

10

Gaussian Discriminant Analysis

高斯判别分析

假设后验概率为高斯分布，最小化分类错误



唯有勇敢者，才能洞见科学的壮美。

The enchanting charms of this sublime science reveal themselves in all their beauty only to those who have the courage to go deeply into it.

—— 卡尔·弗里德里希·高斯 (Carl Friedrich Gauss) | 德国数学家、物理学家、天文学家 | 1777 ~ 1855



- ▶ `matplotlib.pyplot.contour()` 绘制等高线图
- ▶ `matplotlib.pyplot.contourf()` 绘制填充等高线图
- ▶ `matplotlib.pyplot.scatter()` 绘制散点图
- ▶ `numpy.array()` 创建 `array` 数据类型
- ▶ `numpy.c_()` 按列叠加两个矩阵
- ▶ `numpy.linspace()` 产生连续均匀向量数值
- ▶ `numpy.meshgrid()` 创建网格化数据
- ▶ `numpy.r_()` 按行叠加两个矩阵
- ▶ `numpy.ravel()` 将矩阵扁平化
- ▶ `seaborn.scatterplot()` 绘制散点图
- ▶ `sklearn.datasets.load_iris()` 加载鸢尾花数据集
- ▶ `sklearn.discriminant_analysis.LinearDiscriminantAnalysis` 线性判别分析函数
- ▶ `sklearn.discriminant_analysis.QuadraticDiscriminantAnalysis` 二次判别分析函数

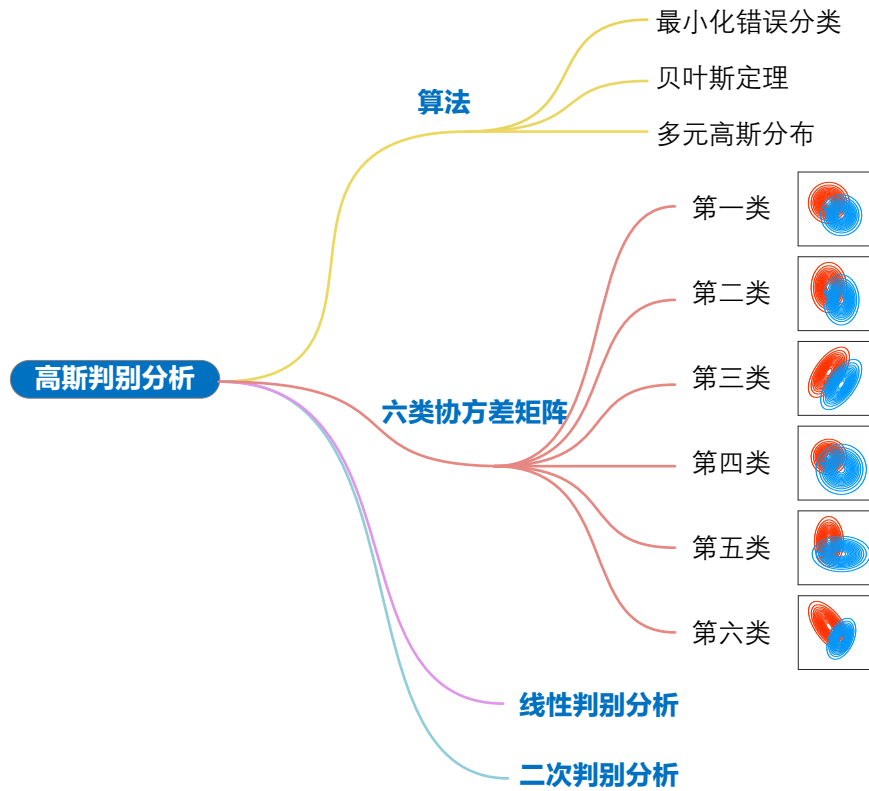
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com



本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

10.1 又见高斯

本章介绍**高斯判别分析** (Gaussian Discriminant Analysis, GDA)。高斯判别分析中，似然概率采用高斯多元分布估计，这便是其名称的由来。GDA 是一种监督学习算法，用于分类和判别问题。它基于假设，即每个类别的数据都服从高斯分布。具体来说，对于每个类别，GDA 通过估计该类别的均值和协方差矩阵来建模该类别的高斯分布。在训练过程中，算法学习这些参数，并使用它们来计算给定输入数据点属于哪个类别的后验概率。在测试时，算法使用这些后验概率来进行分类。

原理

图 1 所示为高斯判别分析的原理，大家可能已经发现六幅子图椭圆等高线呈现不同形态，这代表着高斯多元分布展现不同特点。

此外，图中决策边界包含一次函数和圆锥曲线。相信图 1 各种细节已经引起了大家好奇心，本章将为大家一一解密。

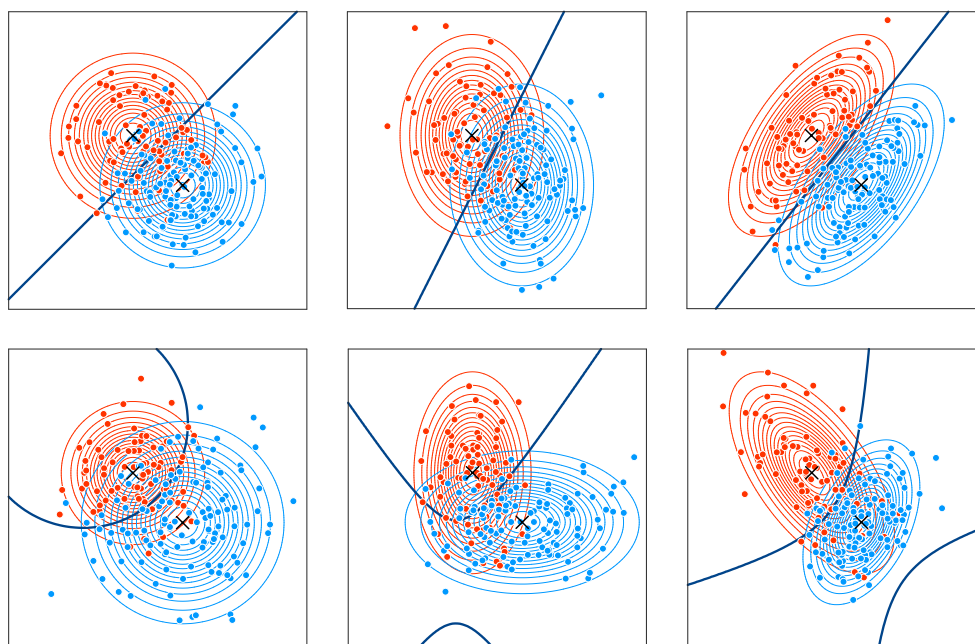


图 1. 高斯判别分析原理

分类

高斯判别分析，又细分为**线性判别分析** (Linear Discriminant Analysis, LDA) 和**二次判别分析** (Quadratic Discriminant Analysis, QDA)。

高斯判别分析算法得到的决策边界有解析解。从它们各自的名字上就可以看出，LDA 和 QDA 决策边界分别为线性式和二次式。

此外，二次判别分析 QDA 和上一章介绍的**高斯朴素贝叶斯** (Gaussian Naïve Bayes) 有着紧密关系。高斯判别分析分类算法和本书后续介绍的**高斯混合模型** (Gaussian Mixture Model, GMM) 也有千丝万缕的联系。

优化问题

高斯判别分析优化目标如下，**预测分类** (predicted classification) \hat{y} 可以通过下式求得：

$$\hat{y} = \arg \min_{C_m} \sum_{k=1}^K f_{Y|X}(C_k | \mathbf{x}) \cdot c(C_m | C_k) \quad (1)$$

其中， K 为类别数量， m 和 k 均为类别序数—— $1, 2, \dots, K$ 。

$f_{Y|X}(C_k | \mathbf{x})$ 为任意一点 \mathbf{x} 被预测分类为 C_k 类的**后验概率** (posterior)。

$c(C_m | C_k)$ 为惩罚因子，代表 \mathbf{x} 正确分类为 C_k ，但被预测分类为 C_m 对应的代价，具体计算如下：

$$c(C_m | C_k) = \begin{cases} 1 & m \neq k \\ 0 & m = k \end{cases} \quad (2)$$

$m \neq k$ 时，也就是当某一点真实类别为 C_k ，但是却被错误地分类为 C_m 时， $c(C_m | C_k) = 1$ 。而当分类正确时，即 $m = k$ ， $c(C_m | C_k) = 0$ 。

(1) 中蕴含着高斯判别分析重要的思路——最小化错误分类。这个思路和朴素贝叶斯恰好相反。

计算后验概率

根据**贝叶斯定理** (Bayes theorem)，后验概率 $f_{Y|X}(C_k | \mathbf{x})$ 可以通过下式计算获得：

$$\underbrace{f_{Y|X}(C_k | \mathbf{x})}_{\text{Posterior}} = \frac{\overbrace{f_{X,Y}(\mathbf{x}, C_k)}^{\text{Joint}}}{\underbrace{f_X(\mathbf{x})}_{\text{Evidence}}} = \frac{\overbrace{f_{X|Y}(\mathbf{x} | C_k)}^{\text{Likelihood}} \overbrace{p_Y(C_k)}^{\text{Prior}}}{\underbrace{f_X(\mathbf{x})}_{\text{Evidence}}} \quad (3)$$

证据因子 $f_X(\mathbf{x})$ 可以通过下式求得：

$$f_X(\mathbf{x}) = \sum_{k=1}^K \overbrace{f_{X|Y}(\mathbf{x} | C_k)}^{\text{Likelihood}} \overbrace{p_Y(C_k)}^{\text{Prior}} \quad (4)$$

和本书之前介绍的朴素贝叶斯一样，证据因子 $f_X(\mathbf{x})$ 也可以不求，后验 \propto 似然 \times 先验：

$$\underbrace{f_{Y|X}(C_k | \mathbf{x})}_{\text{Posterior}} \propto \underbrace{f_{X,Y}(\mathbf{x}, C_k)}_{\text{Joint}} \quad (5)$$

证据因子 $f_X(\mathbf{x})$ 相当于对 $f_{X,Y}(\mathbf{x}, C_k)$ 归一化处理。

引入多元高斯分布

高斯判别分析假设，似然概率 $f_{x|y}(\mathbf{x}|C_k)$ 服从多元高斯分布，因此 $f_{x|y}(\mathbf{x}|C_k)$ 具体表达式如下：

$$f_{x|y}(\mathbf{x}|C_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right)}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} \quad (6)$$

其中， D 为特征数量， \mathbf{x} 为列向量， $\boldsymbol{\mu}_k$ 为 C_k 类数据质心位置， $\boldsymbol{\Sigma}_k$ 为 C_k 类样本协方差矩阵。

可以说，此处便是高斯判别分析和高斯朴素贝叶斯分道扬镳之处！

10.2 六类协方差矩阵

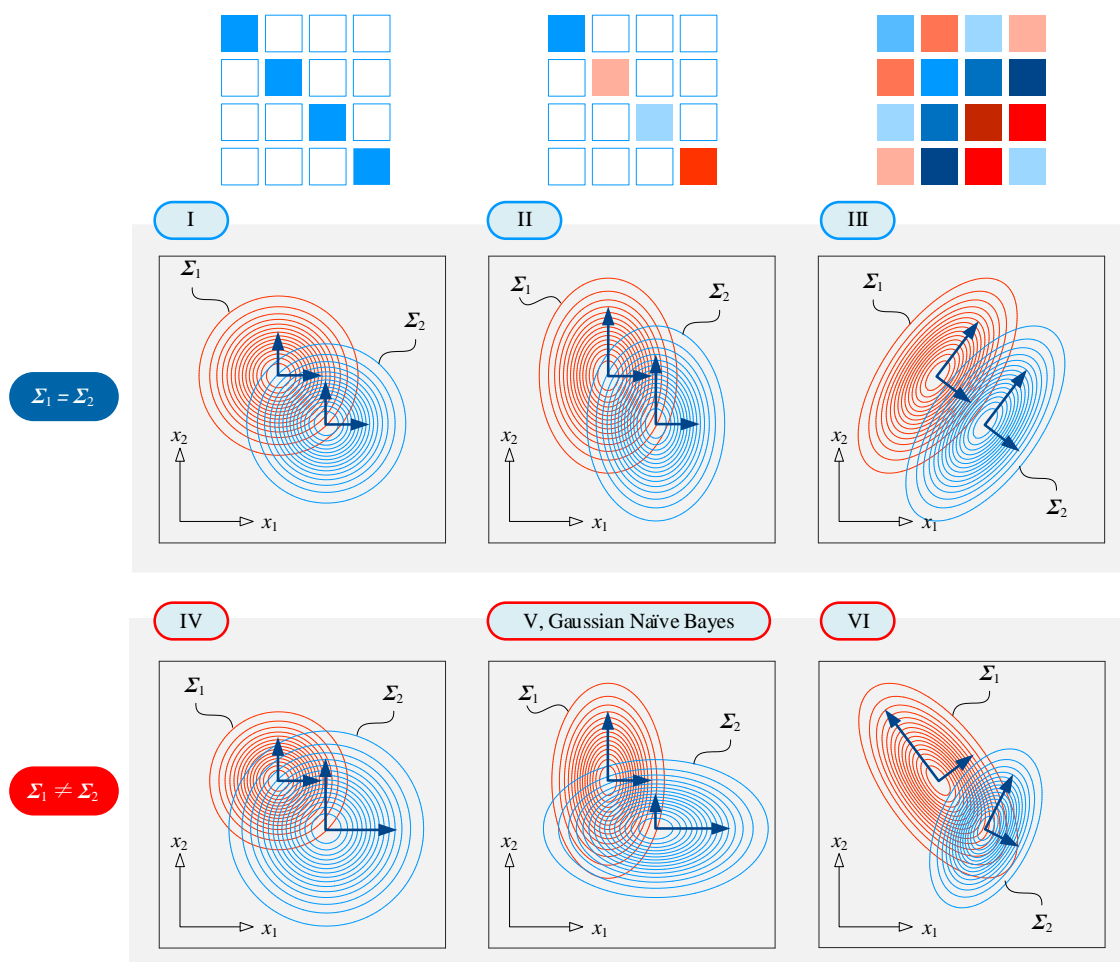
高斯朴素贝叶斯，假设“特征条件独立”。

然而，高斯判别分析，根据 $\boldsymbol{\Sigma}_k$ 形态将算法分成六个类别。这六个类别中，有些特征条件独立，有些特征满足特殊条件。表 1 总结了六类高斯判别分析对应的协方差矩阵特点。

表 1. 根据协方差矩阵特点将高斯判别分析问题分成 6 类

	$\boldsymbol{\Sigma}_k$	特征方差 ($\boldsymbol{\Sigma}_k$ 对角线元素)	$\boldsymbol{\Sigma}_k$ 特点	似然概率 PDF 等高线	决策边界
第一类	相同	相同	对角阵	正圆，形状相同	直线
第二类		不限制	(特征条件独立)	正椭圆，形状相同	
第三类			非对角阵	任意椭圆，形状相同	
第四类	不同	相同	对角阵	正圆	正圆
第五类		不限制	(特征条件独立)	正椭圆	正圆锥曲线
第六类			非对角阵	任意椭圆	圆锥曲线

图 2 所示为六大类判别分析高斯分布椭圆形状。

图 2. 六大类判别分析高斯分布椭圆形状, $K=2$, $D=2$

前三类：决策边界为直线

前三类 (I、II 和 III) GDA 有一个共同特点，假设各个类别协方差矩阵 Σ_k ，完全一致；因此，这三类的决策边界为直线，因此它们被称作线性判别分析 LDA。这一点，本章后续将展开讲解；这里先给大家结论，希望读者学完本章回过头来再看一遍。

- 第一类 GDA 的重要特点是， Σ_k 为对角阵 (除主对角线之外元素为 0)，即特征之间“条件独立”。并且， Σ_k 对角线元素相同，即假设各个特征方差相同。因此，图 2 所示第一类 GDA 中，红色和蓝色 PDF 等高线为正圆，且大小相同。
- 第二类 GDA， Σ_k 为对角阵，特征条件独立；但是，对 Σ_k 对角线元素大小不做限制。如图 2 所示，红色和蓝色 PDF 等高线为大小相等的正椭圆。
- 第三类 GDA，仅仅假设各个类别协方差矩阵 Σ_k ，完全一致。方差和条件独立不做任何限制。如图 2 所示，红色和蓝色 PDF 等高线为大小相等旋转椭圆。

后三类：决策边界为二次曲线

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

后三类 (IV、V 和 VI) GDA，各个类别协方差矩阵 Σ_k 不相同；这三类 GDA 决策边界为二次曲线，因此被称作二次判别分析 QDA。

- 第四类 GDA， Σ_k 为对角阵，也假设特征之间“条件独立”；同时假设每个类别协方差矩阵 Σ_k 对角线元素相同，即假设类别内样本数据各个特征方差相同。图 2 所示第四类 GDA 中，红色和蓝色 PDF 等高线为正圆，但是大小不同。
- 第五类 GDA， Σ_k 为对角阵，特征条件独立；不限制 Σ_k 对角线元素大小；如图 2 所示，红色和蓝色 PDF 等高线为正椭圆，但是大小不同。可以发现，第五类 (V) 对应高斯朴素贝叶斯分类。从几何图像上解释，高斯朴素贝叶斯中，条件概率曲面等高线为正椭圆。
- 第六类 GDA，对 Σ_k 不做任何限制。如图 2 所示，红色和蓝色 PDF 等高线旋转椭圆，大小不等。

此外，非监督学习中高斯混合模型 (Gaussian Mixture Model, GMM) 也会使用到本章介绍的协方差矩阵特点和决策边界关系。

10.3 决策边界解析解

本节推导决策边界解析解一般形式。

判别函数

定义 C_k 类判别函数 $g_k(\mathbf{x})$ 如下：

$$\begin{aligned} g_k(\mathbf{x}) &= \ln(f_{\mathbf{x}|Y}(\mathbf{x}, C_k)) = \ln(f_{\mathbf{x}|Y}(\mathbf{x}|C_k) p_Y(C_k)) \\ &= \ln\left(\frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right)}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} p_Y(C_k)\right) \\ &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k) - \frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_k| + \ln p_Y(C_k) \end{aligned} \quad (7)$$

判别函数就是联合概率密度函数的自然对数。白话说，这个运算是为了去掉多元高斯分布中的 $\exp()$ 。

两特征、两分类问题

为了方便讨论，本章以两个特征 ($D = 2$) 二分类 ($K = 2$) 为例。 C_1 和 C_2 的判别函数分别为：

$$\begin{cases} g_1(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) - \frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_1| + \ln p_Y(C_1) \\ g_2(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_2^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) - \frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_2| + \ln p_Y(C_2) \end{cases} \quad (8)$$

其中

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (9)$$

对于二分类 ($K=2$) 问题, 高斯判别分析的决策边界取决于如下等式:

$$g_1(\mathbf{x}) = g_2(\mathbf{x}) \quad (10)$$

将 (8) 代入 (10) 并整理得到决策边界对应的解析式:

$$\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_2^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) = \ln p_Y(C_1) - \ln p_Y(C_2) + \left(\frac{1}{2} \ln |\boldsymbol{\Sigma}_2| - \frac{1}{2} \ln |\boldsymbol{\Sigma}_1| \right) \quad (11)$$

容易发现上式为二次式, 甚至是一次式。决策边界解析解次数和具体参数, 和两个协方差矩阵 ($\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$) 取值直接相关。而先验概率 $p_Y(C_1)$ 和 $p_Y(C_2)$ 影响常数项。

决策边界形态

图 3 所示为用 Streamlit 搭建的展示高斯判别分析边界的 App。

Streamlit_Bk7_Ch10_01.py 搭建图 3 中 App, 请大家自行学习。

图 4 所示为各种高斯判别分析 QDA 二分类常见决策边界形态。观察图 4 可以发现, 决策边界可以是直线、正圆、椭圆、抛物线、双曲线, 以及各种蜕化二次曲线。下一节开始讲逐个讲解各类 QDA。

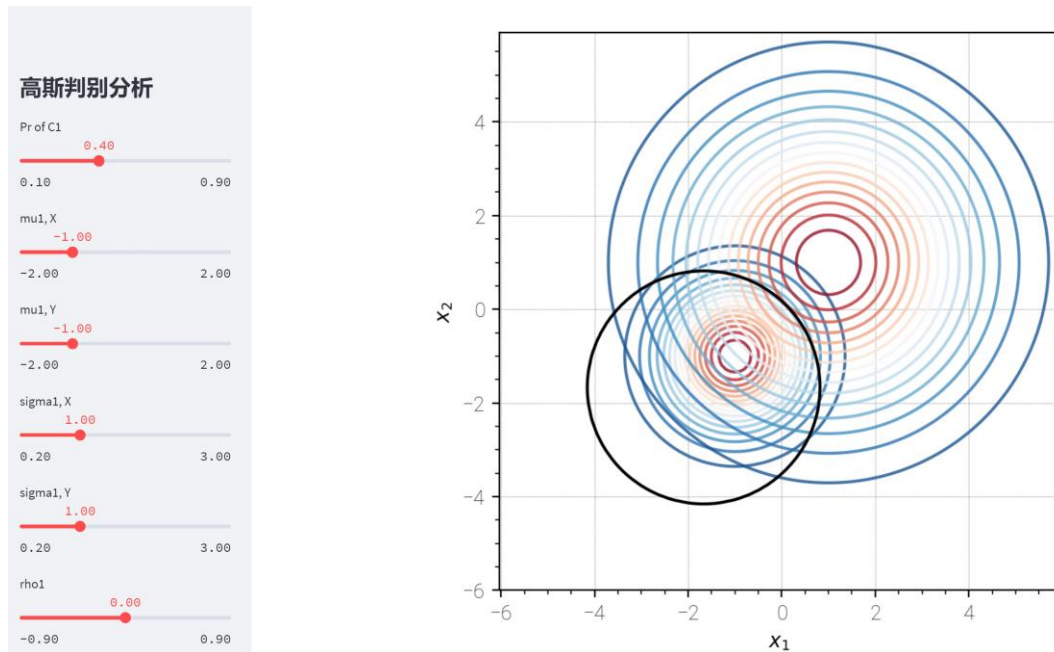



图 3. 展示高斯判别分析边界的 App, Streamlit 搭建 |  Streamlit_Bk7_Ch10_01.py

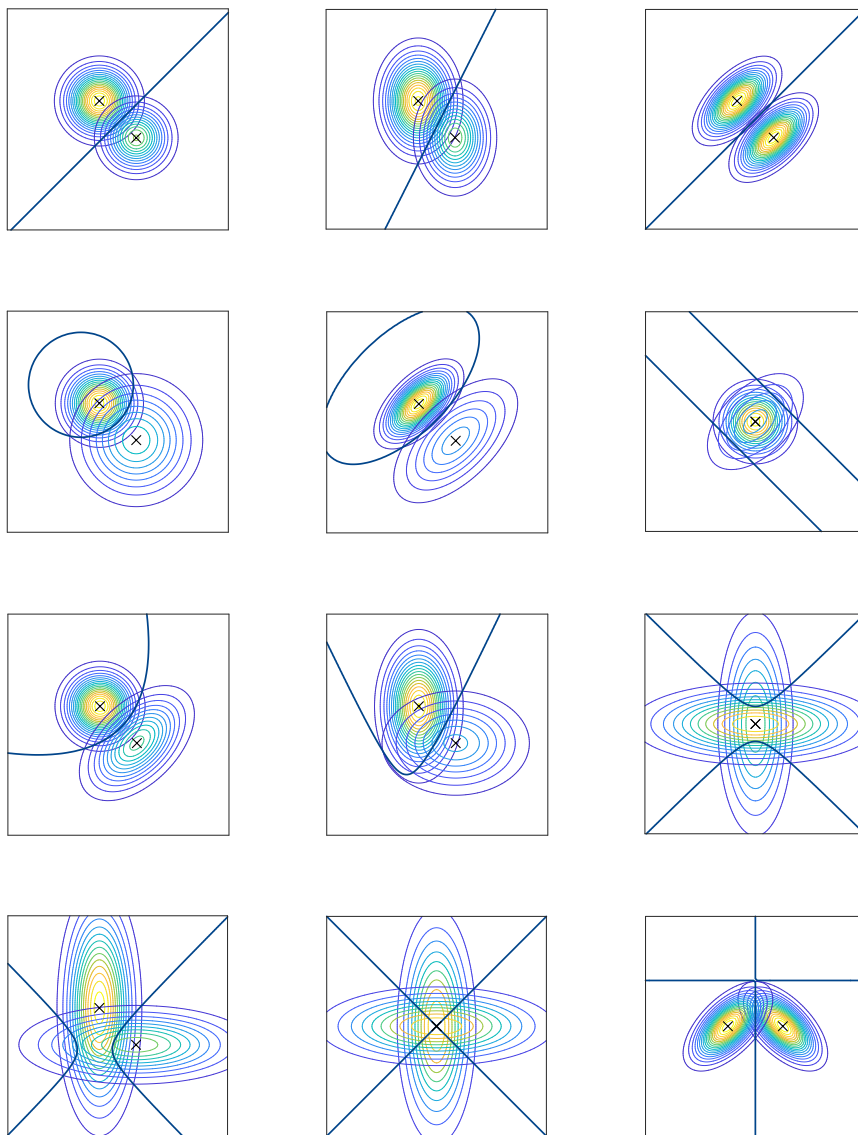


图 4.判别分析常见决策边界

10.4 第一类

第一类高斯判别分析，假设数据特征条件独立。协方差矩阵 Σ_k 为对角阵，即相关系数为 0，且假设 Σ_k 对角元素相同，即特征方差相同。

两特征、两分类

如下例 ($K=2$ 且 $D=2$):

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \sigma^2 I \quad (12)$$

两个协方差矩阵 (Σ_1 和 Σ_2) 的逆矩阵如下:

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载: <https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱: jiang.visualize.ml@gmail.com

$$\Sigma_1^{-1} = \Sigma_2^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{\mathbf{I}}{\sigma^2} \quad (13)$$

将 (13) 代入 (11) 得到：

$$\begin{aligned} \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \frac{\mathbf{I}}{\sigma^2} (\mathbf{x} - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \frac{\mathbf{I}}{\sigma^2} (\mathbf{x} - \boldsymbol{\mu}_2) &= \ln p_Y(C_1) - \ln p_Y(C_2) \\ \Rightarrow (\mathbf{x} - \boldsymbol{\mu}_1)^T (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_2)^T (\mathbf{x} - \boldsymbol{\mu}_2) &= 2\sigma^2 (\ln p_Y(C_1) - \ln p_Y(C_2)) \end{aligned} \quad (14)$$

决策边界

整理得到的决策边界解析解如下：

$$(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \mathbf{x} - \left[\sigma^2 (\ln p_Y(C_1) - \ln p_Y(C_2)) + \frac{1}{2}(\boldsymbol{\mu}_2^T \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^T \boldsymbol{\mu}_1) \right] = 0 \quad (15)$$

回忆《矩阵力量》介绍的空间直线矩阵运算表达式：

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (16)$$

\mathbf{w} 为该直线法向量，也是梯度向量。比较 (15) 和 (16) 可以得到直线参数：

$$\begin{cases} \mathbf{w} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \\ b = - \left[\sigma^2 (\ln p_Y(C_1) - \ln p_Y(C_2)) + \frac{1}{2}(\boldsymbol{\mu}_2^T \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^T \boldsymbol{\mu}_1) \right] \end{cases} \quad (17)$$

先验概率

特别的，当 $p_Y(C_1) = p_Y(C_2)$ ，且 $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ 时，代入 (15) 可以得到：

$$\begin{aligned} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_2^T \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^T \boldsymbol{\mu}_1) &= 0 \\ \Rightarrow (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T (\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1) &= 0 \\ \Rightarrow (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \left[\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1) \right] &= 0 \end{aligned} \quad (18)$$

特别提醒读者的是，式 (18) 中 $(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T$ 不能消去。

当 $p_Y(C_1) = p_Y(C_2)$ 时，观察 (18) 可以发现，决策边界直线通过 $\boldsymbol{\mu}_1$ 和 $\boldsymbol{\mu}_2$ 两点中点 $(\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1)/2$ ，并垂直于两点连线，对应 $(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ 向量。也就是说，决策边界为 C_1 和 C_2 类质心 $\boldsymbol{\mu}_1$ 和 $\boldsymbol{\mu}_2$ 中垂线。

再次注意，因为 $p_Y(C_1) = p_Y(C_2)$ ，所以决策边界距离 C_1 和 C_2 两类样本数据质心 ($\boldsymbol{\mu}_1$ 和 $\boldsymbol{\mu}_2$) 等距。给大家提一个小问题，如果 $p_Y(C_1) > p_Y(C_2)$ ，决策边界更靠近 C_1 ，还是 C_2 ？

举个例子

采用如下具体数值讨论第一类高斯判别分析：

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad p_Y(C_1) = 0.6, \quad p_Y(C_2) = 0.4, \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (19)$$

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

图 5 直接比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面高度，任意一点 \mathbf{x} ，如果 $f_{Y|X}(C_1, \mathbf{x}) > f_{Y|X}(C_2, \mathbf{x})$ ，则该点分类可以被判定为 C_1 ；反之， $f_{Y|X}(C_1, \mathbf{x}) < f_{Y|X}(C_2, \mathbf{x})$ ，则该点分类可以被判定为 C_2 。 $f_{Y|X}(C_1, \mathbf{x}) = f_{Y|X}(C_2, \mathbf{x})$ 处便是决策边界。

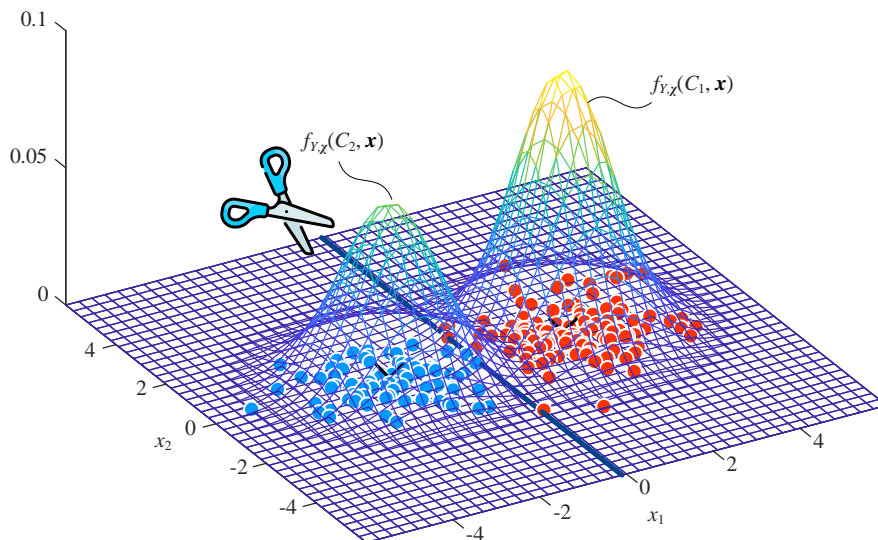


图 5. 第一类高斯判别分析，比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面， $\Sigma_1 = \Sigma_2 = [1 \ 0; 0 \ 1]$ ， $p_Y(C_1) = 0.6$ ， $p_Y(C_2) = 0.4$

平面等高线更方便探讨高斯分布形状和决策边界。图 6 (a) 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 两个曲面等高线；两个曲面交线便是决策边界 (图 6 (a) 中深蓝色线)。 $p_Y(C_1)$ 大于 $p_Y(C_2)$ ，因此 C_1 类数据的影响更大，决策边界便远离质心 μ_1 ；也就是说 C_1 “势力”更大。

观察图 6 (a) 等高线，发现 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 同心圆大小不同。再次，注意图 6 (a) 等高线为联合概率密度函数，而图 2 为似然概率。

由于每一类数据在每个特征上方差相同，且条件独立；因此 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 等高线为正圆。

图 6 (b) 比较后验概率 $f_{Y|X}(C_1 | \mathbf{x})$ 和 $f_{Y|X}(C_2 | \mathbf{x})$ 曲面。由于图 6 所示为二分类问题，因此只要 $f_{Y|X}(C_1 | \mathbf{x}) > 0.5$ ，则可判定 \mathbf{x} 分类为 C_1 。

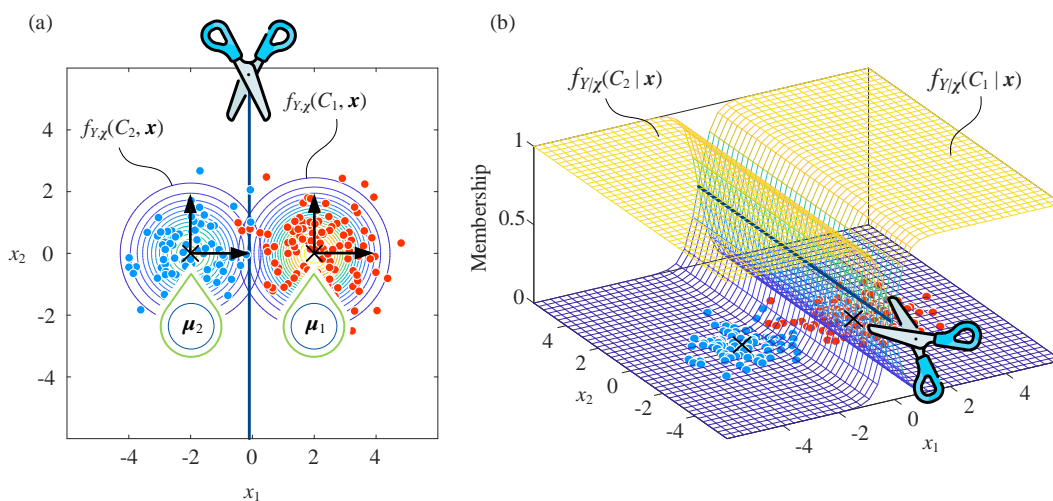


图 6. 第一类高斯判别分析， $\Sigma_1 = \Sigma_2 = [1 \ 0; 0 \ 1]$ ， $p_Y(C_1) = 0.6$ ， $p_Y(C_2) = 0.4$

10.5 第二类

第二类高斯判别分析， Σ_k 相等且为对角阵（协方差矩阵除主对角线外，其他元素为 0，即特征条件独立）；但是，主对角线元素不相等。

两特征、两分类

对于 $D=2$, $K=2$ 的情况， Σ_1 和 Σ_2 可以写成如下形式：

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad (20)$$

其中， $\sigma_1 \neq \sigma_2$ 。

由于 Σ_1 和 Σ_2 相等，代入 (11)，可以发现二次项消去；因此，确定第二类高斯判别分析的决策边界也是直线。

举个例子

下面，举个例子分析第二类高斯判别：

$$\mu_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.4, \quad p_Y(C_2) = 0.6, \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad (21)$$

图 7 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面，这两个曲面的交线为决策边界； $p_Y(C_2) > p_Y(C_1)$ ， $f_{Y|X}(C_2, \mathbf{x})$ 曲面高度高于 $f_{Y|X}(C_1, \mathbf{x})$ 。

观察图 8 (a)，发现 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面等高线为形状相似正椭圆。图 8 (b) 所示为 $f_{Y|X}(C_1 | \mathbf{x})$ 和 $f_{Y|X}(C_2 | \mathbf{x})$ 后验概率曲面，以及两个曲面交线，即决策边界。

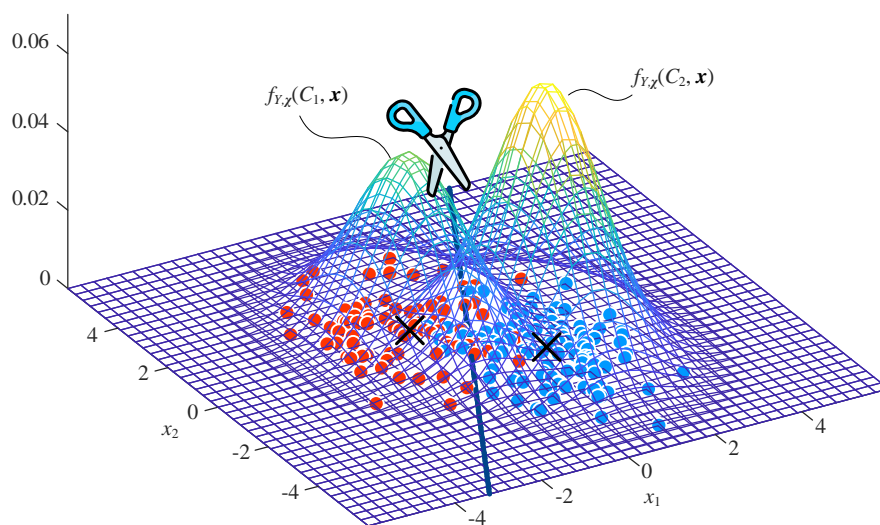
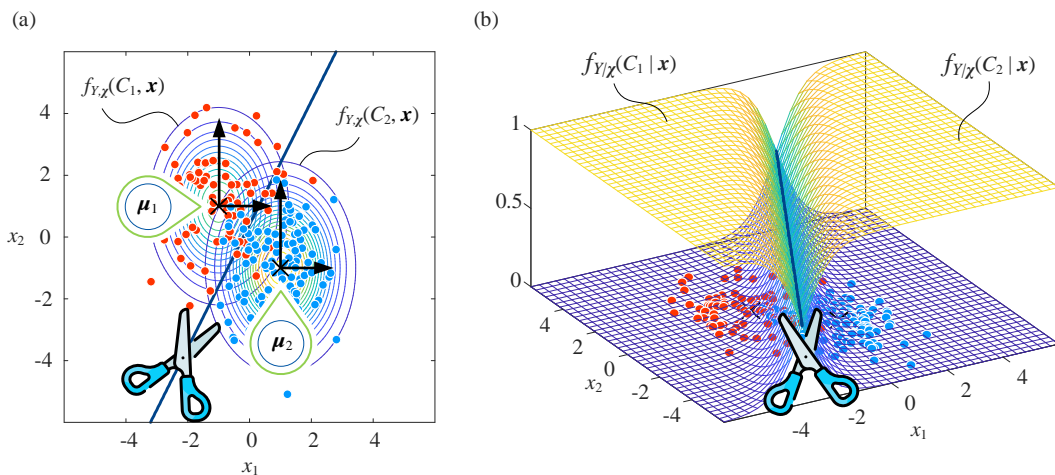


图 7. 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面；第二类高斯判别分析， $\Sigma_1 = \Sigma_2 = [1 \ 0; 0 \ 2]$ ， $p_Y(C_1) = 0.4$ ， $p_Y(C_2) = 0.6$

图 8. 第二类高斯判别分析, $\Sigma_1 = \Sigma_2 = [1 \ 0; 0 \ 2]$, $p_Y(C_1) = 0.4$, $p_Y(C_2) = 0.6$

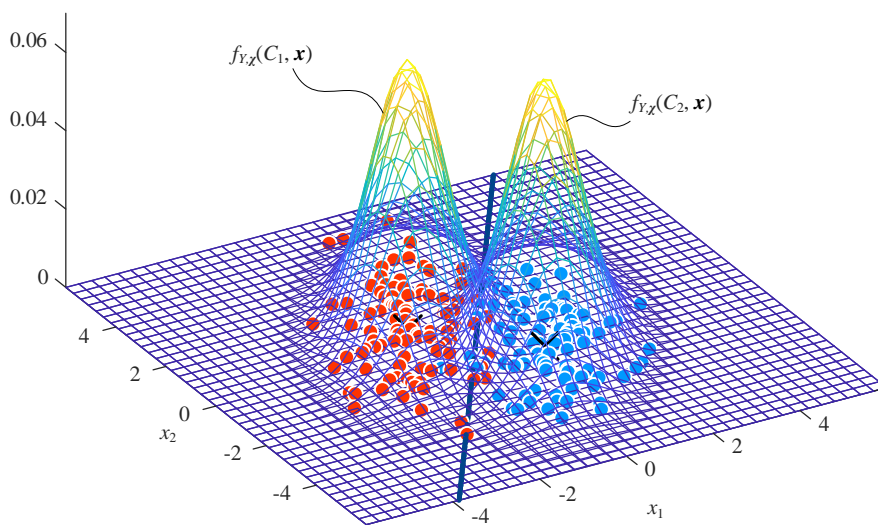
10.6 第三类

第三类高斯判别分析特点是, 仅假设类别协方差矩阵 Σ_k 完全一致; 方差和条件独立不做任何限制。

举个例子

$$\mu_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.5, \quad p_Y(C_2) = 0.5, \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 2 \end{bmatrix} \quad (22)$$

图 9 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 。根据 $\Sigma_1 = \Sigma_2$ 这个条件, 可以判定决策边界为直线。由于假设 $p_Y(C_1) = p_Y(C_2)$, $f_{Y|X}(C_1, \mathbf{x})$ 曲面和 $f_{Y|X}(C_2, \mathbf{x})$ 高度相等。

图 9. 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面; 第三类高斯判别分析, $\Sigma_1 = \Sigma_2 = [1 \ 0.8; 0.8 \ 2]$, $p_Y(C_1) = 0.5$, $p_Y(C_2) = 0.5$

观察图 10 (a) 可以发现, $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 等高线为旋转椭圆, 这是因为 Σ_1 和 Σ_2 两个矩阵协方差不为 0。图 10 (b) 所示为 $f_{Y|X}(C_1 | \mathbf{x})$ 和 $f_{Y|X}(C_2 | \mathbf{x})$ 两个后验概率曲面, 以及决策边界。

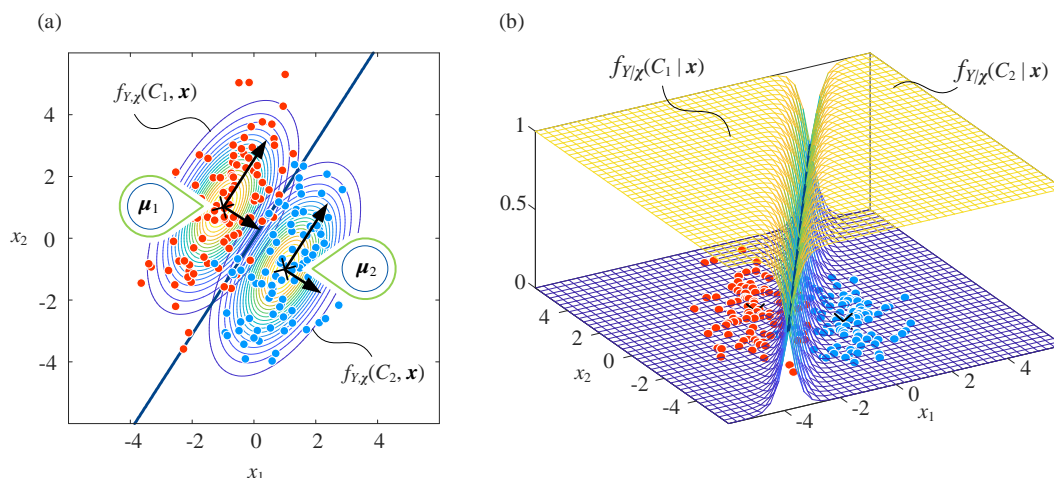


图 10. 第三类高斯判别分析, $\Sigma_1 = \Sigma_2 = [1 \ 0.8; 0.8 \ 2]$, $p_Y(C_1) = 0.5$, $p_Y(C_2) = 0.5$

10.7 第四类

第四类高斯判别分析, Σ_k 为对角阵 (除主对角线之外元素为 0), 即类别内特征“条件独立”; 并且, Σ_k 各自对角线元素相同。不同于第一类, 第四类不同类别 Σ_k 不同。

举个例子

给定如下条件:

$$\mu_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.3, \quad p_Y(C_2) = 0.7, \quad \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \quad (23)$$

图 11 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面。投影在 x_1x_2 平面上, 第四类高斯判别分析的决策边界为正圆, 如图 12 (a) 所示。图 12 (b) 比较两个后验曲面高度。

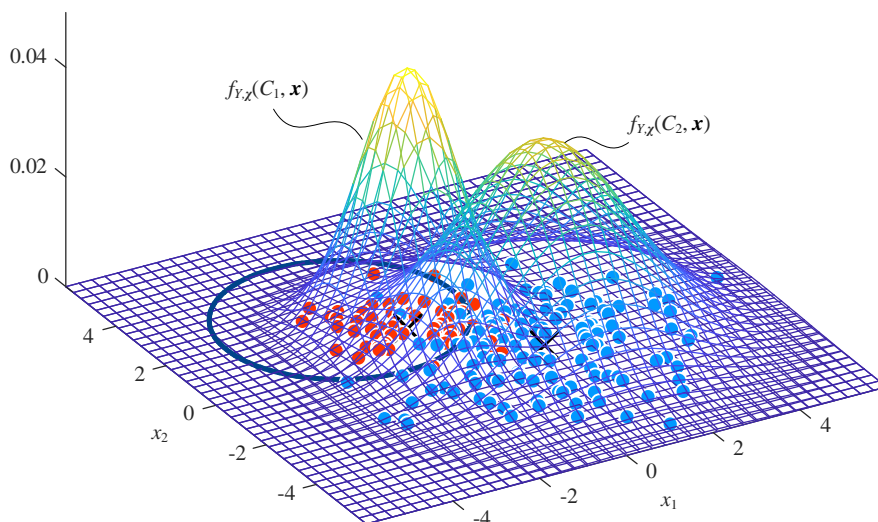


图 11. 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面；第四类高斯判别分析， $\Sigma_1 = [1 \ 0; 0 \ 1]$, $\Sigma_2 = [3 \ 0; 0 \ 3]$, $p_Y(C_1) = 0.3$, $p_Y(C_2) = 0.7$

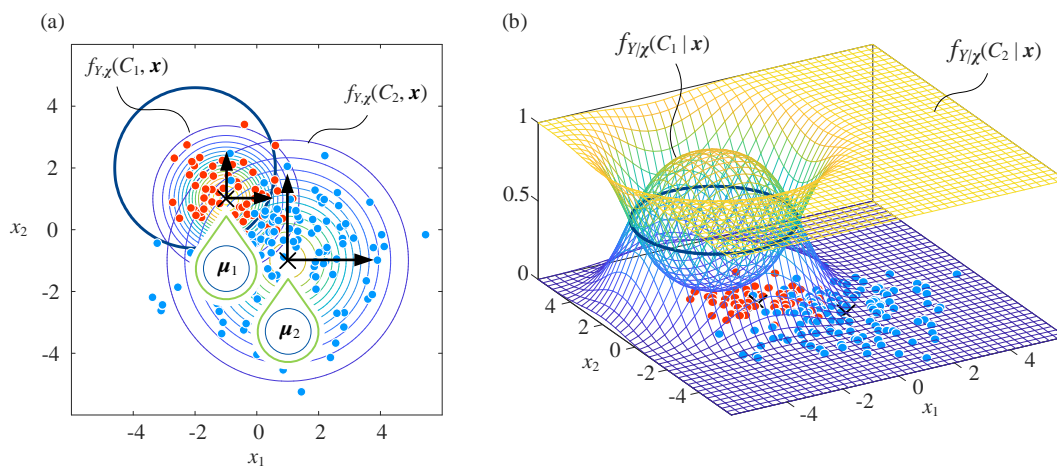


图 12. 第四类高斯判别分析， $\Sigma_1 = [1 \ 0; 0 \ 1]$, $\Sigma_2 = [3 \ 0; 0 \ 3]$, $p_Y(C_1) = 0.3$, $p_Y(C_2) = 0.7$

10.8 第五类

第五类高斯判别分析，不同类别 Σ_k 不同； Σ_k 为对角阵，类别内特征条件独立。但是，第五类高斯判别分析对 Σ_k 对角线元素大小不做限制。

再次请大家注意，第五类高斯判别分析对应高斯朴素贝叶斯分类。大家可以自己推导第五类高斯判别分析决策边界一般式。

举个例子

下面举个例子方便可视化：

$$\mu_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_Y(C_1) = 0.4, \quad p_Y(C_2) = 0.6, \quad \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \quad (24)$$

由于 Σ_1 和 Σ_2 均为对角阵，决策边界解析式中没有 $x_1 x_2$ 项。因此，决策边界为正圆锥曲线，如图 13 所示。图 14 (a) 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面等高线；图 14 (b) 比较后验曲面等高线。

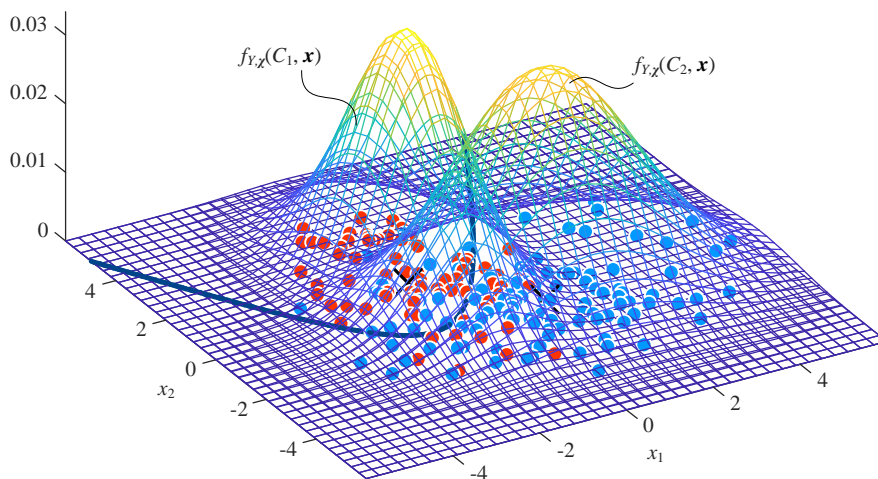


图 13. 比较 $f_{Y|X}(C_1, \mathbf{x})$ 和 $f_{Y|X}(C_2, \mathbf{x})$ 曲面；第五类高斯判别分析， $\Sigma_1 = [1 \ 0; 0 \ 3]$ ， $\Sigma_2 = [4 \ 0; 0 \ 2]$ ， $p_Y(C_1) = 0.4$ ， $p_Y(C_2) = 0.6$

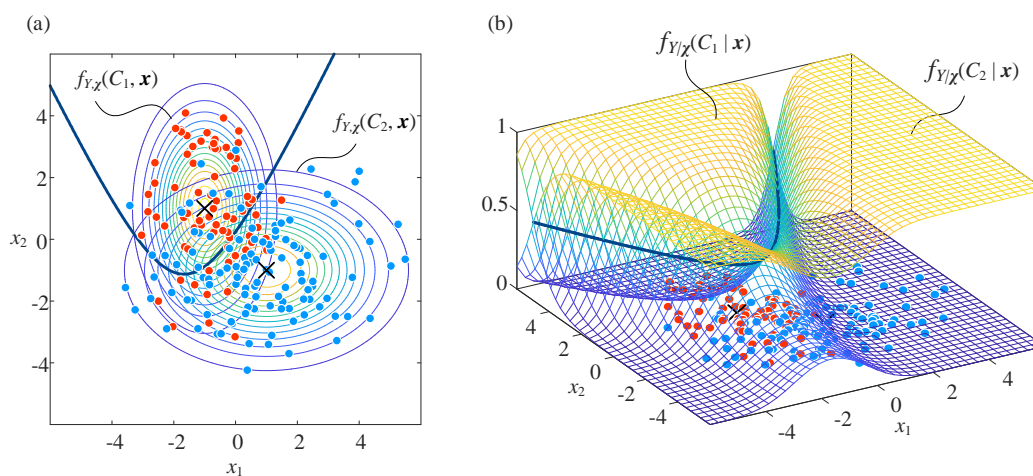


图 14. 第五类高斯判别分析， $\Sigma_1 = [1 \ 0; 0 \ 1]$ ， $\Sigma_2 = [3 \ 0; 0 \ 3]$ ， $p_Y(C_1) = 0.4$ ， $p_Y(C_2) = 0.6$

10.9 第六类

第六类高斯判别分析对 Σ_k 不做任何限制。后验概率 PDF 等高线为任意椭圆。第六类高斯判别分析的决策边界可以是单条直线、平行直线、椭圆、双曲线、蜕化双曲线、抛物线等等。下面看几个例子。

椭圆

如下参数条件得到的决策边界为椭圆：

$$\mu_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \begin{cases} p_Y(C_1) = 0.2 \\ p_Y(C_2) = 0.8 \end{cases}, \quad \Sigma_1 = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 3 & 1.8 \\ 1.8 & 3 \end{bmatrix} \quad (25)$$

如图 15 所示，(25) 对应决策边界为椭圆。

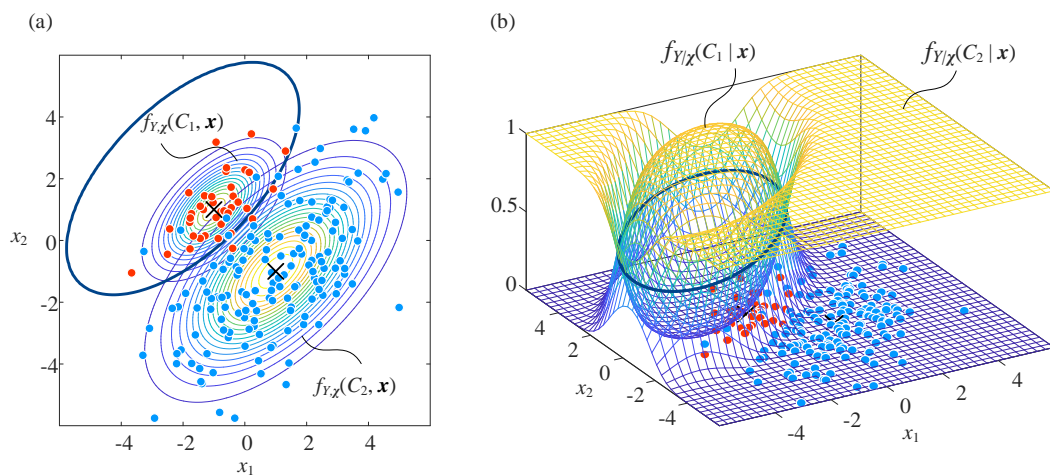


图 15. 决策边界为椭圆

双曲线

下列给出的参数条件得到双曲线决策边界，如图 16 所示：

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{cases} p_Y(C_1) = 0.4 \\ p_Y(C_2) = 0.6 \end{cases}, \quad \Sigma_1 = \begin{bmatrix} 1 & -0.6 \\ -0.6 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} \quad (26)$$

有了以上铺垫，本章最后介绍线性判别分析和二次判别分析。

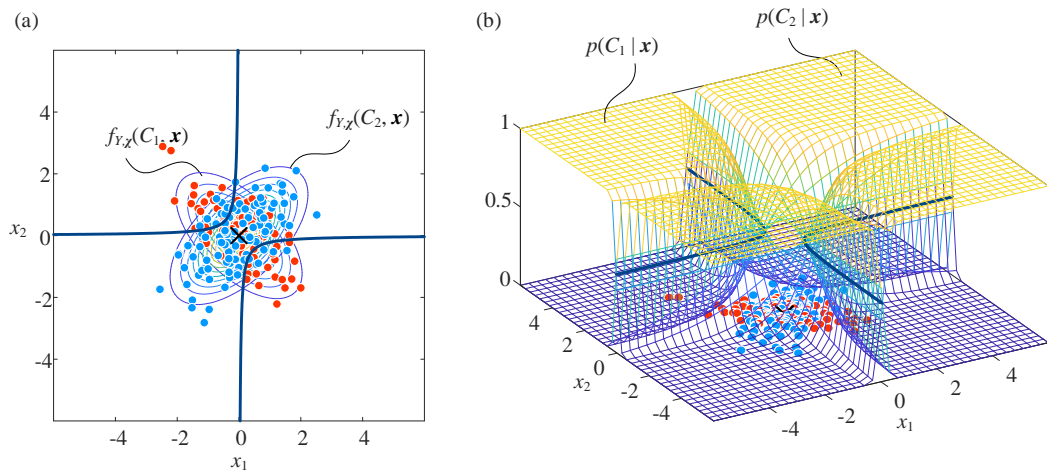


图 16. 决策边界为双曲线

10.10 线性和二次判别分析

线性判别分析 (Linear Discriminant Analysis, LDA) 是一种监督学习算法，用于分类和降维问题。它基于假设，即每个类别的数据都满足高斯分布，并且不同类别之间的协方差矩阵相等。具体来说，LDA

通过寻找一个投影方向，可以将数据从高维空间投影到低维空间，并最大程度地保留不同类别之间的差异，同时最小化同一类别内部的方差。

这个投影方向可以被认为是线，称为“判别线”，可以用于分类或降维。在训练过程中，算法学习这个判别线，并使用它来计算给定输入数据点属于哪个类别的后验概率。LDA 是 GDA 的一种特殊形式。

投影

本节先以“降维”这种思路讨论线性判别分析。

如图 17 所示，采用高斯分布描述 $f_{Y,X}(C_1, \mathbf{x})$ 和 $f_{Y,X}(C_2, \mathbf{x})$ 这两个联合概率。四个参数刻画 $f_{Y,X}(C_1, \mathbf{x})$ 和 $f_{Y,X}(C_2, \mathbf{x})$ —— (1) C_1 质心位置 μ_1 ; (2) C_2 质心位置 μ_2 ; (3) C_1 形状 Σ_1 ; (4) C_2 形状 Σ_2 。

决策边界的解析式为：

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (27)$$

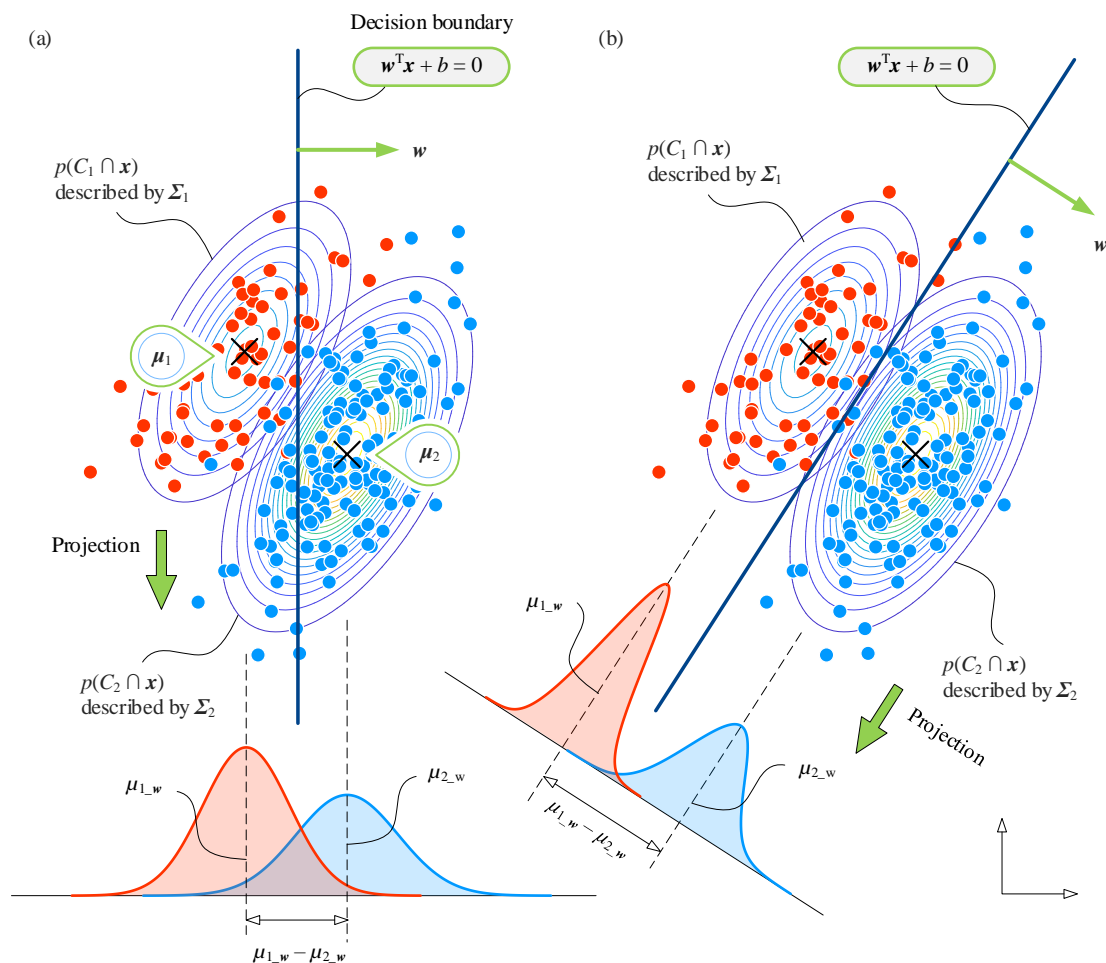
其中， \mathbf{w} 为直线梯度向量，且为单位矩阵。

C_1 质心位置 μ_1 和 C_2 质心位置 μ_2 沿着决策边界方向投影，也就是向 \mathbf{w} 投影，可以得到：

$$\mu_{1_w} = \mathbf{w}^T \mu_1, \quad \mu_{2_w} = \mathbf{w}^T \mu_2 \quad (28)$$

Σ_1 和 Σ_2 向 \mathbf{w} 向量方向投影，得到：

$$\sigma_{1_w}^2 = \mathbf{w}^T \Sigma_1 \mathbf{w}, \quad \sigma_{2_w}^2 = \mathbf{w}^T \Sigma_2 \mathbf{w} \quad (29)$$

图 17. 从投影角度解释线性判别分析，两个子图中 w 代表不同方向

比较图 17 两个子图，可以发现图 17 (b) 的分类效果更好。线性判别分析相当于样本数据投影后，最大化类间差异 $(\mu_{1,w} - \mu_{2,w})$ ，且最小化类内差异 $(\sigma_{1,w}^2 + \sigma_{2,w}^2)$ 。

目标函数

从投影角度，线性判别分析的目标函数为：

$$\arg \max_w \frac{(\mu_{1,w} - \mu_{2,w})^2}{\sigma_{1,w}^2 + \sigma_{2,w}^2} \quad (30)$$

将 (28) 和 (29) 代入 (30)，得到：

$$\arg \max_w \frac{w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w}{w^T (\Sigma_1 + \Sigma_2) w} \quad (31)$$

(31) 分子描述的是类间距离，分母描述的是类内聚集程度。看到这个式子，大家是否眼前一亮？



《矩阵力量》从瑞利商、特征值分解、拉格朗日乘子法几个不同角度讲解过上式，建议大家回顾。

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

分类器函数

`sklearn.discriminant_analysis.LinearDiscriminantAnalysis` 为线性判别分析函数。这个函数采用本章第三类 GDA，即仅仅假设各个类别协方差矩阵 Σ_k 完全一致；方差和条件独立不做任何限制。

`sklearn.discriminant_analysis.QuadraticDiscriminantAnalysis` 为二次判别分析算法函数。这个函数采用本章介绍的第六类 GDA，对 Σ_k 不做任何限制。

图 18 和图 19 分别展示采用线性判别分析 LDA 和二次判别分析 QDA 分类鸢尾花结果。简单来说，二次判别分析是一种监督学习算法，用于分类问题。它基于假设，即每个类别的数据都满足二次高斯分布，即每个类别的协方差矩阵都不相等。具体来说，QDA 通过估计每个类别的均值和协方差矩阵来建模该类别的二次高斯分布。

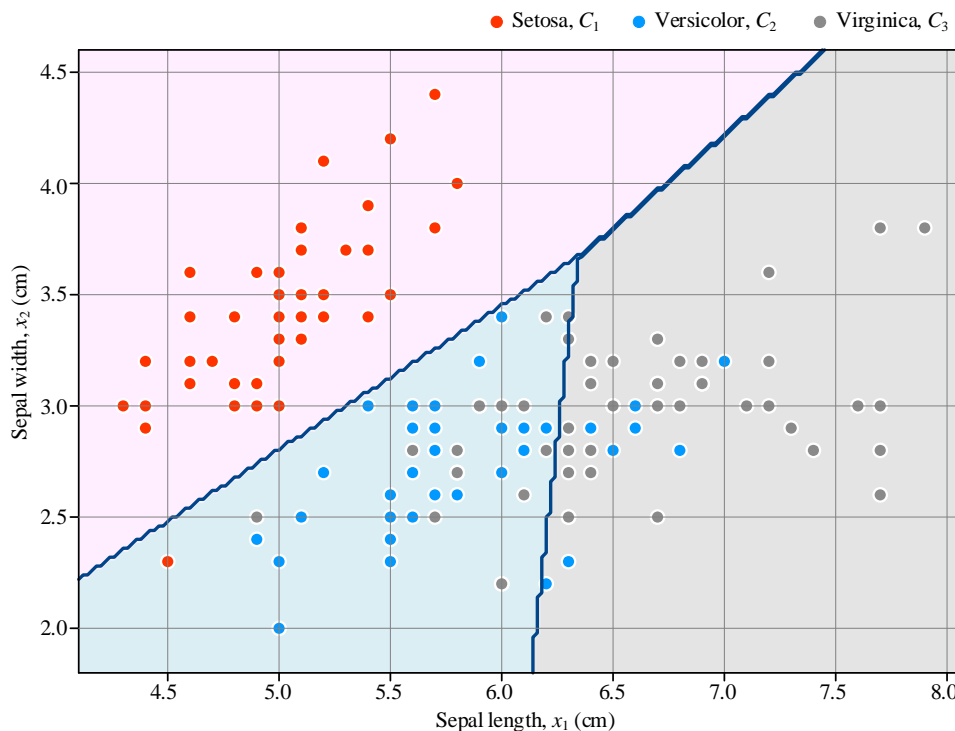


图 18. 线性判别分析分类鸢尾花

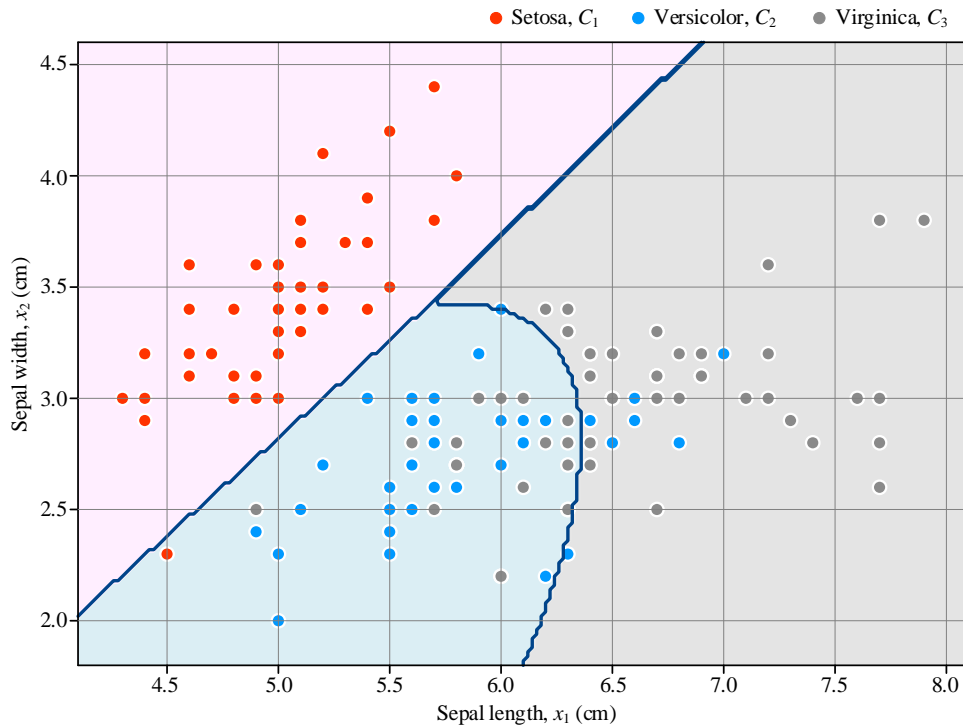


图 19. 二次判别分析分类鸢尾花

代码 Bk7_Ch10_02.ipynb 利用判别分析分类器分类鸢尾花数据，并绘制图 18 和图 19。代码 1 是完成线性判别分析的部分核心代码，请大家阅读注释并自行学习。

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn import datasets

# 导入鸢尾花数据
iris = datasets.load_iris()
X = iris.data[:, 0:2]
y = iris.target

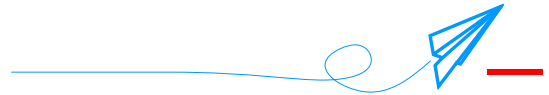
# 线性判别分析模型
a lda = LinearDiscriminantAnalysis()

# 拟合数据
b lda.fit(X, y)

# 查询点
c q = np.c_[xx1.ravel(), xx2.ravel()]

# 预测
d y_predict_LDA = lda.predict(q)
y_predict_LDA = y_predict_LDA.reshape(xx1.shape)
```

代码 1. 线性判别分析 | Bk7_Ch10_02.ipynb

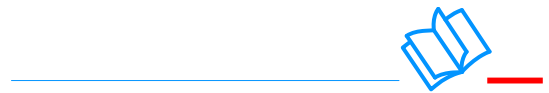


高斯判别分析 GDA，一次判别分析 LDA、二次判别分析 QDA 是常见的监督学习算法，用于分类和判别问题。它们都基于不同的假设和数学模型。LDA、QDA 相当于 GDA 的特殊形式。

GDA 假设每个类别的数据都服从高斯分布，然后估计每个类别的均值和协方差矩阵来建模高斯分布，最后使用这些参数计算后验概率进行分类。

LDA 假设每个类别的数据都满足高斯分布，并且不同类别之间的协方差矩阵相等。然后，它寻找一个投影方向，可以将数据从高维空间投影到低维空间，并最大程度地保留不同类别之间的差异，同时最小化同一类别内部的方差。

QDA 假设每个类别的数据都满足二次高斯分布，即每个类别的协方差矩阵都不相等。然后，它估计每个类别的均值和协方差矩阵，最后使用这些参数计算后验概率进行分类。



建议大家研究下面这个官方示例。示例代码生成随机数，进行 LDA 和 QDA 分析，并绘制旋转椭圆来表达数据的协方差矩阵形状。

https://scikit-learn.org/stable/auto_examples/classification/plot_lda_qda.html