模式识别作业 5

贝叶斯分类

姓名：罗雁天

院系：清华大学电子系

学号：2018310742

日期：April 28, 2019

目录

第 1 章 问题描述

设有符合正太分布的两类样本，并且设 $P(\omega_1) = P(\omega_2) = 0.5$.

$$
\begin{aligned}
\omega_1 &= \{(3,4),(3,8),(2,6),(4,6)\} \\
\omega_2 &= \{(3,0),(3,-4),(1,-2),(5,-2)\}
\end{aligned}
\tag{1.1}
$$

求解以下问题：

- 求识别函数
- 求识别界面方程
- 绘制识别界面

第 2 章 分类结果

首先计算两个类的类均值：

$$M_1 = [3, 6], M_2 = [3, -2] \tag{2.1}$$

根据无偏估计的协方差矩阵计算方法：

$$\Sigma_i = \frac{1}{N-1}(\omega_i - M_i)^T(\omega_i - M_i) \quad i = 1, 2 \tag{2.2}$$

计算两个类的协方差矩阵：

$$\Sigma_1 = \begin{bmatrix} 0.6667 & 0 \\ 0 & 2.6667 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2.6667 & 0 \\ 0 & 2.6667 \end{bmatrix} \tag{2.3}$$

计算正态分布时贝叶斯判别准则所需要的参数如下：

$$\begin{aligned}
\mathbf{W}_1 &= -\frac{1}{2}\Sigma_1^{-1} = \begin{bmatrix} -0.7500 & 0 \\ 0 & -0.1875 \end{bmatrix} \\
\mathbf{W}_2 &= -\frac{1}{2}\Sigma_2^{-1} = \begin{bmatrix} -0.1875 & 0 \\ 0 & -0.1875 \end{bmatrix} \\
\mathbf{w}_1 &= \Sigma_1^{-1}M_1 = [4.50, 2.25]^T; \\
\mathbf{w}_2 &= \Sigma_2^{-1}M_2 = [1.125, -0.75]^T; \\
w_{10} &= -\frac{1}{2}M_1^T\Sigma_1^{-1}M_1 - \frac{1}{2}\ln|\Sigma_1| + \ln P(\omega_1) = -14.4808 \\
w_{20} &= -\frac{1}{2}M_2^T\Sigma_1^{-1}M_2 - \frac{1}{2}\ln|\Sigma_2| + \ln P(\omega_2) = -4.1115
\end{aligned} \tag{2.4}$$

对任意数据点 $\mathbf{x} = [x_1, x_2]^T$，计算两类的识别函数如下：

$$\begin{aligned}
d_1(\mathbf{x}) &= \mathbf{x}^T\mathbf{W}_1\mathbf{x} + \mathbf{w}_1^T\mathbf{x} + w_{10} \\
&= -0.75x^2 - 0.1875x_2^2 + 4.5x_1 + 2.25x_2 - 14.4808 \\
d_2(\mathbf{x}) &= \mathbf{x}^T\mathbf{W}_2\mathbf{x} + \mathbf{w}_2^T\mathbf{x} + w_{20} \\
&= -0.1875x^2 - 0.1875x_2^2 + 1.125x_1 - 0.75x_2 - 4.1115
\end{aligned} \tag{2.5}$$

则判别函数如下：

$$f(\mathbf{x}) = \begin{cases} x \in \text{class1} & \text{if} \quad d_1(\mathbf{x}) > d_2(\mathbf{x}) \\ x \in \text{class2} & \quad else \end{cases} \tag{2.6}$$

计算识别界面如下：

$$d_1(\mathbf{x}) = d_2(\mathbf{x}) \Rightarrow -0.5625x_1^2 + 3.375x_1 + 3x_2 - 10.3693 = 0 \tag{2.7}$$

由此可以看出，分类界面在此种情况下是抛物线。
绘制出两个二维高斯分布的曲面如图2.1所示，识别界面如图2.2所示

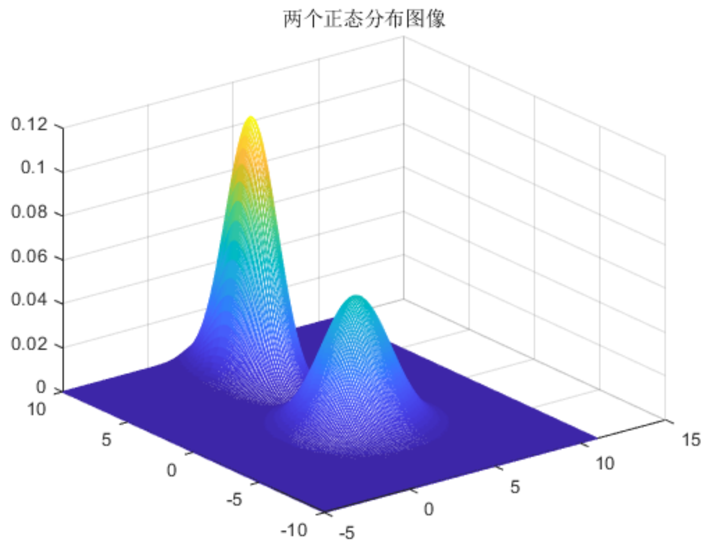

图 2.1: 二维高斯分布密度函数曲面

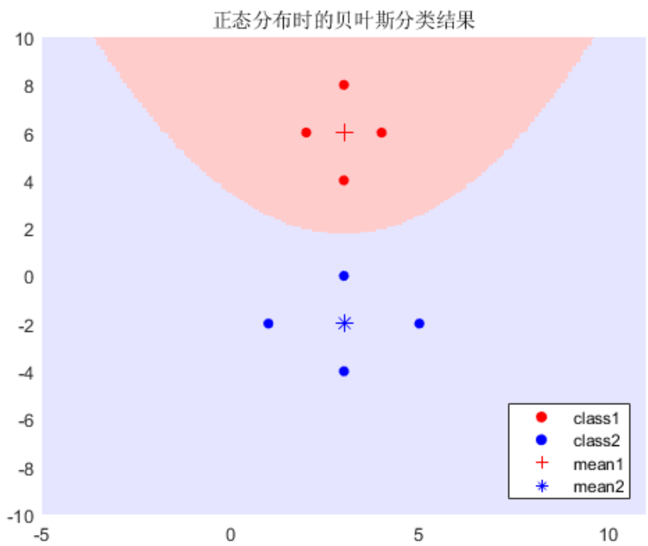

图 2.2: 识别界面

第 3 章 总结与反思

根据课上的结论我们可以知道，对于二维数据的二分类问题，根据正态分布时贝叶斯判别准则我们可以知道，分类界面主要有三种情况：

- 当 $\Sigma_i = \sigma^2 I, i = 1,2$ 时，分类界面是两类均值连线的中垂线；
- 当 $\Sigma_1 = \Sigma_2$ 时，分类界面是过两类均值连线中点的直线 (一般不垂直)；
- 当 $\Sigma_1 \neq \Sigma_2$ 时，分类界面是二次曲线 (圆、椭圆、双曲线、抛物线等)。

以下，我们通过修改数据分别模拟以上几种情况的分类界面。

3.1 $\Sigma_1 = \Sigma_2 = \sigma^2 I$

修改数据为：

$$\begin{aligned} \omega_1 &= \{(3,4),(3,8),(1,6),(5,6)\} \\ \omega_2 &= \{(1,0),(1,-4),(-1,-2),(3,-2)\} \end{aligned} \tag{3.1}$$

绘制出两个二维高斯分布的曲面如图3.1所示，识别界面如图3.2所示，从中可以看出，分类界面是两类均值连线的中垂线；

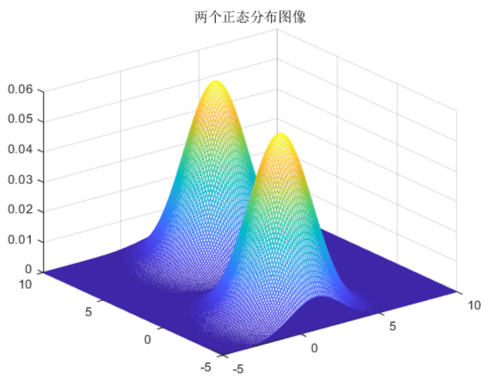

图 3.1: 二维高斯分布密度函数曲面

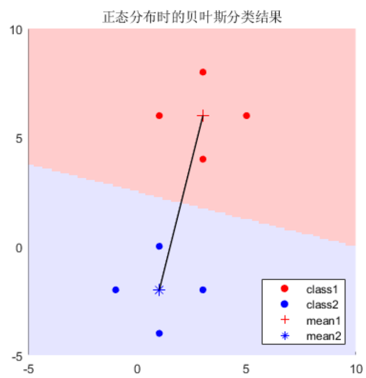

图 3.2: 识别界面

3.2 $\Sigma_1 = \Sigma_2 \neq \sigma^2 I$

修改数据为：

$$\begin{aligned} \omega_1 &= \{(3,4),(3,8),(2,6),(4,6)\} \\ \omega_2 &= \{(1,0),(1,-4),(0,-2),(2,-2)\} \end{aligned} \tag{3.2}$$

绘制出两个二维高斯分布的曲面如图3.3所示，识别界面如图3.4所示，从中可以看出，分类界面是过两类均值连线中点的直线 (一般不垂直)；

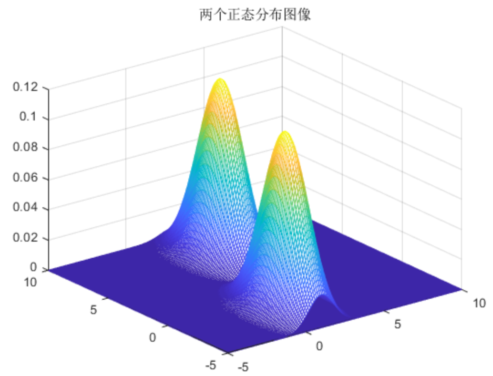

图 3.3: 二维高斯分布密度函数曲面

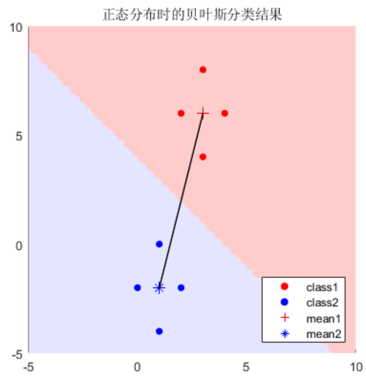

图 3.4: 识别界面

3.3 $\Sigma_1 \neq \Sigma_2$

原问题即是分类界面为抛物线的情况，在此不再考虑。

3.3.1 分类界面为圆

修改数据为：

$$\begin{aligned} \omega_1 &= \{(3,5),(3,7),(2,6),(4,6)\} \\ \omega_2 &= \{(3,0),(3,-4),(1,-2),(5,-2)\} \end{aligned} \tag{3.3}$$

绘制出两个二维高斯分布的曲面如图3.5所示，识别界面如图3.6所示，从中可以看出，分类界面是圆；

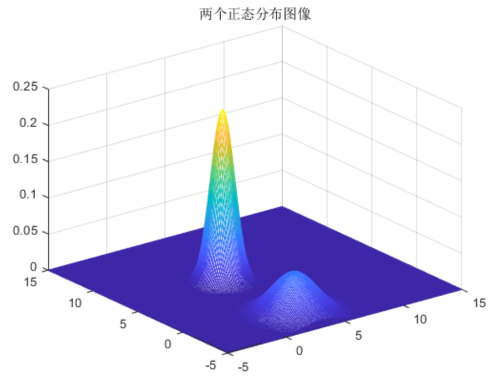

图 3.5: 二维高斯分布密度函数曲面

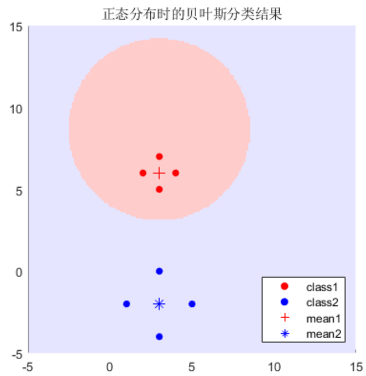

图 3.6: 识别界面

3.3.2 分类界面为椭圆

修改数据为：

$$
\begin{aligned}
\omega_1 &= \{(3,4),(3,8),(2,6),(4,6)\} \\
\omega_2 &= \{(3,1),(3,-5),(1,-2),(5,-2)\}
\end{aligned} \tag{3.4}
$$

绘制出两个二维高斯分布的曲面如图3.7所示，识别界面如图3.8所示，从中可以看出，分类界面是椭圆；

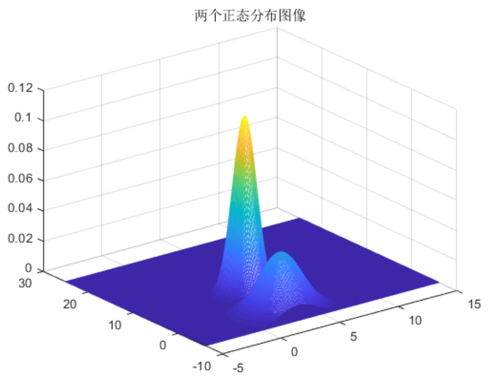

图 3.7: 二维高斯分布密度函数曲面

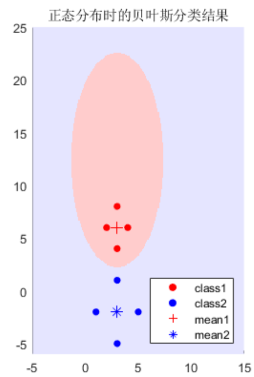

图 3.8: 识别界面

3.3.3 分类界面为双曲线

修改数据为：

$$
\begin{aligned}
\omega_1 &= \{(3,4),(3,8),(2,6),(4,6)\} \\
\omega_2 &= \{(3,-1),(3,-3),(1,-2),(5,-2)\}
\end{aligned} \tag{3.5}
$$

绘制出两个二维高斯分布的曲面如图3.9所示，识别界面如图3.10所示，从中可以看出，分类界面是双曲线；

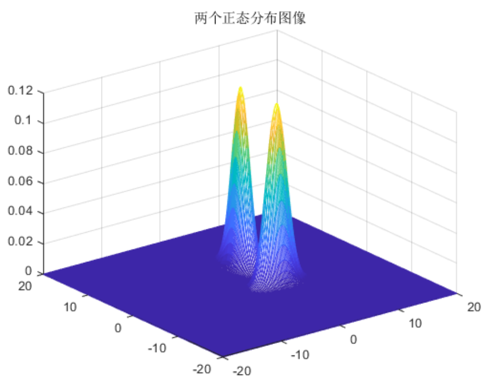

图 3.9: 二维高斯分布密度函数曲面

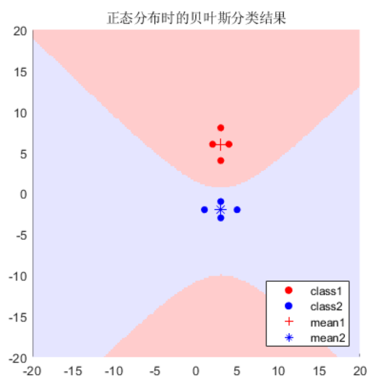

图 3.10: 识别界面

3.3.4 分类界面为退化的双曲线 (双曲线的渐近线)

修改数据为：

$$\begin{aligned} \omega_1 &= \{(3,4),(3,8),(2,6),(4,6)\} \\ \omega_2 &= \{(4,3),(8,3),(6,2),(6,4)\} \end{aligned} \tag{3.6}$$

绘制出两个二维高斯分布的曲面如图3.11所示，识别界面如图3.12所示，从中可以看出，分类界面是两条直线；

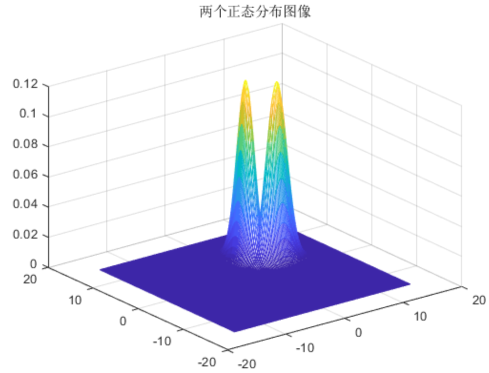

图 3.11: 二维高斯分布密度函数曲面

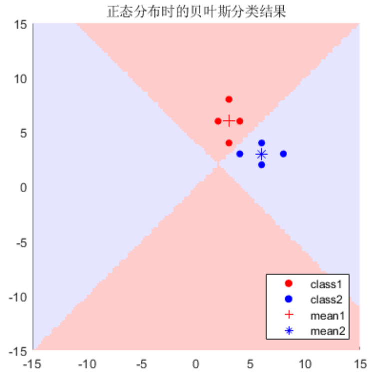

图 3.12: 识别界面

第 4 章 代码说明

本次实验使用 Matlab 语言编写，所有代码放置在“code/”文件夹下:

- main.m: 以上讨论的各种情况执行的主函数，运行即可得到所有的分类界面图像以及二维正态密度函数的图像；
- Bayes_Gauss.m: 使用高斯分布时的贝叶斯判别准则绘制分类界面的函数。