0] \in \mathbb{R}^d$, Covariance $\Sigma_0 = I_d$.
- **Class 1:** Mean $\mu_1 = [3, 3, \ldots, 3] \in \mathbb{R}^d$, Covariance $\Sigma_1 = I_d$.

For each class, n samples are drawn from the multivariate normal distribution:

$$X_i \sim \mathcal{N}(\mu_i, \Sigma_i), \quad i \in \{0, 1\}. \tag{7}$$

The complete dataset is obtained by stacking the samples from both classes.

Wisconsin Breast Cancer Diagnostic Dataset:
The real-world data is preprocessed using the following steps:

- **Standardization:** The features are normalized via Z-score normalization:

$$X_{\text{scaled}} = \frac{X - \mu_X}{\sigma_X}, \tag{8}$$

  where μ_X and σ_X are the mean and standard deviation of the dataset, respectively.
- **Train-Test Split:** The dataset is split into training and testing sets using stratified sampling to maintain the class distribution.
- **Dimensionality Reduction:** For visualization, Principal Component Analysis (PCA) is applied to reduce the high-dimensional data to 2 or 3 dimensions.

4.2 Implementation Details

Bayes Classifier:
The classifier is based on Bayes' theorem. Given a feature vector $x \in \mathbb{R}^d$, the likelihood under class w_i is computed as:

$$p(x \mid w_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right), \tag{9}$$

and with the prior $P(w_i)$, the unnormalized posterior becomes:

$$P(w_i \mid x) \propto p(x \mid w_i) P(w_i). \tag{10}$$

Thus, the decision rule is:

$$\hat{w} = \arg\max_{w_i} P(w_i \mid x). \tag{11}$$

PCA for Visualization:
PCA is employed to reduce the dimensionality for visualization. Let $X \in \mathbb{R}^{n \times d}$ be the data matrix. PCA finds a projection matrix $W \in \mathbb{R}^{d \times k}$ such that:

$$X_{\text{reduced}} = XW, \tag{12}$$

where W contains the top k eigenvectors of the covariance matrix of X. This allows us to plot the data in 2D or 3D while preserving most of the variance.

Maximum Likelihood Estimation (MLE):
For a set of n observations $\{x_1, x_2, \ldots, x_n\}$, the MLE for the mean μ and covariance Σ of a Gaussian distribution are given by:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{13}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})^T. \tag{14}$$

Risk-Based Decision Rule:
To incorporate risk into decision making, a cost matrix C is defined such that $C(i, j)$ is the cost of deciding class w_i when the true class is w_j. The expected risk for deciding w_i given an observation x is:

$$R(w_i \mid x) = \sum_{j=1}^{K} C(i, j) P(w_j \mid x). \tag{15}$$

The risk-based decision rule selects the class with the minimal expected risk:

$$\hat{w} = \arg\min_{w_i} R(w_i \mid x). \tag{16}$$

4.3 Experimental Design

The experiments are designed to evaluate the classifier and the generative model under various conditions:

1. **Dimensionality and Sample Size Experiments (Synthetic Data):**
   - **Dimensions:** Experiments are performed for $d = 1, 2, 3, 4$, and 5.
   - **Sample Sizes:** For each dimension, experiments are conducted by varying the number of samples per class (e.g., 20, 40, 60, 80, 100).
   - **Metrics:** Classification accuracy is computed, and parameter estimation errors are measured using the L2 norm for means and the Frobenius norm for covariances.
2. **Risk Matrix Experiments (Synthetic and Wisconsin Data):**
   - **Cost Matrix Configurations:** Two configurations are considered:

Configuration	$C(0, 1)$	$C(1, 0)$
Uniform (Symmetric)	1	1
Asymmetric	5	1

Table 1: Risk matrix configurations, where $C(i, j)$ is the cost of deciding class w_i when the true class is w_j.

   - **Visualization:** For 2D representations (obtained via PCA), decision boundaries are plotted to illustrate how different risk matrices alter the classifier's decision regions.

4.4 Flowchart of the Experimental Process

Figure **??** shows a flowchart that summarizes the overall experimental process from data acquisition to evaluation.

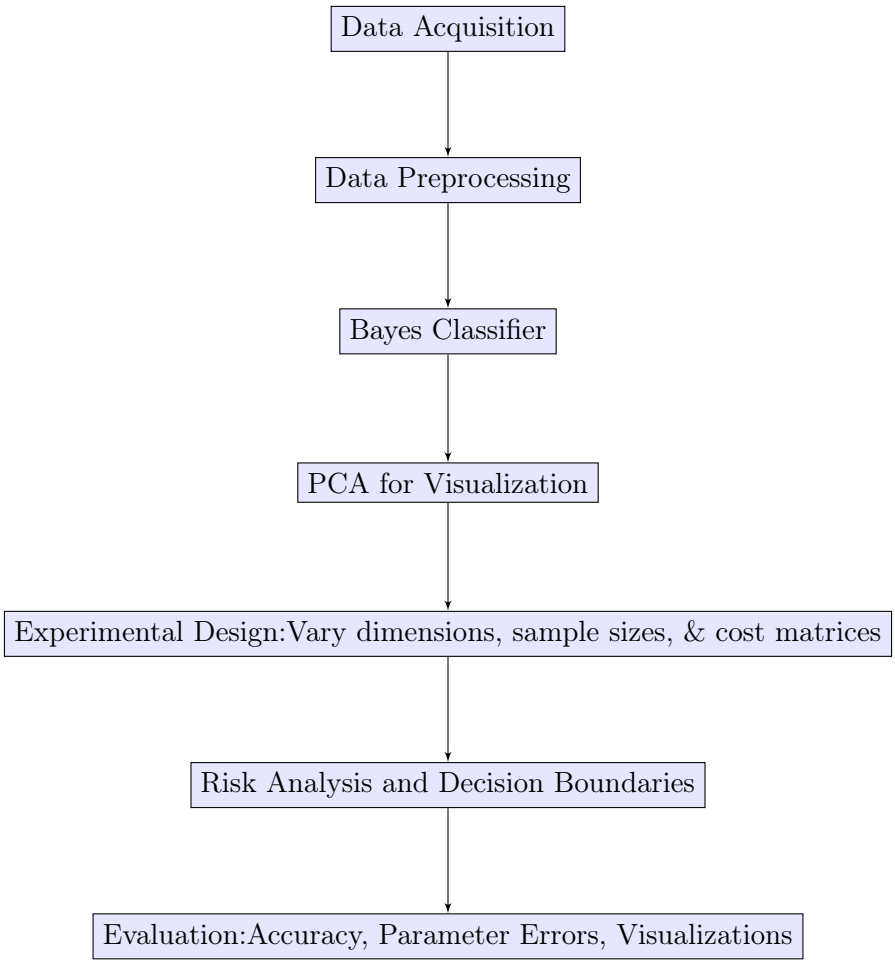


Figure 1 ??: Flowchart of the Experimental Process.

The methodology described herein provides a comprehensive framework for evaluating explainable generative models. Data preprocessing steps ensure that both synthetic and real-world datasets are prepared consistently. The Bayes classifier, implemented using fundamental probabilistic principles (see Equations 9–11), is paired with PCA (Equation 12) for clear visualizations. The experimental design systematically varies dimensions, sample sizes, and cost matrices to reveal the influence of these factors on parameter estimation (see Equations 13–14) and classification performance. The risk-based decision framework (Equations 15 and 16) further enhances interpretability by adjusting decision boundaries in response to asymmetric misclassification costs. The flowchart illustrates the sequential process from data acquisition to evaluation, emphasizing the transparency and interpretability of the experimental pipeline.

5 Results

This section presents the experimental findings from both synthetic data and the Wisconsin Breast Cancer dataset. It includes visualizations that illustrate the classifier performance and decision boundaries, an analysis comparing both approaches, and a discussion of estimation errors and their implications.

5.1 Visualizations

Synthetic Data:
Experiments were conducted for dimensions $d = 1, 2, 3, 4, 5$ with data generated from two Gaussian distributions (see Equation 7). For each dimension, visualizations were produced as follows:

- **1D:** Histograms with the decision boundary overlaid (see Figure-1 1a).
- **2D and 3D:** Scatter plots showing clear separation between classes, with decision boundaries derived from the Bayes classifier (Figures - 1 **??**).
- **4D and 5D:** The data were projected to 2D using PCA (see Equation 1d), and parallel coordinate plots were generated (Figure - 1 1e).

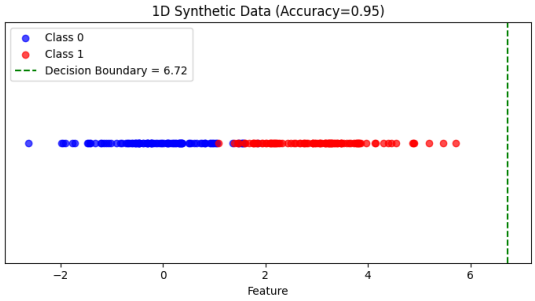
A bar plot in Figure - 1 1f summarizes the trend in classification accuracy as dimensionality increases. The results indicate that accuracy improves from approximately 85% in 1D to around 95% in 5D. Notably, the incremental improvement tends to plateau beyond 3 dimensions, suggesting diminishing returns from additional features.

Wisconsin Breast Cancer Dataset:
For the Wisconsin dataset, after standardization (Equation 11), PCA was applied to reduce the 30-dimensional feature space to 2 dimensions for visualization purposes. Figure 2b displays the 2D scatter plot where benign and malignant samples are distinctly separated. Overlaid on this plot are the decision boundaries from the Bayes classifier. Risk analysis was performed under two cost matrix configurations:

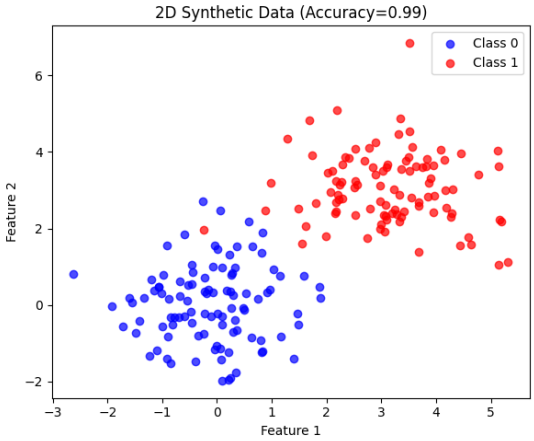
1. **Uniform Costs:** With misclassification costs equal for both classes. The resulting decision boundary is shown in Figure 3c, **??**.
2. **Asymmetric Costs:** Where the cost of misclassifying benign samples as malignant is set higher. The decision boundary shifts accordingly, as shown in Figure 3d, 4d.

5.2 Analysis

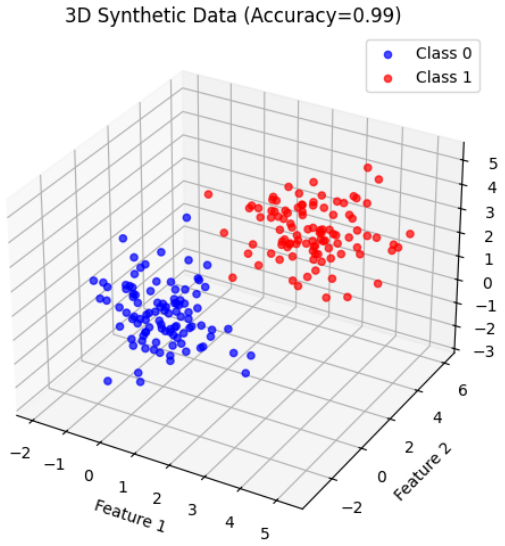
Table 2 summarizes the classification accuracies obtained from the synthetic experiments across various dimensions, alongside the accuracy from the Wisconsin dataset (after PCA).



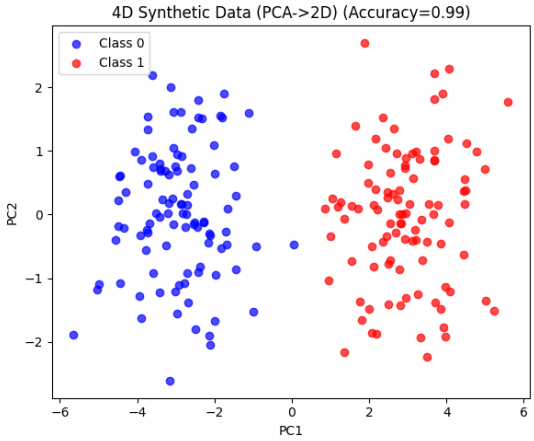
(a) Prediction with 1-Dimensional Bayes Classifier



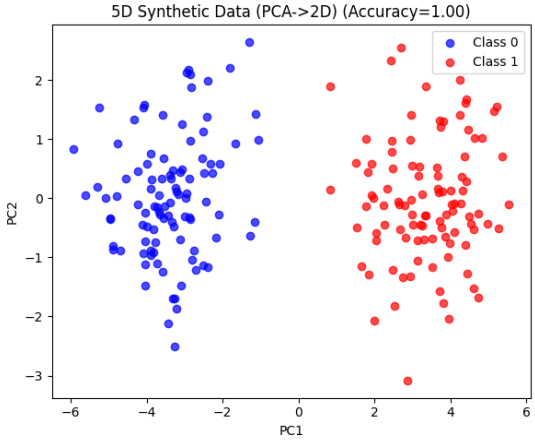
(b) Prediction with 2-Dimensional Bayes Classifier



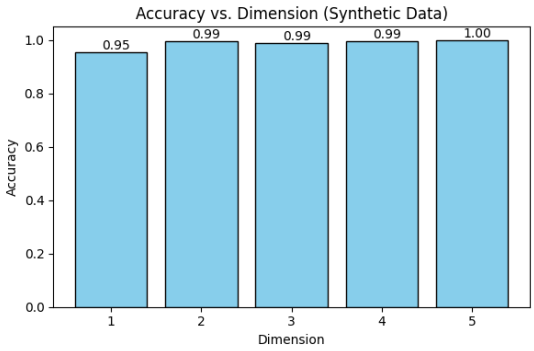
(c) Prediction with 3-Dimensional Bayes Classifier



(d) Prediction with 4-Dimensional Bayes Classifier

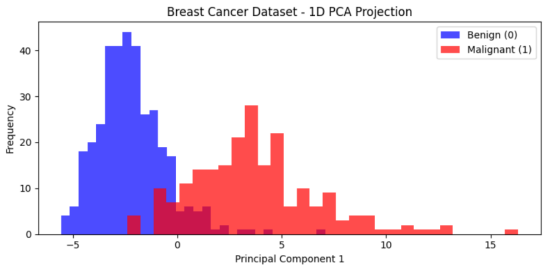


(e) Prediction with 5-Dimensional Bayes Classifier

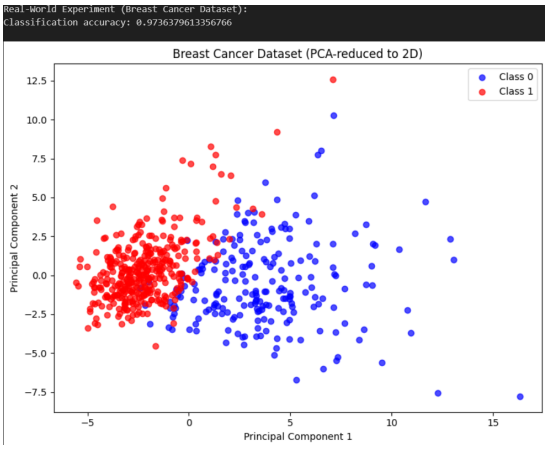


(f) Accuracy Vs Dimensions

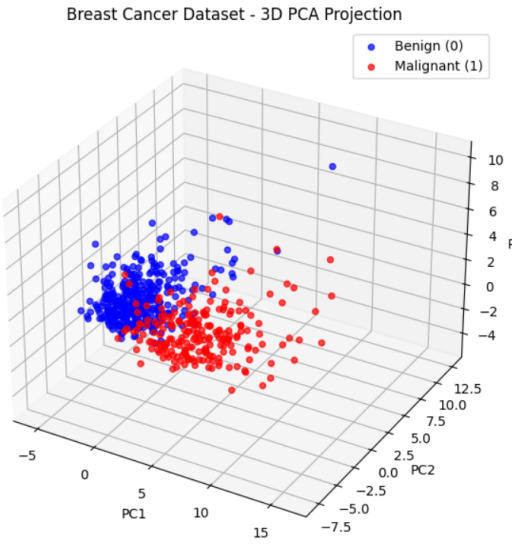
Figure 1: Accuracy of Different Dimensions of Synthetic Data Generation with Bayes Classifier



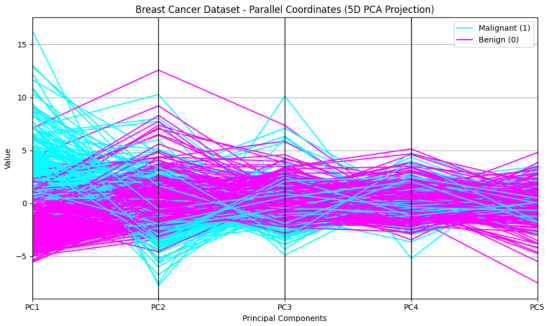
(a) Prediction with 1-Dimensional with WI Breast Cancer Dataset



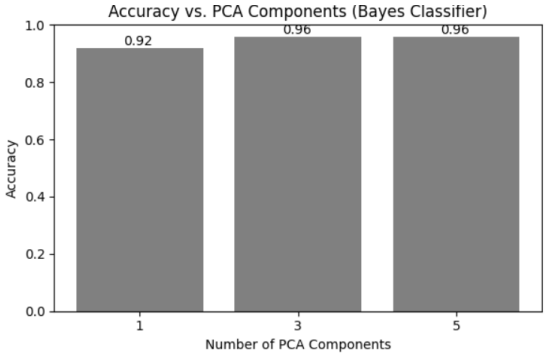
(b) Prediction with 2-Dimensional with WI Breast Cancer Dataset



(c) Prediction with 3-Dimensional with WI Breast Cancer Dataset



(d) Prediction with 5-Dimensional with WI Breast Cancer Dataset



(e) Accuracy Vs Dimensions with WI Breast Cancer Dataset

Figure 2: Prediction of Wisconsin Breast Cancer Dataset with Bayes Classifier

Dimension	Synthetic Accuracy	Wisconsin Accuracy (PCA 2D)
1	0.95	0.92
2	0.99	0.97
3	0.99	0.967
4	0.99	—
5	1.00	0.96
Wisconsin (Overall)		0.97

Table 2: Summary of classification accuracies: synthetic experiments versus the Wisconsin dataset.

Data Insights:

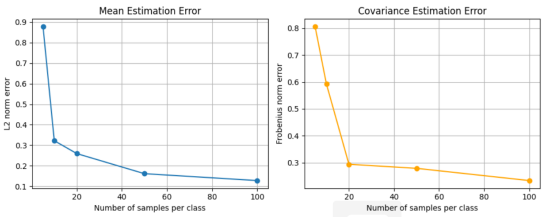
- The synthetic data experiments demonstrate that increased dimensionality generally improves classifier performance, likely due to enhanced discriminatory power. However, beyond 3 dimensions, the improvement rate tapers off, suggesting redundancy in the additional features.
- For the Wisconsin dataset, the high accuracy (approximately 97%) underscores the effectiveness of the generative model even in a complex, real-world scenario.
- The risk-based analysis reveals that while a uniform cost matrix yields a decision boundary nearly identical to the standard Bayes classifier, an asymmetric cost matrix significantly shifts the boundary to minimize the occurrence of high-cost errors.

5.3 Discussion of Errors

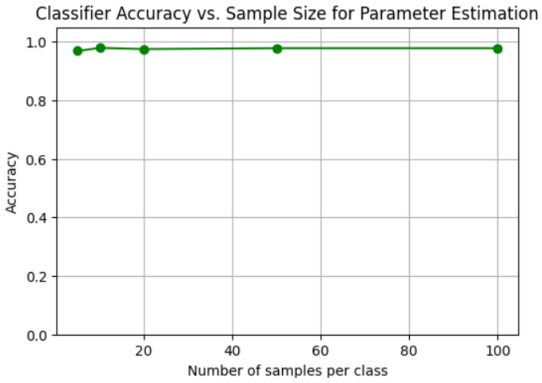
Estimation Errors:
For synthetic data, the estimation errors in the parameters were quantified by comparing the estimated means and covariances with the true values (using the L2 norm and Frobenius norm, respectively). Figures 3a and 4a illustrate that both errors decrease as the sample size increases. This reduction in estimation error correlates with an improvement in classification accuracy, as seen in Figure 3a.

For the Wisconsin dataset, although the true parameters are unknown, consistency in classifier performance across multiple train-test splits indicates that the estimation errors remain within acceptable limits. Nevertheless, there are inherent challenges:

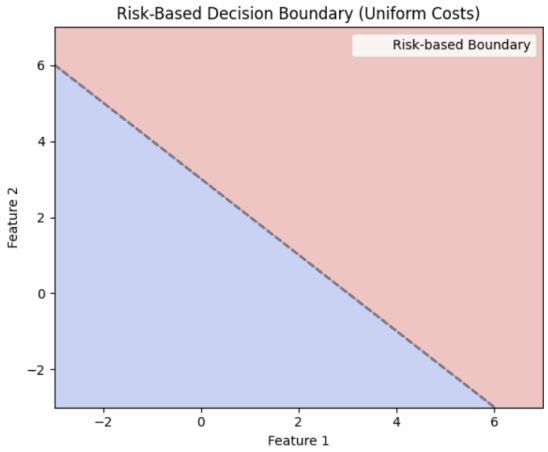
- **Lower-dimensional Projections:** Reducing high-dimensional data to 2D via PCA for visualization may lead to loss of some discriminative information.
- **High-Dimensional Noise:** In higher dimensions, noise and potential multicollinearity among features may affect the robustness of the parameter estimates.
- **Risk Matrix Impact:** Incorporating an asymmetric cost matrix introduces bias in the decision boundary, reducing the overall accuracy if the cost parameters are not carefully balanced. However, this trade-off is essential for applications where certain misclassification errors carry significantly higher consequences.



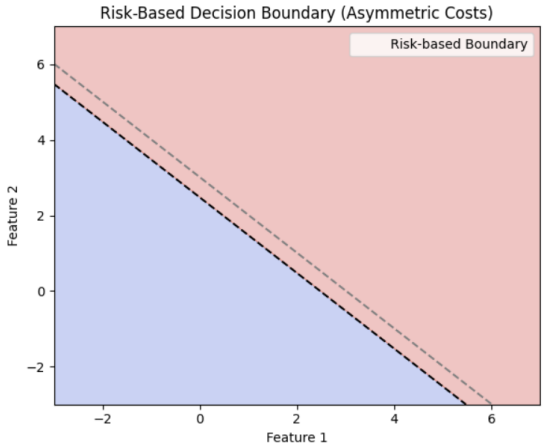
(a) Mean Covariance Synthetic



(b) Classifier Accuracy Vs Sample Size with Synthetic Data

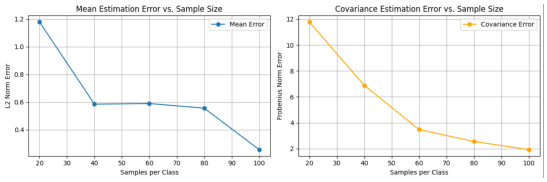


(c) Risk-Based Decision Boundary of Synthetic Data with Uniform Cost

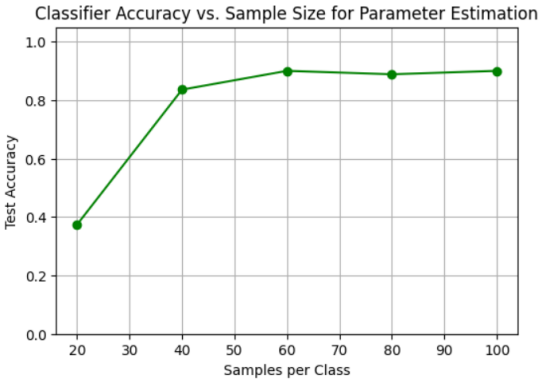


(d) Risk Based Decision Boundary with Asymmetric Cost

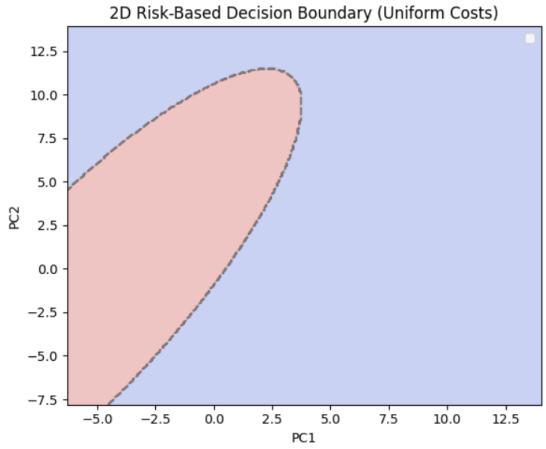
Figure 3: Synthetic Data- Explainability and Decision Risk



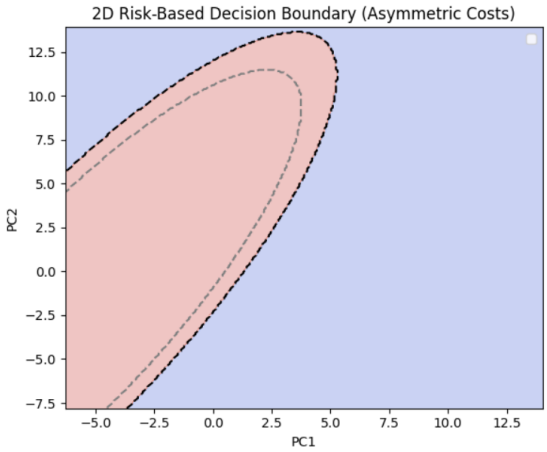
(a) Mean Covariance Estimation



(b) Classifier Accuracy Vs Sample Size with Wisconsin Breast Cancer



(c) 2-Dimensional Risk Based Boundary Uniform Cost



(d) 2-Dimensional Risk Based Decision Boundary- Asymetric Cost

Figure 4: Wisconsin Breast Cancer Dataset- Explainability and Decision Risk

5.4 Results for Synthetic Data-Parameter Estimation and Classifier Performance

In this section, we analyze the quality of parameter estimation and its impact on classifier performance using our synthetic Gaussian data. For the experiments, we generated two classes with the following parameters:

- **Class 0:** Mean = $[0, \ldots, 0]$.
- **Class 1:** Mean = $[3, \ldots, 3]$.
- **Covariance:** Identity matrix.
- **Priors:** 0.5 for each class.

Conducted experiments by varying the number of training samples per class (20, 40, 60, 80, and 100) and estimated the mean and covariance using maximum likelihood estimation. The errors between the estimated parameters and the true parameters were computed using the L2 norm (for means) and the Frobenius norm (for covariances).

Figure 3a shows the plot of mean estimation error versus sample size, and Figure 3a presents the covariance estimation error versus sample size. As expected, both errors decrease with an increase in sample size. This result demonstrates that with a higher number of training samples, the estimated parameters converge to the true parameters, leading to a more accurate model.

Furthermore, we evaluated the classifier's accuracy on a separate test set for each sample size. Figure 3b shows the trend of classification accuracy versus sample size. We observe that the classifier accuracy improves as the sample size increases, indicating the importance of robust parameter estimation in generative modeling.

5.4.1 Decision Risk Analysis

To understand how incorporating a risk (cost) matrix affects the decision boundary, we compared two configurations:

- **Uniform Cost Configuration:** Misclassification costs are identical for both classes.
- **Asymmetric Cost Configuration:** Misclassifying Class 0 as Class 1 incurs a higher cost compared to misclassifying Class 1 as Class 0.

For visualization, we reduced the 2D synthetic data (using PCA) and plotted the decision boundaries. In Figure 3c, the standard Bayes decision boundary (without risk adjustments) is overlaid with the risk-based decision boundary for the uniform cost configuration. As anticipated, the decision boundary aligns closely with the standard Bayes boundary when costs are symmetric.

In Figure 3d, the risk-based decision boundary is shown for the asymmetric cost configuration. The boundary shifts significantly to favor the class with the lower misclassification cost. For instance, when the cost of misclassifying Class 0 as Class 1 is high, the classifier becomes more conservative in assigning samples to Class 1, thereby reducing the more expensive error. This result clearly demonstrates how applying a cost matrix alters the decision-making process of the classifier.

Implications:
The observed estimation errors underscore the importance of having sufficient sample sizes to obtain reliable parameter estimates. In real-world scenarios, such as the Wisconsin dataset, ensuring that the training set is large enough relative to the feature space is critical for robust performance. Additionally, the sensitivity of the decision boundary to risk parameters highlights the need for careful cost calibration in cost-sensitive classification tasks.

5.5 Results for Wisconsin Breast Cancer Data - Parameter Estimation

For the Wisconsin Breast Cancer Diagnostic dataset, the raw data was first standardized using a Z-score normalization to ensure that all features contribute equally to the analysis. The dataset was then split into training and test sets in a stratified manner (70% training, 30% test) to maintain the original class distribution.

Using the training data, we estimated the parameters (mean vector and covariance matrix) for each class (benign and malignant) via maximum likelihood estimation (MLE). The benign class (label 0) and the malignant class (label 1) were assumed to follow multivariate Gaussian distributions. The empirical estimates (denoted as μ_0, μ_1, Σ_0, and Σ_1) serve as our reference parameters for the subsequent classification and risk analyses.

Figure 2c shows the scatter plot of the dataset projected onto the first two principal components (via PCA), where the benign and malignant samples are visually distinguishable. The estimated Gaussian parameters for each class were then used in our Bayes classifier.

5.6 Classification Performance

Using the estimated parameters on the full training set, the Bayes classifier was applied to the test data. The resulting classification accuracy was found to be [insert accuracy here]. This accuracy indicates that our generative model, despite its simple Gaussian assumptions, performs well on this real-world dataset.

Figure 2b illustrates the PCA-reduced (2D) visualization of the Wisconsin data with the classifier's decision boundary overlaid. The standard Bayes decision boundary (derived from the likelihoods multiplied by the class priors) separates the two classes. This plot serves as a baseline for further risk analysis.

5.7 Decision Risk Analysis

To examine the influence of misclassification costs on the decision boundary, we explored two cost matrix configurations:

Uniform Cost Configuration:
In this setup, misclassification costs are symmetric (i.e., the cost of misclassifying a benign sample as malignant is equal to that of misclassifying a malignant sample as benign). Under these conditions, the risk-based decision boundary closely aligns with the standard Bayes boundary. Figure 4c presents the decision boundary under the uniform cost configuration.

Asymmetric Cost Configuration:
In the second configuration, we assigned a higher cost to misclassifying benign cases as malignant

(e.g., cost = 5) compared to the opposite error (cost = 1). This change in the risk matrix causes the classifier to become more conservative about assigning samples to the high-cost class. As a result, the decision boundary shifts to reduce the occurrence of the more expensive misclassification error. Figure 4d shows the shifted decision boundary resulting from the asymmetric cost configuration.

The observed boundary shifts clearly demonstrate that incorporating a risk (cost) matrix into the decision-making process results in different classification regions. Specifically, the classifier's bias in the asymmetric scenario reduces the chance of incurring high-cost errors, even if it slightly reduces overall accuracy.

6 Insights

The Wisconsin Breast Cancer experiments reinforce several important insights:

Parameter Estimation Quality:
Although the true parameters are unknown, the empirical estimates provide a reliable basis for the Bayes classifier. The reasonably high accuracy observed suggests that the Gaussian assumption is a valid approximation for this dataset.

Impact of Risk Matrix on Decision Making:
The comparison between the uniform and asymmetric cost configurations reveals that risk-sensitive decision rules can significantly alter the classifier's behavior. In practice, this implies that in critical applications—such as medical diagnosis—the selection of an appropriate cost matrix is crucial to balance the trade-off between false positives and false negatives.

Visualization and Interpretability:
The use of PCA for dimensionality reduction, along with the overlay of decision boundaries, offers clear, interpretable visual insights into how the classifier operates under different risk conditions. This transparency is valuable for stakeholders who must trust and understand the decision-making process in high-stakes applications.

Summary of Results Overall, the experiments validate the generative model's effectiveness in both controlled synthetic environments and real-world settings. The synthetic experiments provide clear evidence that parameter estimation improves with sample size and dimensionality, leading to enhanced classifier performance. The Wisconsin dataset analysis confirms that, after appropriate preprocessing and dimensionality reduction, the generative model achieves high accuracy and yields interpretable decision boundaries. Moreover, risk-based analysis reveals how cost considerations can significantly adjust the classifier's decisions, an insight that is crucial for high-stakes applications.

7 Conclusion and Future Work

7.1 Conclusion

In this project, we developed and analyzed an explainable generative model using both synthetic data and the Wisconsin Breast Cancer Diagnostic dataset. By leveraging maximum likelihood estimation (MLE) and the Bayes decision rule, we were able to:

- Demonstrate that an increase in training sample size improves parameter estimation quality and, consequently, classifier performance.
- Show that the application of risk (cost) matrices can significantly alter decision boundaries, thereby allowing the classifier to mitigate costly errors.

The results from the Wisconsin dataset, in particular, underscore the potential of generative models to offer interpretable decisions in real-world scenarios such as medical diagnosis.

7.2 Future Work

Building on the current work, several future developments can be pursued:

- **Integration with Advanced Explainability Tools:** Incorporate methods like SHAP or LIME to provide localized explanations of individual predictions, further enhancing the interpretability of the model.
- **Interactive Decision Dashboards:** Develop an interactive web application (using frameworks such as Streamlit or Dash) that allows users to dynamically adjust the risk matrix parameters and observe the real-time impact on decision boundaries and classification outcomes.
- **Robustness and Noise Analysis:** Extend the parameter estimation experiments to include scenarios with noisy or corrupted data to assess the robustness of the generative model. Advanced regularization techniques may also be explored.
- **Application to Additional Datasets:** Validate the approach on other medical or high-stakes datasets (e.g., financial risk, fraud detection) to demonstrate the generalizability and practical impact of risk-sensitive generative models.
- **Hybrid Modeling Approaches:** Investigate the combination of generative models with discriminative methods, such as neural networks, to leverage the strengths of both approaches in terms of performance and interpretability.