

作业 7 - 识别正态分布样本

郭一隆

May 10, 2018

1 问题描述

设有符合正态分布的两类训练样本：

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

并且设

$$P(\omega_1) = P(\omega_2) = 0.5$$

求

- 识别函数
- 识别界面方程
- 作图

2 问题求解

2.1 观察训练样本

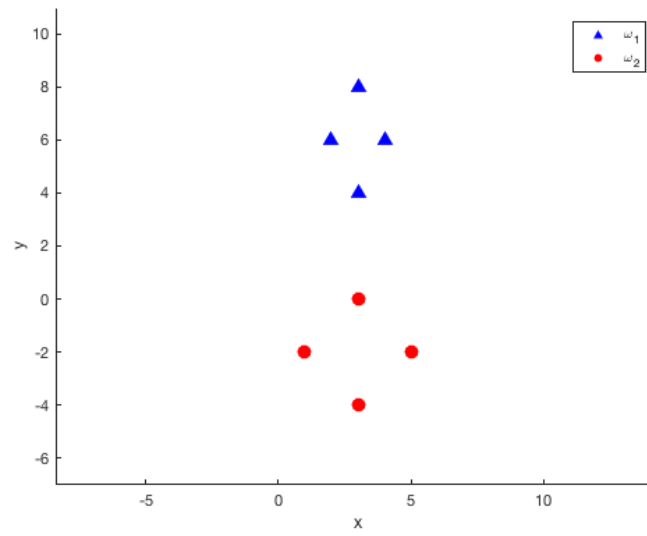


Figure 1: 两类训练样本的分布

假设两类互相独立地各自服从正态分布，即

$$\Sigma_1 \neq \Sigma_2$$

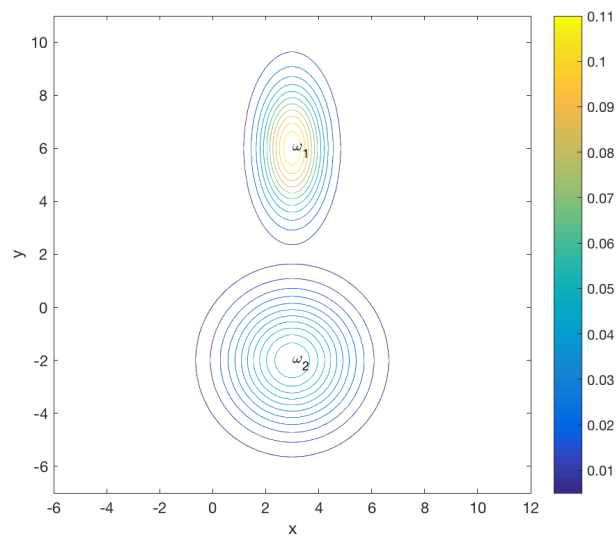


Figure 2: 两类正态分布数据（等概率线）

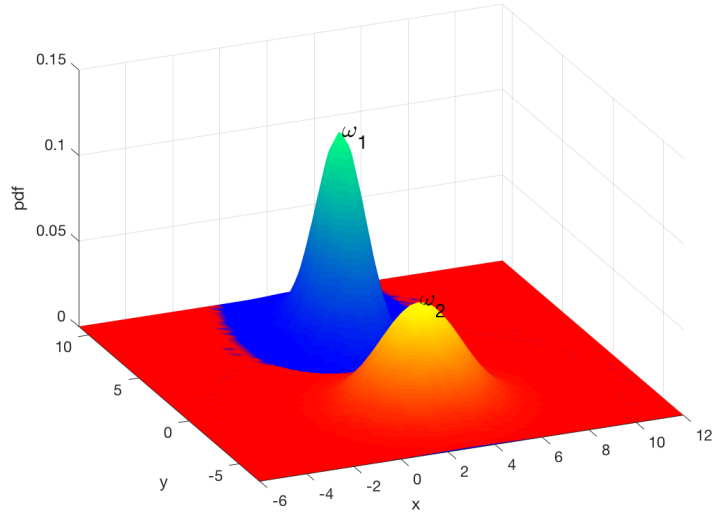


Figure 3: 概率分布函数

2.2 识别函数

对于类别 i ,

$$p(\mathbf{X}|\omega_i) = \frac{1}{2\pi\sqrt{|\Sigma_i|}} \exp \left[-\frac{1}{2}(\mathbf{X} - \mathbf{M}_i)^T \Sigma_i^{-1} (\mathbf{X} - \mathbf{M}_i) \right]$$

识别函数

$$d_i(\mathbf{X}) = \ln [p(\mathbf{X}|\omega_i)p(\omega_i)] = \mathbf{X}^T \mathbf{W}_i \mathbf{X} + \mathbf{w}_i^T \mathbf{X} + w_{0,i}$$

其中

$$\mathbf{W}_i = -\frac{1}{2}\Sigma_i^{-1}$$

$$\mathbf{w}_i = \Sigma_i^{-1}\mathbf{M}_i$$

$$w_{0,i} = -\frac{1}{2}\mathbf{M}_i^T \Sigma_i^{-1} \mathbf{M}_i - \frac{1}{2} \ln |\Sigma_i| + \ln p(\omega_i)$$

于是可在二维平面内绘制两类识别函数的等高线，如图4所示。

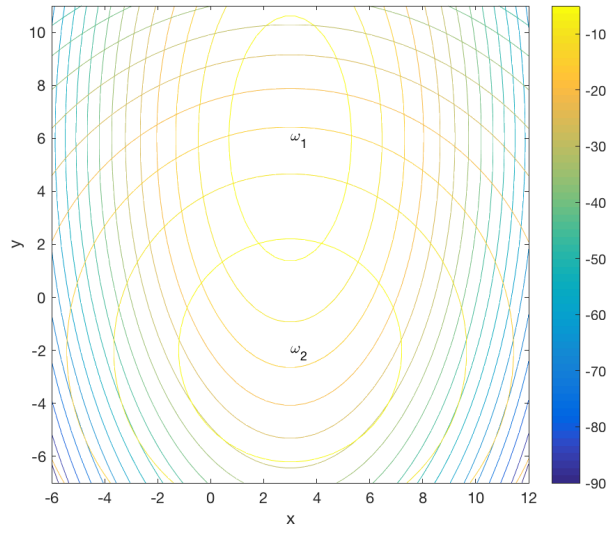


Figure 4: 识别函数等高线

两类识别函数的符号形式为：

$$d_1(\mathbf{X}) = -\frac{3}{4}(x-3)^2 - \frac{3}{16}(y-6)^2 - 0.98$$

$$d_2(\mathbf{X}) = -\frac{3}{16}(x-3)^2 - \frac{3}{16}(y+2)^2 - 1.67$$

2.3 识别界面

观察图4可知，两类识别函数等值等高线相交处构成了识别界面，即令

$$d_1(\mathbf{X}) = d_2(\mathbf{X})$$

联立可得识别界面方程

$$y = \frac{3}{16}(x-3)^2 + 1.77$$

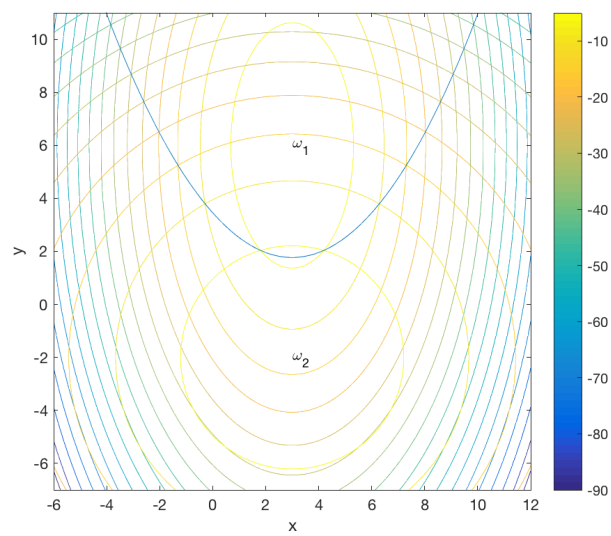


Figure 5: 联立识别函数求得识别界面

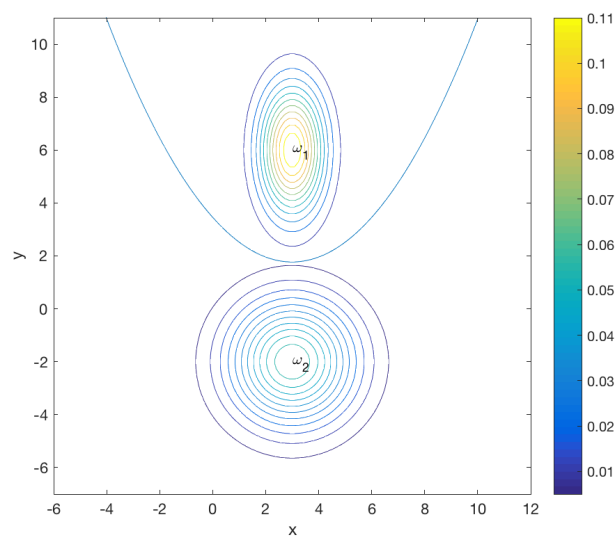


Figure 6: 两类正态分布数据（等概率线 + 识别界面）

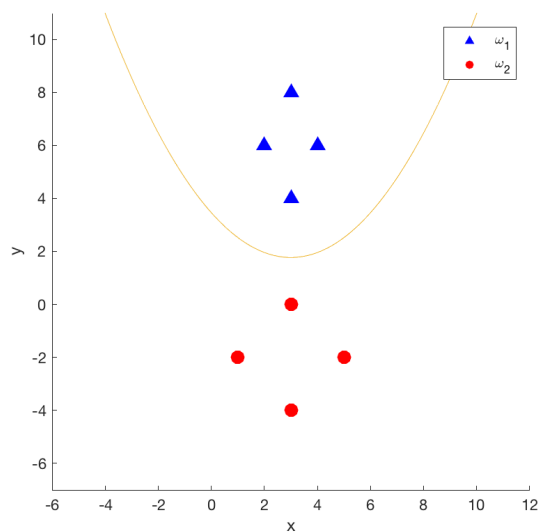


Figure 7: 识别界面 + 两类训练样本散点分布

3 总结

- 本次实验对样本做了较为简单的假设，即假设不同类别、不同维度的特征均相互独立，且各自服从正态分布。实际情况一般没有如此理想，特征与特征之间往往具有相关性，甚至不同类别也不完全独立。

附

- 源码仓库链接：

<https://github.com/Nuullll/pattern-recognition-assignment-7>

- 运行环境：

MATLAB 版本：9.1.0.441655 (R2016b)

MATLAB 许可证编号：1114839

操作系统：Mac OS X Version: 10.13.4 Build: 17E199

Java 版本：Java 1.7.0_75-b13 with Oracle Corporation Java

HotSpot(TM) 64-Bit Server VM mixed mode

- 运行方法：

直接运行 `run.m` 即可。