17 Gaussian Discriminant Analysis: From Circles to Ellipses

17.1 Foundations of Gaussian Discriminant Analysis

17.1.1 What is Gaussian Discriminant Analysis?

Gaussian Discriminant Analysis (GDA) is a probabilistic model for classification that assumes the data from each class follows a multivariate Gaussian (normal) distribution. This is where its name originates. As a supervised learning algorithm, GDA is designed to separate data points into categories based on how likely they are to belong to each class under these Gaussian models.

In essence, GDA assumes that each class has its own mean vector and covariance matrix, which describe the shape and orientation of that class's data distribution in feature space. During training, the algorithm estimates these parameters from the data. Then, for a new input point, GDA computes the posterior probability that the point belongs to each class using Bayes' theorem, and assigns it to the class with the highest posterior probability.

17.1.2 The Principle Behind GDA

Figure 1 illustrates the principle behind Gaussian Discriminant Analysis. Each ellipse represents the contour of a Gaussian distribution in a two-dimensional feature space. Notice how the shapes and orientations of the ellipses differ — this reflects how each class's covariance matrix captures its unique spread and correlation structure. Figure 1 also shows that the decision boundary separating the classes can take different forms — sometimes a straight line, other times a curved (quadratic) one — depending on the assumptions about the covariance matrices.

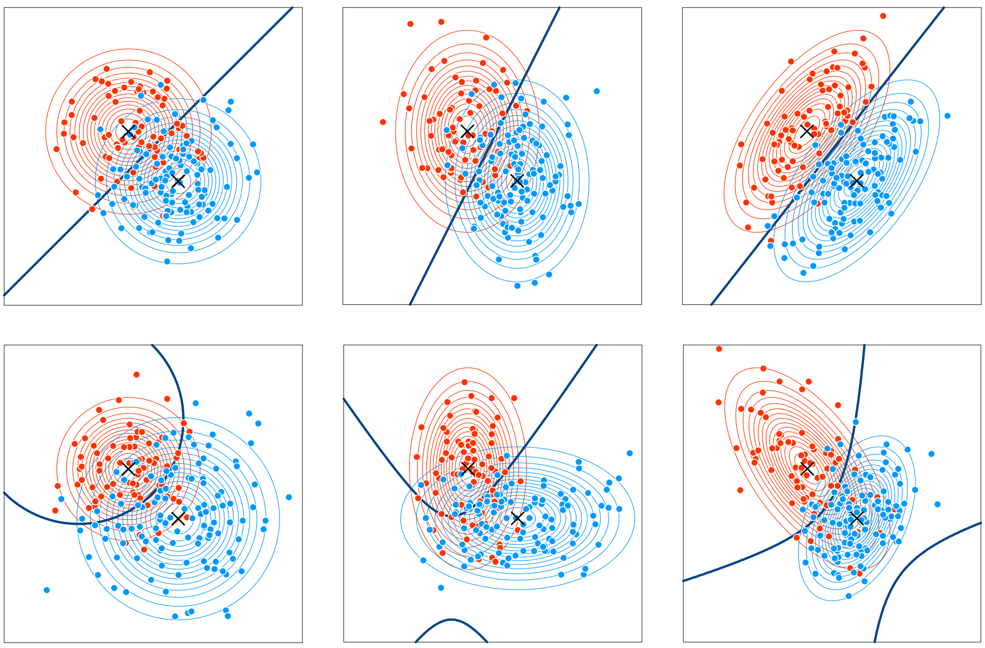

Figure 1. Illustration of Gaussian Discriminant Analysis. Each subplot shows elliptical contours of Gaussian distributions representing different classes. The decision boundary between them can be linear or curved, depending on model assumptions.

17.1.3 From LDA to QDA: One Model, Two Flavors

GDA includes two main variants: **Linear Discriminant Analysis (LDA)** and **Quadratic Discriminant Analysis (QDA)**.

The distinction lies in the assumption about covariance matrices. LDA assumes that all classes share the same covariance matrix, which results in **linear decision boundaries**. In contrast, QDA allows each class to have its own covariance matrix, leading to **quadratic decision boundaries** that better adapt to curved data distributions.

QDA is conceptually related to **Gaussian Naïve Bayes**, which also models each class with a Gaussian distribution but assumes that features are independent. GDA, by contrast, captures the **correlation between features** through its full covariance matrices. Moreover, GDA is also closely related to the **Gaussian Mixture Model (GMM)**, a probabilistic clustering technique discussed later in this book.

17.1.4 Optimization Objective and Posterior Probabilities

The goal of GDA is to minimize misclassification by assigning each input point $\boldsymbol{x}$ to the class C_k that maximizes the expected posterior probability. In general form, the classifier seeks to minimize the expected cost:

$$\hat{y} = \arg\min_{C_m} \sum_{k=1}^{K} f_{Y|\chi}\left(C_k \mid \boldsymbol{x}\right) \cdot c\left(C_m \mid C_k\right) \tag{1}$$

Here, K is the number of classes, and $c(C_m \mid C_k)$ represents the cost of predicting class C_m when the true class is C_k. Typically, $c(C_m \mid C_k) = 0$ for correct classifications ($m = k$) and $c(C_m \mid C_k) = 0$ for misclassifications ($m \neq k$). This formulation emphasizes the minimization of classification errors, contrasting with Naïve Bayes, which maximizes likelihood under independence assumptions.

Using Bayes' theorem, GDA computes the posterior probability that a data point $\boldsymbol{x}$ belongs to class C_k:

$$\underbrace{f_{Y|\chi}\left(C_k \mid \boldsymbol{x}\right)}_{\text{Posterior}} = \frac{\overbrace{f_{\chi,Y}\left(\boldsymbol{x}, C_k\right)}^{\text{Joint}}}{\underbrace{f_{\chi}\left(\boldsymbol{x}\right)}_{\text{Evidence}}} = \frac{\overbrace{f_{\chi|Y}\left(\boldsymbol{x} \mid C_k\right)}^{\text{Likelihood}} \overbrace{p_Y\left(C_k\right)}^{\text{Prior}}}{\underbrace{f_{\chi}\left(\boldsymbol{x}\right)}_{\text{Evidence}}} \tag{2}$$

The denominator is the evidence term, which ensures that all posteriors sum to one. However, since it is the same for all classes, it can be omitted when comparing classes, leading to the proportionality:

$$\underbrace{f_{Y|\chi}\left(C_k \mid \boldsymbol{x}\right)}_{\text{Posterior}} \propto \underbrace{f_{\chi,Y}\left(\boldsymbol{x}, C_k\right)}_{\text{Joint}} \tag{3}$$

This means that the classification decision depends only on the likelihood of the data under each class and the prior probability of each class.

17.1.5 Modeling with Multivariate Gaussians

In GDA, the likelihood term is modeled as a multivariate Gaussian distribution:

$$f_{\chi|Y}\left(\boldsymbol{x}|C_k\right)=\frac{\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}_k\right)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_k\right)\right)}{\sqrt{\left(2\pi\right)^D\left|\boldsymbol{\Sigma}_k\right|}} \tag{4}$$

Here, D is the number of features, $\boldsymbol{\mu}_k$ is the mean vector of class C_k, and $\boldsymbol{\Sigma}_k$ is its covariance matrix. This formulation is where GDA and Gaussian Naïve Bayes diverge: while Naïve Bayes assumes a diagonal covariance matrix (independent features), GDA uses the full covariance matrix, capturing interactions between features.

17.2 Six Covariance Matrix Configurations in GDA

17.2.1 Overview of Covariance Matrix Types

In the previous section, we introduced Gaussian Discriminant Analysis (GDA) as a probabilistic model that assumes each class follows a multivariate Gaussian distribution. The shape of this Gaussian — and therefore the form of the decision boundary — depends entirely on the covariance matrix $\boldsymbol{\Sigma}_k$ of each class.

In contrast to Gaussian Naïve Bayes, which assumes that all features are conditionally independent (that is, $\boldsymbol{\Sigma}_k$ is diagonal and features are uncorrelated), GDA allows more flexible assumptions about the relationships among features. Depending on how we constrain or free the covariance matrices, GDA can take on different forms — six in total. Each form reflects a distinct geometric pattern of the probability density function (PDF) contours and decision boundaries.

Table 1. Six forms of Gaussian Discriminant Analysis classified by covariance matrix structure

Category	Covariance matrix	Feature Dependence	PDF Contours	Decision Boundary Type
Type I	Same across all classes, diagonal with equal variances	Independent features	Perfect circles of equal size	Linear (straight line)
Type II	Same across all classes, diagonal with unequal variances	Independent features	Ellipses of equal size and orientation	Linear
Type III	Same across all classes, full covariance (non-diagonal)	Correlated features allowed	Rotated ellipses of equal shape	Linear
Type IV	Different across classes, diagonal with equal variances	Independent features	Perfect circles, different sizes	Quadratic (curved)
Type V	Different across classes, diagonal with unequal variances	Independent features	Ellipses of different sizes	Quadratic (curved)
Type VI	Different across classes, full covariance (non-diagonal)	Correlated features allowed	Rotated ellipses of different shapes	Quadratic (curved)

The six categories of discriminant analysis of the elliptic shape of the Gaussian distribution are shown in Figure 2.

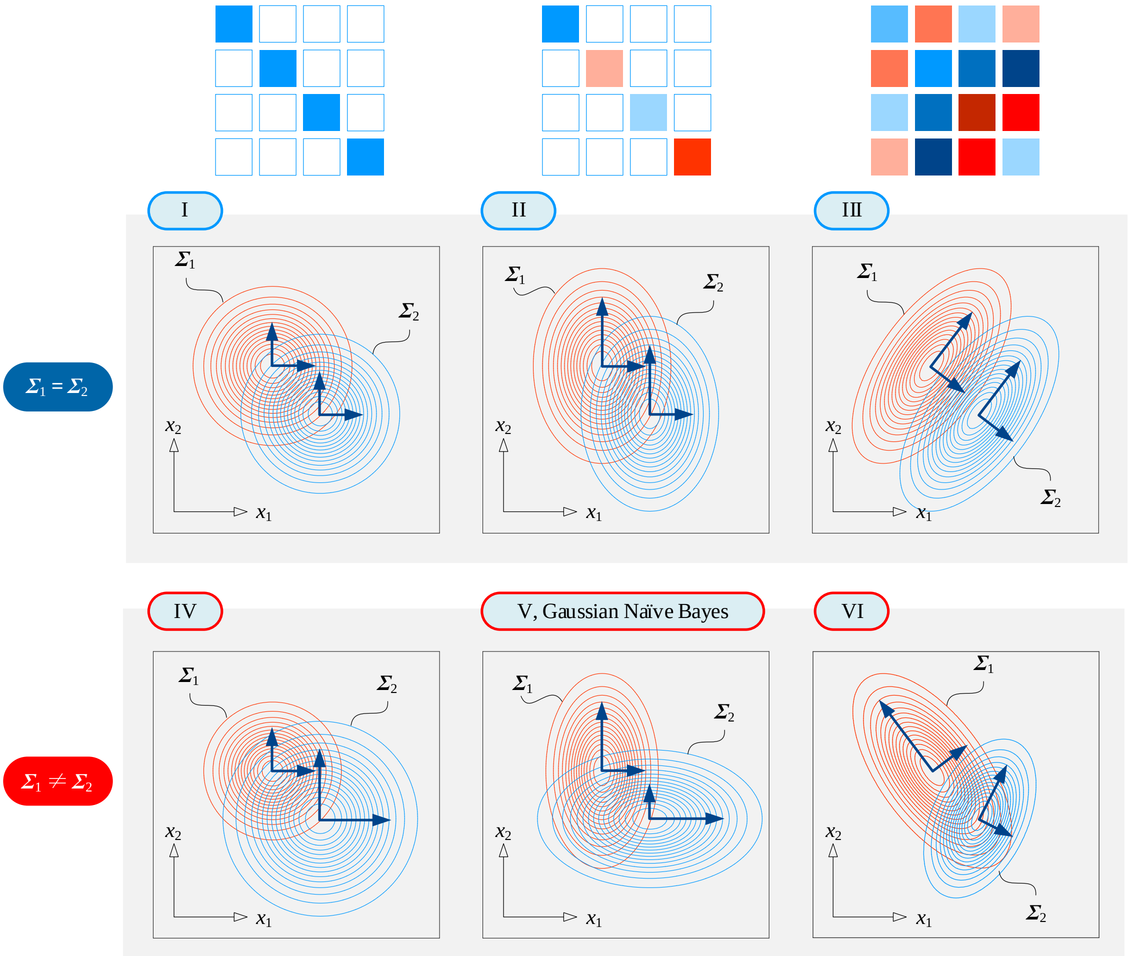

Figure 2. Six categories of Gaussian Discriminant Analysis showing the elliptical contours of class-conditional Gaussian distributions when $K = 2$ and $D = 2$.

Each subplot visualizes how changes in the covariance matrix affect the shape of the probability contours and the type of decision boundary.

17.2.2 Linear Cases (Types I–III): Shared Covariance

The first three categories — Types I, II, and III — share a key assumption: **all classes have the same covariance matrix**. This assumption leads to **linear decision boundaries**, forming the basis of **Linear Discriminant Analysis (LDA)**.

- **Type I** assumes that Σ_k is a diagonal matrix and that all diagonal elements (variances) are equal. In other words, all features are independent and contribute equally. The result is a set of **perfectly circular contours** with identical radii for all classes.
- **Type II** still assumes a diagonal covariance matrix (independent features), but now each feature can have a different variance. The contours become **ellipses** rather than circles, but since all classes share the same Σ_k, the ellipses are of the same size and shape.
- **Type III** removes the diagonal constraint, allowing the covariance matrix to include non-zero off-diagonal terms. This means features can be **correlated**. The contours become **rotated ellipses**, but they are still identical across classes because the covariance structure is shared.

Because these three types share the same covariance across all classes, the log-likelihood ratio simplifies to a linear function of $\boldsymbol{x}$, resulting in **straight decision boundaries**.

17.2.3 Quadratic Cases (Types IV–VI): Distinct Covariance

The last three categories — Types IV, V, and VI — relax the assumption that all classes share the same covariance matrix. When each class has its own $\boldsymbol{\Sigma}_k$, the log-likelihood ratio becomes a quadratic function of $\boldsymbol{x}$, producing **curved decision boundaries**. These cases are collectively known as **Quadratic Discriminant Analysis (QDA)**.

- **Type IV** assumes diagonal covariance matrices (independent features) with **equal variances within each class**, but the overall covariance differs between classes. This produces **circular contours** of different radii.
- **Type V** still assumes diagonal covariance matrices (independent features), but the variances differ both across features and across classes. The contours become **ellipses** of different shapes and sizes. Geometrically, this case corresponds to the **Gaussian Naïve Bayes classifier**, since both assume feature independence.
- **Type VI** removes all restrictions on the covariance matrices, allowing each class to have its own full covariance matrix with possible correlations between features. This produces **rotated ellipses** of varying shapes and orientations — the most general and flexible case of GDA.

The geometric relationships described here form the foundation not only of GDA but also of models in **unsupervised learning** such as the **Gaussian Mixture Model (GMM)**, which uses similar covariance structures to model clusters of data without labeled supervision. Understanding how covariance matrices influence shape and boundary is therefore essential for both classification and clustering models.

17.3 Decision Boundaries in GDA: Geometry Meets Probability

17.3.1 The Discriminant Function

In Gaussian Discriminant Analysis (GDA), the **decision boundary** defines where the model changes its prediction from one class to another. In other words, it is the surface (or line, in two dimensions) that separates regions of the feature space associated with different classes. The shape of this boundary depends on how the class distributions are modeled — particularly on the covariance matrices of the Gaussian functions.

To formalize this, we define a **discriminant function** $g_k(\boldsymbol{x})$ for each class C_k. This function represents how strongly a data point $\boldsymbol{x}$ is associated with class C_k. It is given by the logarithm of the posterior probability (up to a constant):

$$
\begin{aligned}
g_k(\boldsymbol{x}) &= \ln\left(f_{\chi,Y}(\boldsymbol{x}, C_k)\right) = \ln\left(f_{\chi|Y}(\boldsymbol{x}|C_k)\, p_Y(C_k)\right) \\
&= \ln\left(\frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right)}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}_k|}}\, p_Y(C_k)\right) \\
&= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k) - \frac{D}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_k| + \ln p_Y(C_k)
\end{aligned}
\tag{5}
$$

This formulation is often called the **log-discriminant function**.

Taking the logarithm is useful because the original Gaussian distribution involves an exponential term. The log transformation "removes" the exponential and simplifies the expression into a sum of terms involving $\boldsymbol{x}$, making the boundary derivation much easier to handle.

17.3.2 Two Features, Two Classes: An Intuitive Example

To illustrate the idea intuitively, let us focus on a **binary classification problem** (K = 2) with **two features** (D = 2).

We can write discriminant functions for the two classes as $g_1(\boldsymbol{x})$ and $g_2(\boldsymbol{x})$

$$\begin{cases} g_1(\boldsymbol{x}) = -\dfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - \dfrac{D}{2}\ln(2\pi) - \dfrac{1}{2}\ln|\boldsymbol{\Sigma}_1| + \ln p_Y(C_1) \\ g_2(\boldsymbol{x}) = -\dfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) - \dfrac{D}{2}\ln(2\pi) - \dfrac{1}{2}\ln|\boldsymbol{\Sigma}_2| + \ln p_Y(C_2) \end{cases} \tag{6}$$

The **decision boundary** itself occurs where both discriminant functions are equal:

$$g_1(\boldsymbol{x}) = g_2(\boldsymbol{x}) \tag{7}$$

At this point, the model is "undecided," meaning the posterior probabilities of both classes are equal.

Substituting the Gaussian likelihood expressions into the discriminant functions, we can expand the equality above to obtain an analytical form of the boundary:

$$\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) = \ln p_Y(C_1) - \ln p_Y(C_2) + \left(\frac{1}{2}\ln|\boldsymbol{\Sigma}_2| - \frac{1}{2}\ln|\boldsymbol{\Sigma}_1|\right) \tag{8}$$

This equation can represent a **quadratic** or a **linear** surface depending on the structure of the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

If the covariance matrices of the two classes are **identical**, the quadratic terms cancel out, and the decision boundary becomes a **straight line** (as in **Linear Discriminant Analysis**).

If the covariance matrices are **different**, the quadratic terms remain, leading to a **curved boundary** (as in **Quadratic Discriminant Analysis**).

The **prior probabilities** influence only the **constant term**, which slightly shifts the boundary toward one class or the other, depending on how likely each class is before observing the data.

17.3.3 Geometric Interpretation of Boundaries

The geometry of the decision boundary reveals how GDA captures class structure in data. Depending on the covariance matrices, the boundary can take various forms — a straight line, a circle, an ellipse, a parabola, or even a hyperbola. These shapes correspond to different assumptions about feature correlation and variance within each class.

For instance, when both classes have identical, spherical covariance matrices, the boundary is a straight line equidistant between the two means. When one class is more spread out, the boundary curves toward it, reflecting the model's uncertainty in distinguishing overlapping regions.

Figure 3 shows several common decision boundary patterns observed in two-dimensional Gaussian Discriminant Analysis with two classes. The contour lines represent the Gaussian distributions of each class, while the thick curve between them marks the decision boundary. As the covariance structure changes, the boundary transforms smoothly from linear to various quadratic forms, revealing the flexibility of GDA in modeling complex data distributions.

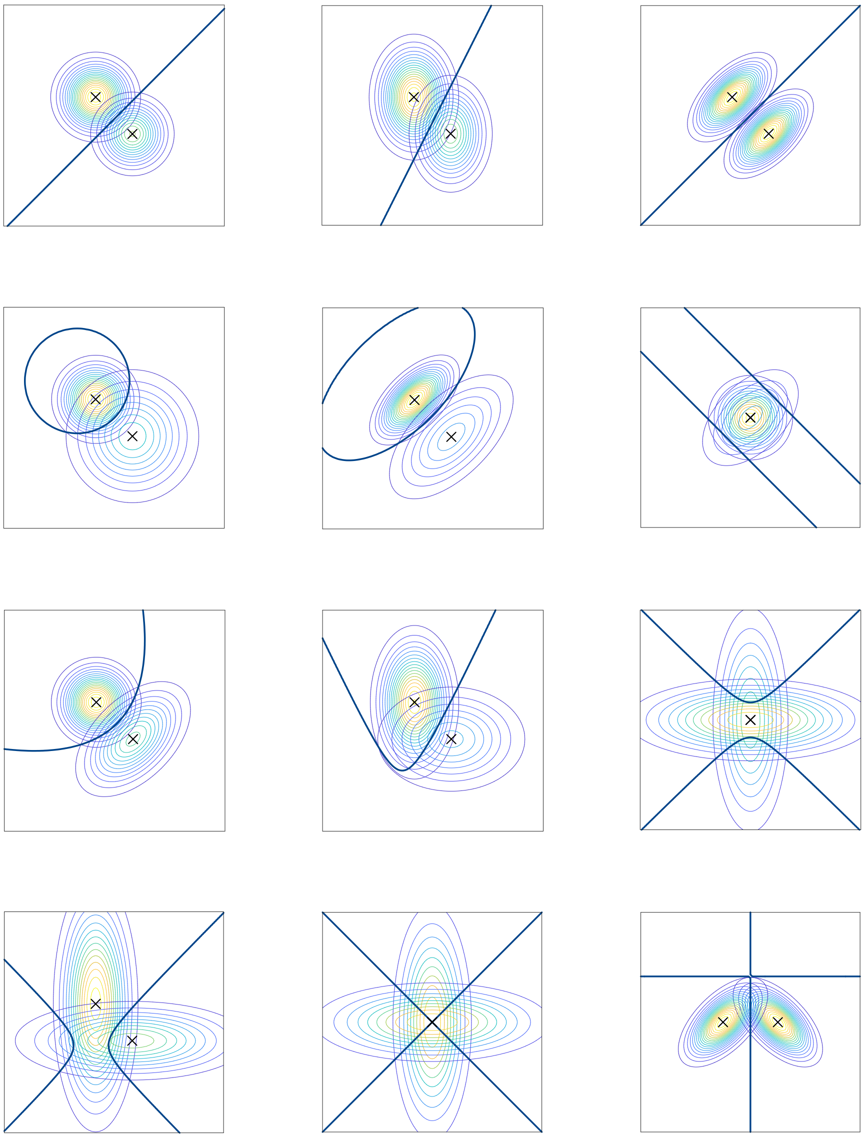

Figure 3. Examples of typical decision boundaries in two-class Gaussian Discriminant Analysis. Depending on the covariance matrices, the boundary may appear as a straight line, a circle, an ellipse, a parabola, a hyperbola, or other quadratic curves.

17.4 Applying Each Type of GDA

17.4.1 Type I: Simple Circles, Linear Boundaries

The **first type of Gaussian Discriminant Analysis (GDA)** assumes that all features are *conditionally independent* and that each feature contributes equally to the overall variance of the data. In mathematical terms, this means the covariance matrix $\boldsymbol{\Sigma}_k$ for each class is a **diagonal matrix** (off-diagonal elements are zero) and that all diagonal elements are **equal** (same variance across features).

Under these assumptions, the contours of each class's Gaussian probability density function (PDF) are **perfect circles** in the feature space, and the resulting **decision boundary** is a **straight line**. This is the simplest and most symmetric case of GDA.

To make this idea concrete, let us consider a two-dimensional, two-class problem ($K = 2$ and $D = 2$).

Each class C_1 and C_2 has a mean vector $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and the same covariance matrix:

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \sigma^2 \boldsymbol{I} \tag{9}$$

The inverse of both covariance matrices is therefore:

$$\boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_2^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{\boldsymbol{I}}{\sigma^2} \tag{10}$$

Substituting this into the general decision boundary equation derived earlier, we obtain a simplified **linear discriminant function**.

$$\begin{aligned} &\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}} \frac{\boldsymbol{I}}{\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu}_1) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}} \frac{\boldsymbol{I}}{\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu}_2) = \ln p_Y(C_1) - \ln p_Y(C_2) \\ &\Rightarrow (\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_1) - (\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_2) = 2\sigma^2\left(\ln p_Y(C_1) - \ln p_Y(C_2)\right) \end{aligned} \tag{11}$$

When the covariance matrices are identical, the quadratic terms in the general decision function cancel out. What remains is a linear relationship between the features, describing a straight decision boundary:

$$(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{x} - \left[\sigma^2\left(\ln p_Y(C_1) - \ln p_Y(C_2)\right) + \frac{1}{2}\left(\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_1\right)\right] = 0 \tag{12}$$

This linear form shows that the boundary depends on the difference between the class means and the prior probabilities of the classes. The vector $\boldsymbol{w} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$ serves as the **normal vector** (or gradient) of the separating line, defining its orientation in the feature space.

Let us examine two cases:

- When the prior probabilities are **equal**, and the class means are different ($\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$), the boundary becomes a line that passes through the **midpoint** between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and is **perpendicular** to the line connecting these two centroids. This means each class occupies an equal region of influence in the feature space.
- When the priors are **unequal**, the boundary **shifts toward the less probable class**. Intuitively, the more likely class with higher prior "claims" a larger region, since the model expects to see it more often. Thus, the boundary moves closer to the class with smaller prior probability.

This shift is one of the most intuitive illustrations of how prior beliefs affect classification even before observing new data.

Consider the following setup:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad p_Y(C_1) = 0.6, \quad p_Y(C_2) = 0.4, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{13}$$

Here, each class has independent features with equal variance in all directions. Figure 4 shows the Gaussian likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$.

At any point $\boldsymbol{x}$, if $f_{Y,\chi}(C_1, \boldsymbol{x}) > f_{Y,\chi}(C_2, \boldsymbol{x})$, the model classifies $\boldsymbol{x}$ as belonging to C_1; otherwise, it belongs to C_2. The decision boundary is where these two surfaces intersect — where the likelihoods are equal.

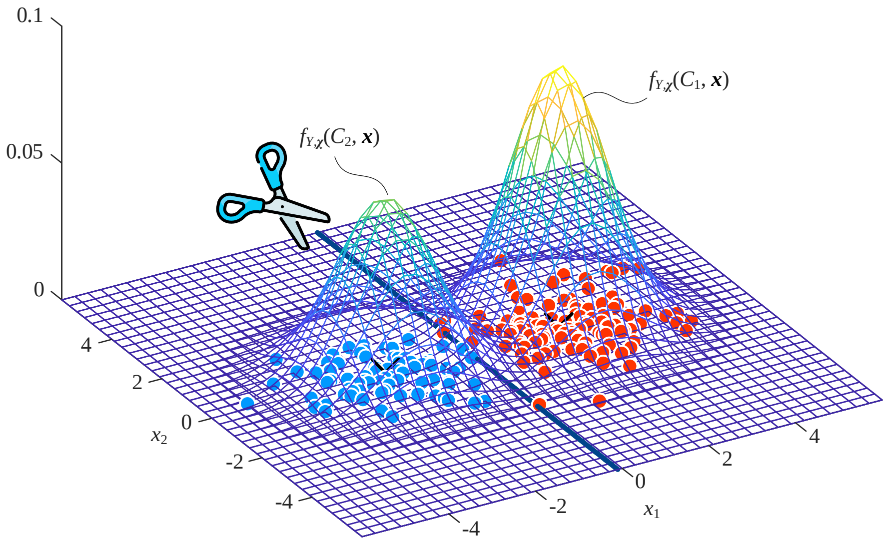

Figure 4. Likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ for the first type of Gaussian Discriminant Analysis.

A clearer way to visualize this is through contour plots. Figure 5 (a) compares the contour lines of $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$. Because both classes have the same spherical covariance matrix, their contours are **concentric circles** of different centers but identical shape. The **dark blue line** marks the decision boundary where the two distributions intersect.

Figure 5 (b) shows the corresponding posterior probability surfaces. Because this is a two-class case, a point is classified as C_1 whenever its posterior is higher than 0.5. The posterior surface illustrates how the model's certainty increases near each class mean and diminishes toward the decision boundary.

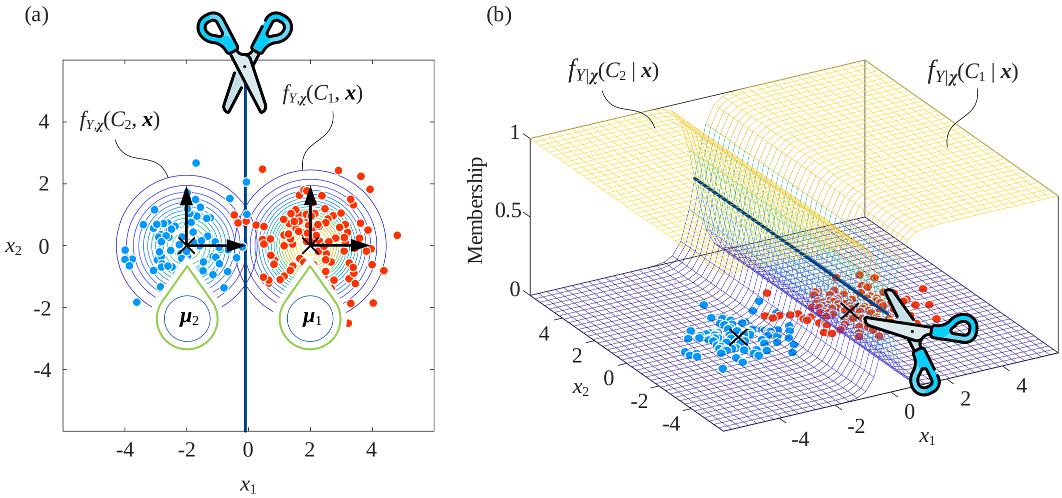

Figure 5. (a) Contours of Gaussian likelihoods for two classes with identical, diagonal covariance matrices. (b) Corresponding posterior probability surfaces.

In summary, **Category I GDA** represents the simplest and most intuitive form of discriminant analysis. Its assumptions — independent features and equal variance — lead to circular class contours and linear decision boundaries. This category provides an important conceptual foundation for understanding more complex types of GDA, where correlations or unequal variances distort these circles into ellipses and bend the decision boundary into nonlinear shapes.

17.4.2 Type II: Ellipses, Still Linear

The second type of GDA assumes that the covariance matrices of all classes are **equal** and **diagonal**, meaning that the features are statistically independent and there are no correlations between them. However, unlike Type I, the **variances along each feature dimension are not identical**. In other words, although the covariance matrix $\boldsymbol{\Sigma}_k$ is diagonal for all classes, its diagonal entries can differ, allowing each feature to have its own scale of variability.

To make the idea concrete, consider the case with two features (D = 2) and two classes (K = 2). The covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are given by

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \tag{14}$$

where $\sigma_1 \neq \sigma_2$. Since $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are identical ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$), substituting them into the discriminant boundary equation (11) eliminates the quadratic terms. As a result, **the decision boundary remains linear**, even though the data distributions may differ in variance along each feature dimension. This property links Type II GDA closely to **Linear Discriminant Analysis (LDA)**, as both produce straight-line decision boundaries.

Consider the following specific setup for Type II GDA:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.4, \quad p_Y(C_2) = 0.6, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \tag{15}$$

Figure 6 compares the surfaces of the likelihood functions $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$. The intersection of these two surfaces represents the decision boundary—the points where the model is equally likely to assign the data to either class. In this example, since $p_Y(C_2) > p_Y(C_1)$, the likelihood surface of class C_2 is higher, meaning the classifier is biased slightly toward C_2.

Looking at Figure 7(a), the contour lines of $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ appear as **ellipses** rather than circles, reflecting the unequal variances along different feature directions. Figure 8(b) shows the posterior probability surfaces, whose intersection again forms a straight-line decision boundary. This confirms that even though each feature dimension contributes differently to the overall spread of data, the separation between classes remains linear as long as their covariance matrices are equal.

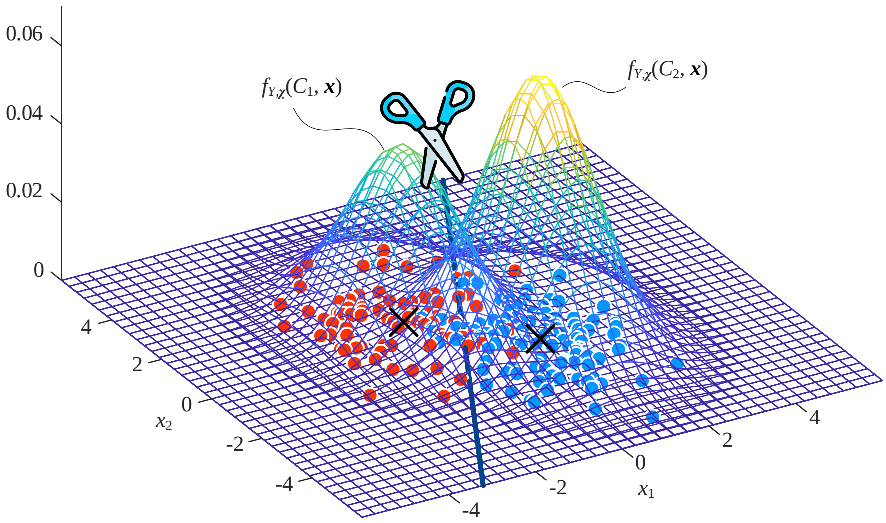

Figure 6. Likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ in Type II Gaussian Discriminant Analysis. The intersection of the two surfaces forms the decision boundary.

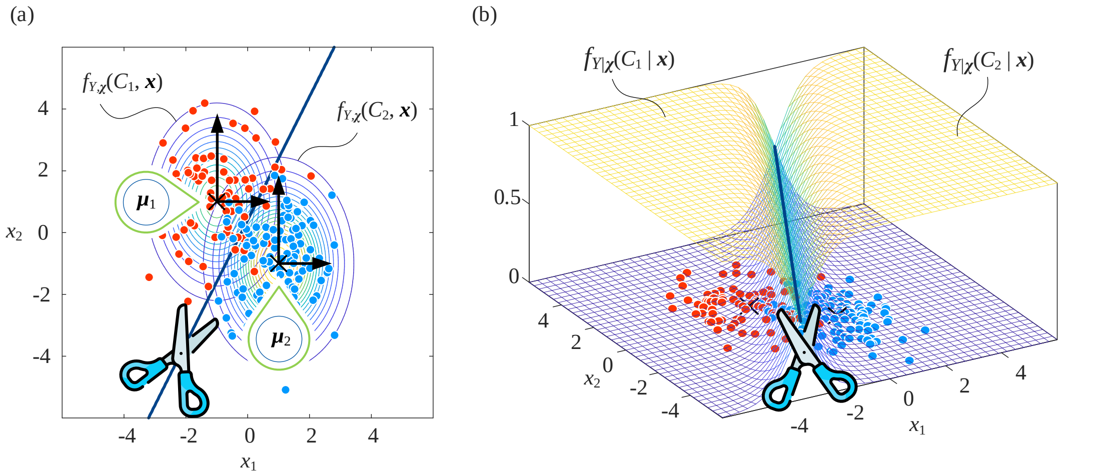

Figure 7. (a) Contour plots of the likelihood functions, forming elliptical patterns; (b) Posterior probability surfaces. The straight-line intersection represents the decision boundary in Type II GDA.

17.4.3 Type III: Correlated Features, Linear Separation

The third type of Gaussian Discriminant Analysis (GDA) assumes that the covariance matrices of all classes are **identical and unrestricted**. In other words, each class shares the same covariance matrix, but this common matrix can contain **non-zero off-diagonal elements**, meaning that the features are allowed to be correlated. Unlike the previous types, there is no requirement for the features to be independent or for the variances to be equal across dimensions. This makes Type III GDA the most general form among the models that still produce **linear decision boundaries**.

When the covariance matrices of the two classes are the same, the quadratic terms in the discriminant function cancel out. The resulting decision boundary is therefore a **linear function of the features**, even though the data distribution may appear elongated or tilted due to feature correlations. The orientation and shape of these correlated ellipses are determined by the shared covariance structure, but since both classes share the same Σ, the boundary depends only on their mean vectors and prior probabilities.

Consider the following example:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y\left(C_1\right) = 0.5, \quad p_Y\left(C_2\right) = 0.5, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 2 \end{bmatrix} \tag{16}$$

Since both classes have identical covariance matrices and equal prior probabilities, the decision boundary between them is a straight line. Figure 8 compares the likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$, where both surfaces reach the same maximum height due to equal priors.

Looking at Figure 9(a), we can see that the contour lines of $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ are **rotated ellipses**, reflecting the non-zero covariance between features (that is, the ellipses are tilted rather than aligned with the coordinate axes). Figure 9(b) shows the posterior probability surfaces, whose intersection defines the linear decision boundary.

This scenario illustrates how Gaussian models can represent correlated data while still maintaining a simple, linear separation between classes. Such cases commonly arise in real-world datasets, where features are not independent but still share a consistent covariance structure across categories.

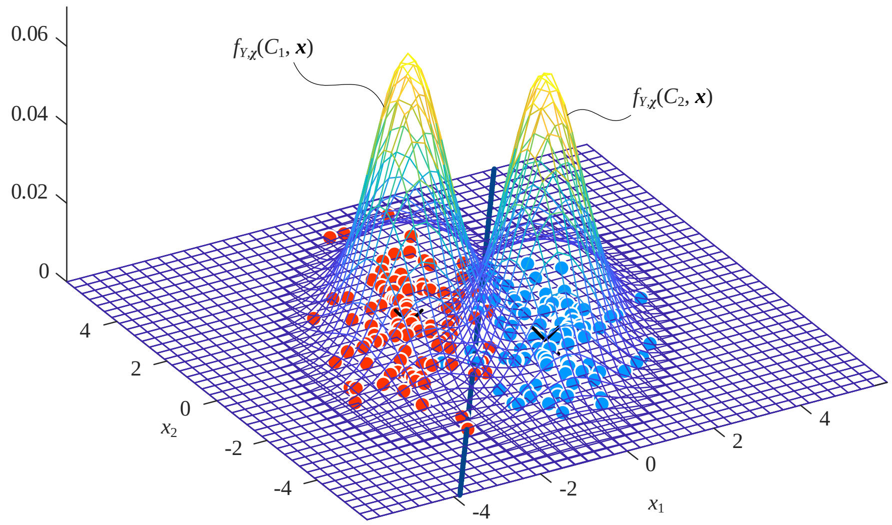

Figure 8. Likelihood surfaces $f_{Y,\chi}(C_1, \mathbf{x})$ and $f_{Y,\chi}(C_2, \mathbf{x})$ for Type III Gaussian Discriminant Analysis. The two surfaces have equal heights, and their intersection forms a linear decision boundary.

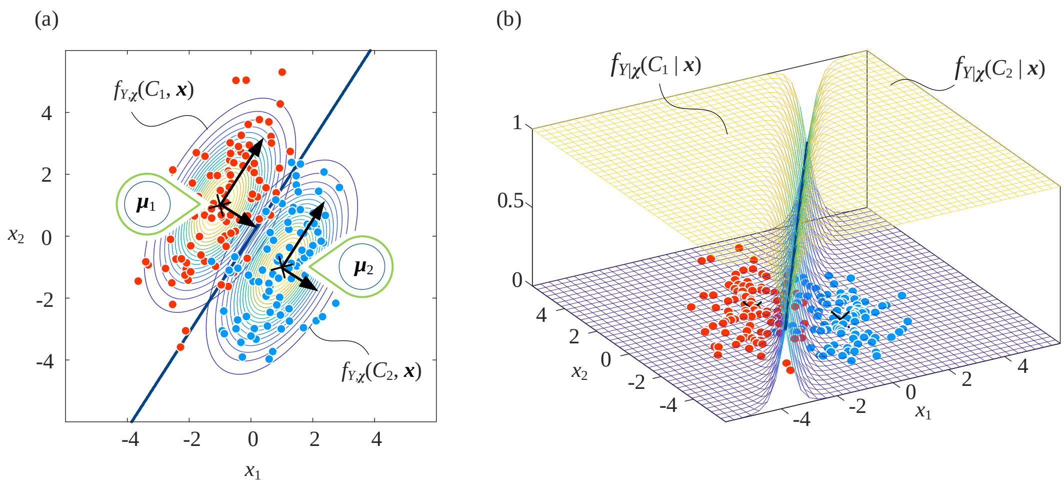

Figure 9. (a) Contour plots of showing rotated ellipses due to feature correlation; (b) Posterior probability surfaces, where their intersection represents the decision boundary.

17.4.4 Type IV: Circular, Quadratic Boundaries

The fourth type of Gaussian Discriminant Analysis (GDA) assumes that the features within each class are **conditionally independent**, meaning that the covariance matrix is **diagonal**—all off-diagonal elements are zero. This reflects the idea that each feature contributes independently to the classification decision, without interacting with others. However, unlike the first type of GDA, the covariance matrices of different classes are **not identical**. Each class can have its own variance magnitude, but within a class, all features share the **same variance value** (the diagonal elements are equal).

This setup produces circular probability density contours for each class, since equal variance in all directions implies spherical symmetry. Yet because the variances differ between classes, the circular contours of the two Gaussian distributions have different radii. As a result, the decision boundary between the two classes becomes a **quadratic curve**, specifically a **circle**. This is a key distinction from the first type of GDA, where both classes share the same covariance matrix, resulting in a linear boundary.

Consider the following example:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.3, \quad p_Y(C_2) = 0.7, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \tag{17}$$

Figure 10 compares the likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$. When projected onto the plane, the decision boundary appears as a **perfect circle**, as shown in Figure 11(a). Because the second class has a larger variance, its contours are more spread out, meaning it covers a broader region of the feature space with lower peak height.

Figure 11(b) shows the posterior probability surfaces. The decision boundary is formed where these two surfaces intersect—at points where the model is equally uncertain about class membership. Intuitively, the boundary tends to shift toward the class with the **smaller variance** or **lower prior probability**, as these regions are more concentrated.

This type of GDA is one of the simplest forms of **Quadratic Discriminant Analysis (QDA)**, where each class has its own covariance matrix but still maintains conditional independence among features. It provides a smooth nonlinear boundary that better adapts to data distributions with differing spreads.

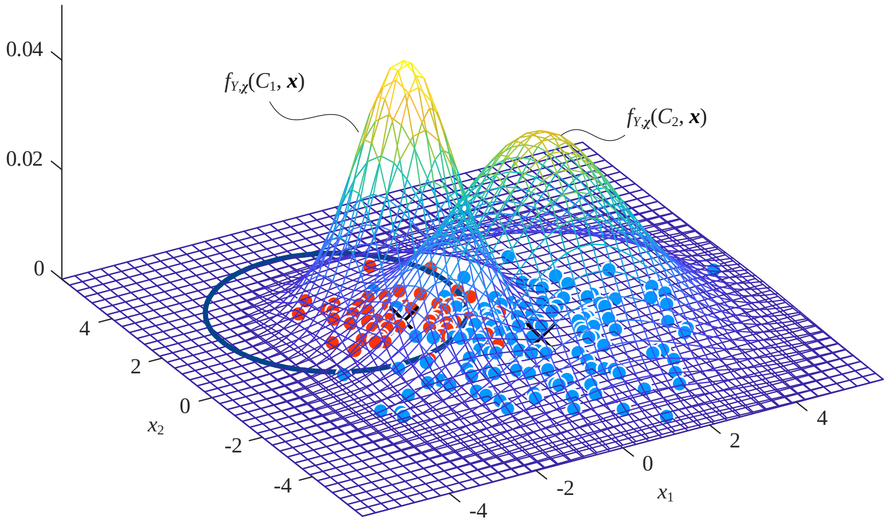

Figure 10. Likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ for Type IV Gaussian Discriminant Analysis.

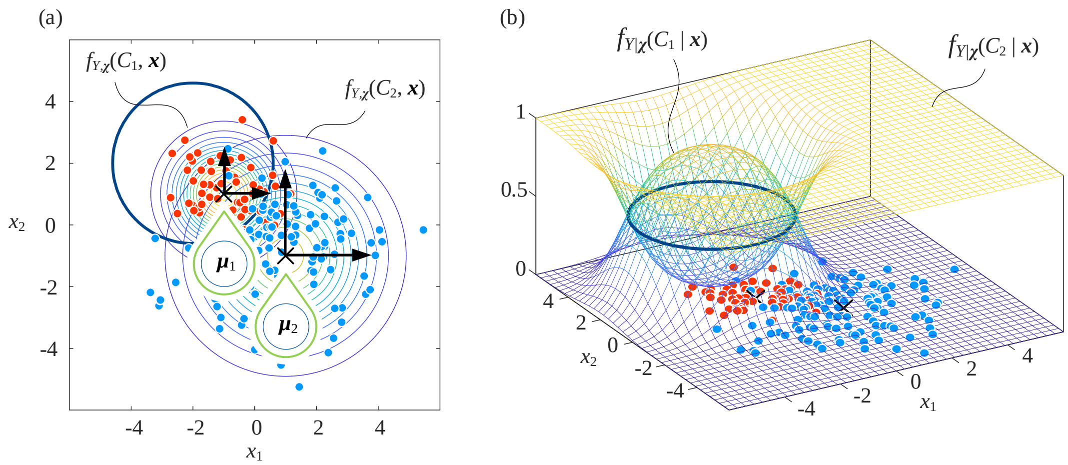

Figure 11. (a) Projection of the two Gaussian distributions on the plane showing a circular decision boundary. (b) Comparison of posterior probability surfaces, where their intersection defines the circular decision boundary.

17.4.5 Type V: Elliptical, Quadratic Boundaries

The fifth type of Gaussian Discriminant Analysis (GDA) allows **different covariance matrices for each class**, but each is still **diagonal**, meaning that the features within a class are **conditionally independent**. Unlike Type IV, the diagonal elements of $\boldsymbol{\Sigma}_k$ are not constrained to be equal, so each feature can have its own variance within a class. This flexibility allows the model to handle classes where different features vary at different scales.

Importantly, Type V GDA is equivalent to **Gaussian Naïve Bayes**, because the diagonal covariance matrices reflect the assumption that features are independent given the class. The decision boundary can be derived from the discriminant functions and, due to the absence of off-diagonal terms, **there are no cross-product terms involving** x_1x_2. This results in **regular conic curves** for the decision boundary, such as ellipses, hyperbolas, or parabolas, depending on the relative values of the variances and class priors.

Consider the following example:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.4, \quad p_Y(C_2) = 0.6, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \tag{18}$$

Figure 12 shows the likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$. The decision boundary forms a regular conic curve, which can be visually identified as the intersection of the two surfaces.

Figure 13(a) presents the contour plots of the likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$. The contours are elliptical, reflecting the different variances along the feature axes for each class. Figure 13(b) shows the posterior probability surfaces, where the intersection of the surfaces defines the decision boundary.

This type of GDA illustrates how feature independence combined with class-specific variances leads to flexible quadratic boundaries that can adapt to differences in spread across classes, which is a hallmark of Gaussian Naïve Bayes.

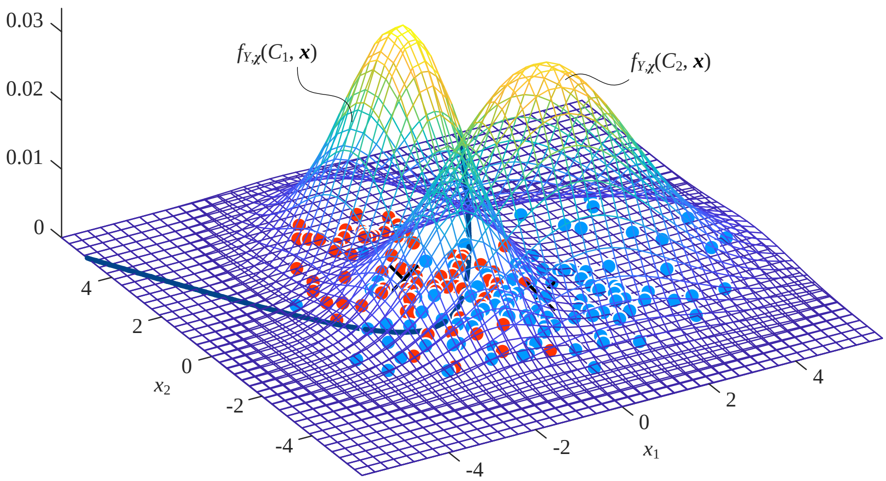

Figure 12. Likelihood surfaces $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ for Type V Gaussian Discriminant Analysis. The intersection forms a regular conic decision boundary.

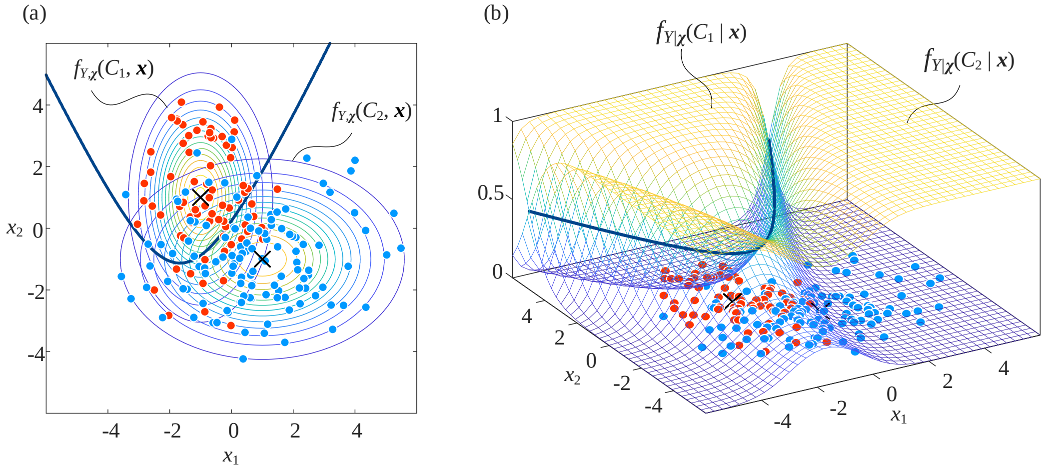

Figure 13. (a) Contour plots of $f_{Y,\chi}(C_1, \boldsymbol{x})$ and $f_{Y,\chi}(C_2, \boldsymbol{x})$ showing elliptical shapes reflecting class-specific variances; (b) Posterior probability surfaces, with their intersection representing the decision boundary.

17.4.6 Type VI: Fully General, Flexible Boundaries

The sixth type of Gaussian Discriminant Analysis (GDA) is the most general form, where **no restrictions are imposed on the covariance matrices $\boldsymbol{\Sigma}_k$**. Each class can have an arbitrary covariance matrix, allowing for **any orientation, scale, and correlation among features**. As a result, the posterior probability density function (PDF) contours of each class can take the form of **rotated, stretched, or skewed ellipses**. This flexibility allows the model to handle highly complex data distributions.

The decision boundaries in Type VI GDA can take a wide variety of forms, including **single straight lines, parallel lines, ellipses, hyperbolas, parabolas, or more complex metamorphosing quadratic curves**. The specific shape of the boundary depends on the covariance matrices of the classes and their prior probabilities, as well as the location of the mean vectors.

Under the following parameter settings,

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \begin{cases} p_Y(C_1) = 0.2 \\ p_Y(C_2) = 0.8 \end{cases}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 3 & 1.8 \\ 1.8 & 3 \end{bmatrix} \tag{19}$$

the decision boundary forms an **ellipse**, as shown in Figure 14. This occurs when the class covariance matrices are positive definite and similar in scale but oriented differently.

Alternatively, the decision boundary can become a **hyperbola** under different parameter conditions, as shown in Figure 15.

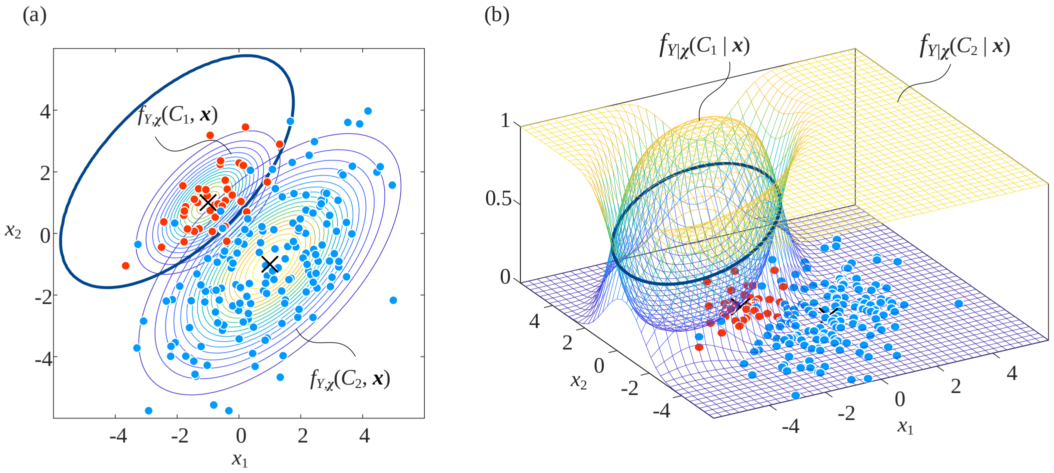

Figure 14. Elliptical decision boundary in Type VI Gaussian Discriminant Analysis under specific covariance and mean parameters.

Hyperbolic boundaries occur when the covariance matrices differ significantly in scale or orientation, given the following

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{cases} p_Y(C_1) = 0.4 \\ p_Y(C_2) = 0.6 \end{cases}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & -0.6 \\ -0.6 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} \tag{20}$$

Which creates saddle-shaped regions where the classifier switches between classes (see Figure 15).

These examples illustrate the full expressive power of Gaussian discriminant analysis. By allowing unrestricted covariance matrices, Type VI GDA can adapt to highly irregular class distributions, making it a foundation for **Quadratic Discriminant Analysis (QDA)**. This general form naturally leads to the distinction between **Linear Discriminant Analysis (LDA)**, which assumes shared covariance matrices, and QDA, which allows class-specific covariance structures.

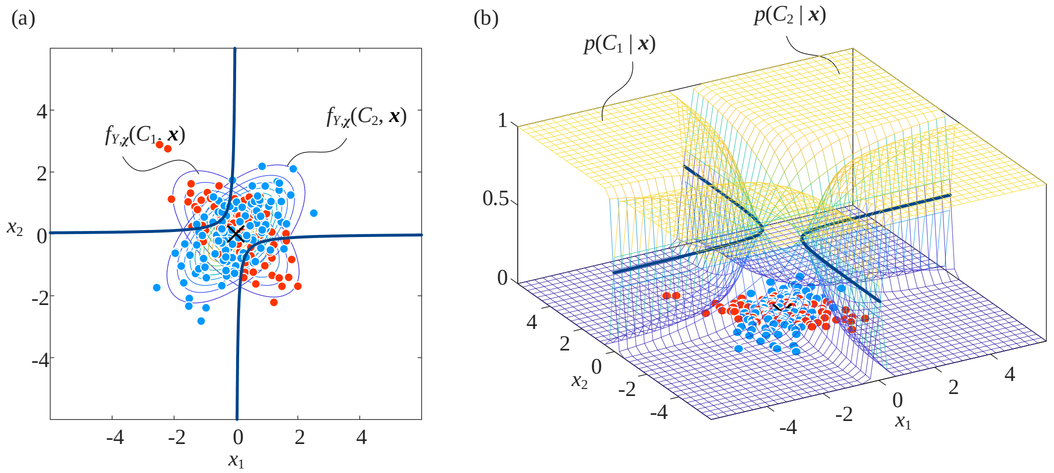

Figure 15. Hyperbolic decision boundary in Type VI Gaussian Discriminant Analysis with distinct class covariance matrices.

17.5 Practical Implementation with Python

In practice, Gaussian Discriminant Analysis (GDA) can be implemented directly using **scikit-learn**. The sklearn.discriminant_analysis.LinearDiscriminantAnalysis function corresponds to **Linear Discriminant Analysis (LDA)**, which is based on the **third type of GDA** introduced in this chapter. In this setting, the covariance matrix is assumed to be the same for all classes, but there are no restrictions on feature correlations or variance magnitudes. Because the shared covariance simplifies the discriminant function, LDA produces **linear decision boundaries**, even when the features are correlated.

On the other hand, sklearn.discriminant_analysis.QuadraticDiscriminantAnalysis implements **Quadratic Discriminant Analysis (QDA)**. This algorithm corresponds to the **sixth type of GDA**, where there are **no restrictions on Σ_k**. Each class can have its own full covariance matrix, which allows QDA to model **arbitrary elliptical or quadratic class boundaries**.

In essence, QDA is a supervised learning algorithm that estimates both the mean vector and the full covariance matrix for each class. By doing so, it accommodates more complex class distributions where the assumption of shared covariance is not valid.

Figure 16 and Figure 17 illustrate the results of classifying the famous **Iris dataset** using LDA and QDA. The LDA decision boundaries are straight lines separating the classes, while the QDA boundaries are curved, reflecting differences in class covariance. These visualizations highlight how the choice of GDA type influences the shape of the classification boundaries and the model's flexibility.

This book will provide a dedicated chapter on LDA later, exploring its derivation, geometric intuition, and applications in detail.

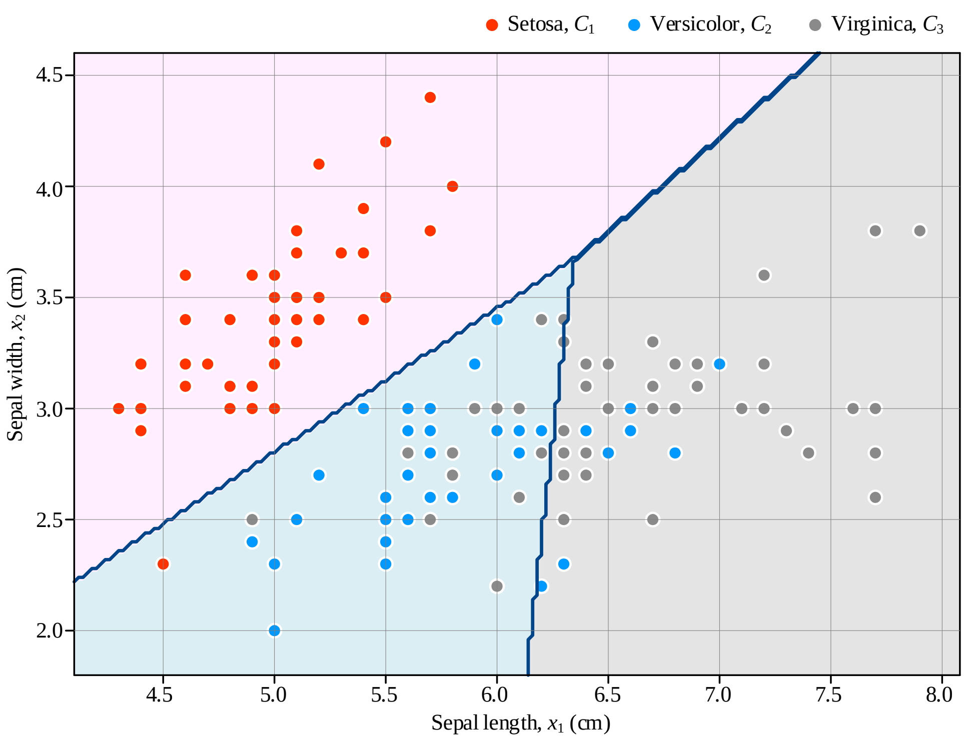

Figure 16. Linear Discriminant Analysis (LDA) applied to classify iris flowers, showing linear decision boundaries. Figure generated by Ch17_01_Gaussian_Discriminant_Analysis.ipynb.

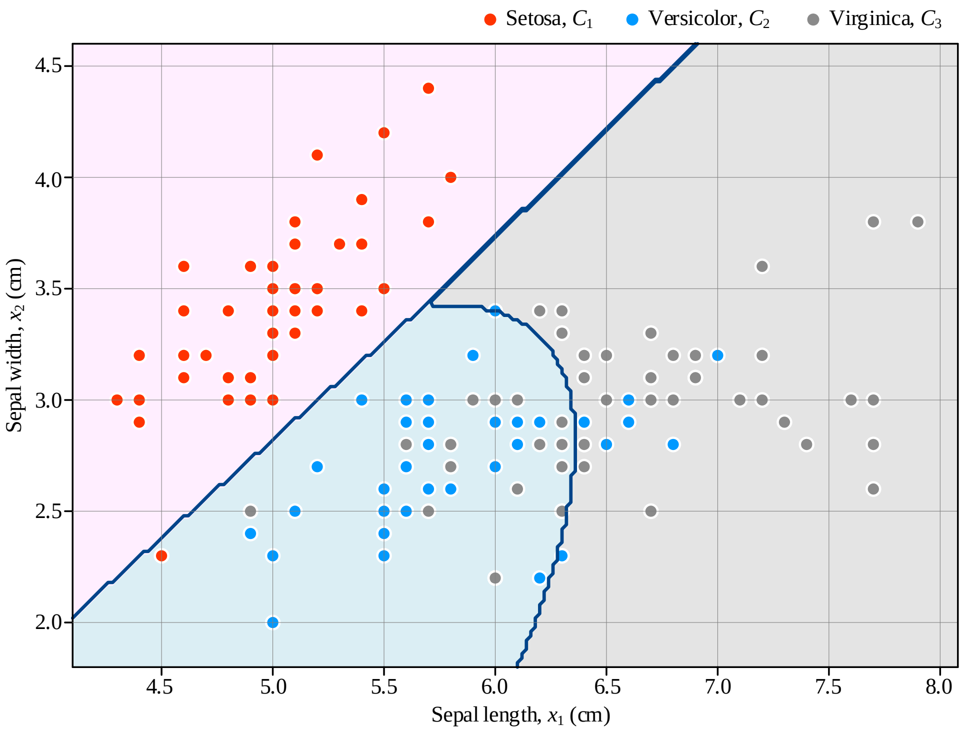

Figure 17. Quadratic Discriminant Analysis (QDA) applied to classify iris flowers, showing curved, quadratic decision boundaries. Figure generated by Ch17_01_Gaussian_Discriminant_Analysis.ipynb.

17.6 Conclusion

Gaussian Discriminant Analysis is a probabilistic classification method that models each class as a multivariate Gaussian distribution. It estimates the mean and covariance matrix for each class and assigns new data points to the class with the highest posterior probability.

The shape of the Gaussian distributions, determined by the covariance matrices, influences the form of the decision boundary. Linear Discriminant Analysis assumes all classes share the same covariance matrix, producing straight boundaries, while Quadratic Discriminant Analysis allows class-specific covariance matrices, resulting in curved boundaries.

GDA has six types based on assumptions about covariance structure, ranging from independent features with equal variance to fully general covariance with correlations. The decision boundary can be linear, circular, elliptical, or more complex depending on these assumptions. GDA also relates to Gaussian Naïve Bayes and Gaussian Mixture Models.

Practical implementation in Python uses scikit-learn's LDA and QDA functions, which illustrate how different covariance assumptions affect classification results.