模式识别 第一次作业

2022 年 9 月 13 日

说明

- 作业用中文撰写，务必注明题号，鼓励使用 LaTeX；
- 编程题需要提交源码，并指出运行环境以及环境依赖以方便查看。编程语言建议使用 Python, 源码中提供简单注释；
- 文档按"学号 _ 姓名.pdf"命名，".pdf"和代码文件全部打包成"学号 _ 姓名.zip"提交；
- 编程题要在".pdf"中简要说清思路和方法；
- 本次作业截止时间为 2022 年 9 月 28 日，请到课程网站及时提交。

题目 1. 设 $\omega_{\max}$ 为类别状态，此时对所有的 $i(i=1,\ldots,c)$，有 $P\left(\omega_{\max}|\boldsymbol{x}\right) \geq P\left(\omega_i|\boldsymbol{x}\right)$。

(1) 证明 $P\left(\omega_{\max}|\boldsymbol{x}\right) \geq 1/c$。

(2) 证明对于最小错误率判定规则，平均错误概率为

$$P\left(error\right) = 1 - \int P\left(\omega_{\max}|\boldsymbol{x}\right) p\left(\boldsymbol{x}\right) d\boldsymbol{x}$$

(3) 利用这两个结论证明 $P\left(error\right) \leq \left(c-1\right)/c$。

(4) 描述一种情况，在此情况下有 $P\left(error\right)=\left(c-1\right)/c$。

题目 2. 对于一个 c 类分类问题，假设各类先验概率为 $P(\omega_i)$, $i=1,\ldots,c$; 条件概率密度为 $P(\boldsymbol{x}|\omega_i)$, $i=1,\ldots,c$, ($\boldsymbol{x}$ 表示特征向量); 将第 j 类样本判别为第 i 类的损失为 λ_{ij}。

(1) 请写出贝叶斯风险最小决策和最小错误率决策的决策规则；

(2) 引入拒识 (表示为第 $c+1$ 类), 假设决策损失为

$$\lambda(\alpha_i|\omega_j)=\begin{cases}0, & i=j \quad i,j=1,\ldots,c\\ \lambda_r, & i=c+1\\ \lambda_s, & otherwise\end{cases}$$

请写出最小风险决策的决策规则 (包括分类规则和拒识规则)。

题目 3. 考虑三维正态分布 $p\left(\boldsymbol{x}|\omega\right)\sim N\left(\boldsymbol{\mu},\Sigma\right)$，其中

$$\boldsymbol{\mu}=\begin{pmatrix}1\\2\\2\end{pmatrix},\Sigma=\begin{pmatrix}1&0&0\\0&5&2\\0&2&5\end{pmatrix}$$

(1) 构造白化变换 $\mathrm{A}_\omega=\Phi\Lambda^{-1/2}$，计算分别表示本征向量和本征值的矩阵 Φ 和 Λ；接下来，将此分布转换为以原点为中心、协方差矩阵为单位阵的分布，即 $p\left(\boldsymbol{x}|\omega\right)\sim N\left(\boldsymbol{0},\mathrm{I}\right)$。

(2) 将 (1) 中的白化变换应用于点 $\boldsymbol{x}_0=(0.5,0,1)^t$，求其经过白化变换后的点 $\boldsymbol{x}_\omega$。

(3) 通过详细计算，证明原分布中从 $\boldsymbol{x}_0$ 到均值 $\boldsymbol{\mu}$ 的 Mahalanobis 距离与变换后的分布中从 $\boldsymbol{x}_\omega$ 到 $\boldsymbol{0}$ 的 Mahalanobis 距离相等。

(4) 概率密度在一个一般的线性变换下是否保持不变？换句话说，对于某线性变换 T，是否有 $p\left(\boldsymbol{x}_0|N\left(\boldsymbol{\mu},\Sigma\right)\right)=p\left(T^t\boldsymbol{x}_0|N\left(T^t\boldsymbol{\mu},T^t\Sigma T\right)\right)$？解释原因。

题目 4. 对一个 c 类分类问题, 特征向量 $\boldsymbol{x} \in \mathcal{R}^d$，假设各类先验概率相等，每一类条件概率密度为高斯分布。

(1) 请写出类条件概率密度函数的数学形式；

(2) 请写出在下面两种情况下的最小错误率决策判别函数: (a) 类协方差矩阵不等; (b) 所有类协方差矩阵相等.

(3) 在基于高斯概率密度的二次判别函数中，当协方差矩阵为奇异时，判别函数变得不可计算。请说出两种克服协方差奇异的方法。

题目 5. 编程题

(1) 写一个程序产生 d 维空间的样本点，服从均值为 $\boldsymbol{\mu}$ 和协方差矩阵 Σ 的正态分布。

(2) 考虑正态分布

$$p(\boldsymbol{x}|\omega_1) \sim N\left(\left(\begin{array}{c}1\\0\end{array}\right), \mathbf{I}\right), p(\boldsymbol{x}|\omega_2) \sim N\left(\left(\begin{array}{c}-1\\0\end{array}\right), \mathbf{I}\right)$$

其中，$\mathbf{I}$ 为单位矩阵，且 $P(\omega_1) = P(\omega_2) = 0.5$。说明贝叶斯判决边界。

(3) 产生 $n = 100$ 个点 (50 个 ω_1 类的点，50 个 ω_2 类的点)，并计算经验误差。

(4) 对于不断增加的 n 值重复以上步骤，$100 \leq n \leq 1000$，步长为 100，并绘出所得的经验误差。