

模式识别与机器学习大作业

PRML

王滑 尹超 伍昱衡 陈志强 崔祯徐 郑子辰
(按照姓氏笔画排序)

中国科学院大学，北京 10004

University of Chinese Academy of Sciences, Beijing 100049, China

2025.6.3 - 2025.6.30

序言

本文为笔者模式识别与机器学习的大作业。
望老师批评指正。

目录

序言	I
目录	II
1 分类与估计问题的解答	1
1.1 问题 1: 贝叶斯分类与风险决策	1
1.1.1 问题描述	1
1.1.2 解答	1
1.2 问题 2: 高斯分布下的分类	1
1.2.1 问题描述	1
1.2.2 解答	2
1.3 问题 3: 隐马尔可夫模型示例	2
1.3.1 问题描述	2
1.3.2 解答	2
1.4 问题 4: 参数与非参数估计	2
1.4.1 问题描述	2
1.4.2 解答	2
1.5 问题 5: USPS 数据集分类	3
1.5.1 问题描述	3
1.5.2 解答	3
方法	3
Python 代码实现	4
代码说明	6
结果分析与讨论	6
总结	6
附录 A. 中英文对照表	7
A.1 中英文对照表	7

Chapter 1 分类与估计问题的解答

1.1 问题 1: 贝叶斯分类与风险决策

1.1.1 问题描述

对于 c 类分类问题, 假设各类先验概率为 $P(\omega_i), i = 1, \dots, c$, 类条件概率密度为 $p(\mathbf{x} | \omega_i), i = 1, \dots, c$ (\mathbf{x} 为样本特征向量), 将第 j 类样本误判为第 i 类的风险为 λ_{ij} 。

- (1) 写出贝叶斯最小错误率决策和最小风险决策的决策规则。
- (2) 引入拒识选项 (第 $c+1$ 类)。假设正确分类风险 $\lambda(\alpha_i | \omega_j) = 0 (i = j)$; 错误分类风险 $\lambda(\alpha_i | \omega_j) = \lambda_s (i \neq j)$; 拒识风险 $\lambda(\alpha_i | \omega_j) = \lambda_r (i = c+1)$ 。写出带拒识的最小风险决策规则。
- (3) 说明拒识的意义。

1.1.2 解答

(1) 决策规则

- 贝叶斯最小错误率决策: 对样本 \mathbf{x} , 选择最大化后验概率的类别:

$$\hat{\omega} = \arg \max_{\omega_i} P(\omega_i | \mathbf{x}) = \arg \max_{\omega_i} p(\mathbf{x} | \omega_i) P(\omega_i)$$

- 最小风险决策: 选择最小化期望风险的类别:

$$\hat{\omega} = \arg \min_{\omega_i} \sum_{j=1}^c \lambda_{ij} P(\omega_j | \mathbf{x})$$

- (2) 带拒识的最小风险决策拒识为 α_{c+1} 。分类风险:

$$R(\alpha_i | \mathbf{x}) = \lambda_s (1 - P(\omega_i | \mathbf{x})), \quad i = 1, \dots, c$$

拒识风险:

$$R(\alpha_{c+1} | \mathbf{x}) = \lambda_r$$

决策规则:

$$\hat{\alpha} = \arg \min_{\alpha_k} R(\alpha_k | \mathbf{x}), \quad k = 1, \dots, c+1$$

若 $P(\omega_i | \mathbf{x}) > 1 - \frac{\lambda_r}{\lambda_s}$, 分类为 ω_i ; 否则拒识。

- (3) 拒识的意义拒识允许分类器在不确定性高时避免决策, 降低误分类风险。例如, 在医疗诊断中, 拒识可避免错误治疗。

1.2 问题 2: 高斯分布下的分类

1.2.1 问题描述

对于 c 类分类问题, 特征向量 $\mathbf{x} \in \mathbb{R}^d$, 假设各类先验概率相等, 类条件概率密度为 d 维高斯分布。

- (1) 写出类条件概率密度函数的数学形式。
- (2) 写出协方差矩阵相等和不等情况下的最小错误率判别函数。
- (3) 两类等协方差情况下, 决策面特点? 何时通过两类均值中点?
- (4) 当协方差矩阵奇异时, 如何克服?

1.2.2 解答

(1) 类条件概率密度

$$p(\mathbf{x} | \omega_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right)$$

(2) 判别函数

– 协方差相等 ($\Sigma_i = \Sigma$):

$$g_i(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_{i0}, \quad \mathbf{w} = \Sigma^{-1} \mu_i, \quad w_{i0} = -\frac{1}{2} \mu_i^T \Sigma^{-1} \mu_i$$

– 协方差不等:

$$g_i(\mathbf{x}) = -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)$$

(3) 决策面特点为超平面; 当先验概率相等时, 通过均值中点 $\frac{\mu_1 + \mu_2}{2}$ 。

(4) 克服奇异协方差在协方差矩阵对角线上加小正数 (岭正则化)。

1.3 问题 3: 隐马尔可夫模型示例

1.3.1 问题描述

举出生活中符合隐马尔可夫模型 (HMM) 的例子, 说明隐状态和观测变量含义, 及观测序列评价、状态序列解码的实际问题。

1.3.2 解答

示例: 语音识别

- 隐状态: 音素序列 (如 /k/, /æ/, /t/ 表示“cat”)。
- 观测变量: 音频信号的 MFCC 特征。
- 观测序列评价: 计算音频对应特定单词的概率。
- 状态序列解码: 确定音频的最可能音素序列, 识别单词。

1.4 问题 4: 参数与非参数估计

1.4.1 问题描述

- (1) 总结最大似然估计、贝叶斯估计、k 近邻估计和 Parzen 估计的原理和步骤。
- (2) 一维特征空间样本点 $\{-1, 2, 3, 6\}$, 服从均匀分布, 用最大似然法估计概率密度函数。
- (3) 样本服从高斯分布, 协方差 Σ 已知, 均值 μ 未知, 先验为 $\mathcal{N}(m_0, \Sigma_0)$, 写出贝叶斯估计过程。
- (4) 说明 k 近邻密度估计与分类规则的关系。

1.4.2 解答

(1) 估计方法

- 最大似然估计 (MLE): 最大化似然函数 $L(\theta) = p(\mathbf{x} | \theta)$ 。步骤: 定义似然, 对数化, 求导, 解参数。
- 贝叶斯估计: 使用后验 $p(\theta | \mathbf{x}) \propto p(\mathbf{x} | \theta)p(\theta)$, 取均值或众数。步骤: 指定先验, 计算后验, 估计参数。

- **k 近邻估计**: 基于样本局部密度估计。步骤: 选择 k , 计算距离, 估计密度为 $\frac{k}{nV}$ 。
 - **Parzen 窗估计**: 使用核函数估计密度。步骤: 选择核和带宽, 累加样本贡献。
- (2) 均匀分布的 MLE 样本: $\{-1, 2, 3, 6\}$, 假设 $x \sim U(a, b)$ 。似然: $L(a, b) = \frac{1}{(b-a)^4}$, 约束 $a \leq -1, b \geq 6$ 。
最大化得: $\hat{a} = -1, \hat{b} = 6$ 。密度: $p(x) = \frac{1}{7}, -1 \leq x \leq 6$ 。
- (3) 贝叶斯估计后验:

$$p(\mu | \mathbf{x}) = \mathcal{N} \left((\Sigma_0^{-1} + n\Sigma^{-1})^{-1} (\Sigma_0^{-1}m_0 + n\Sigma^{-1}\bar{\mathbf{x}}), (\Sigma_0^{-1} + n\Sigma^{-1})^{-1} \right)$$

估计: $\hat{\mu} = (\Sigma_0^{-1} + n\Sigma^{-1})^{-1} (\Sigma_0^{-1}m_0 + n\Sigma^{-1}\bar{\mathbf{x}})$ 。

- (4) **k 近邻关系** 密度估计通过局部样本统计密度, 分类通过多数投票; 两者均依赖局部信息, 密度估计可间接支持分类。

1.5 问题 5: USPS 数据集分类

1.5.1 问题描述

在 USPS 数据集中, 选择两个类别 (例如数字 0 和 1), 实现贝叶斯分类器:

- (1) 假设两类协方差矩阵相等, 使用线性判别函数分类, 并与 Fisher 判别准则的线性分类器比较。
- (2) 分别估计两类协方差矩阵, 使用二次判别函数分类。
- (3) 分析两种情况的结果。

附录: USPS 数据集 手写数字数据集, 包含 7291 个训练样本和 2007 个测试样本, 每样本为 16×16 灰度图像 (256 维)。数据以 HDF5 格式提供, 包含 `train` 和 `test` 组, 各有 `data` (特征) 和 `target` (标签)。

读取代码:

```
1 import h5py
2 with h5py.File(path, 'r') as hf:
3     train = hf.get('train')
4     X_tr = train.get('data')[:]
5     y_tr = train.get('target')[:]
6     test = hf.get('test')
7     X_te = test.get('data')[:]
8     y_te = test.get('target')[:]
```

1.5.2 解答

方法

- **数据预处理**:
 - 使用 `h5py` 加载数据集。
 - 选择数字 0 和 1, 过滤训练和测试数据。
 - 特征维度: 256 维向量。
- **第一部分: 线性判别分析 (LDA) 与 Fisher 判别**:
 - **假设**: 两类共享协方差矩阵。
 - **LDA**: 线性判别函数:

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

其中：

$$\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_0), \quad w_0 = -\frac{1}{2}(\mu_0 + \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_0)$$

– **Fisher 判别**：投影向量：

$$\mathbf{w} = S_W^{-1}(\mu_1 - \mu_0)$$

在此情况下与 LDA 等价。

– **注意**：特征维度 (256) 可能导致协方差矩阵奇异，需正则化。

• **第二部分：二次判别分析 (QDA)：**

– **假设**：两类协方差矩阵不同。

– **判别函数**：

$$g_i(\mathbf{x}) = -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i)$$

– **实现**：分别计算协方差矩阵，加入正则化。

• **评估与分析**：

– 在测试集上计算准确率。

– 讨论高维和小样本的影响。

Python 代码实现

```

1  import h5py
2  import numpy as np
3  from numpy.linalg import pinv, det
4
5  # 加载 USPS 数据集
6  with h5py.File('usps.h5', 'r') as hf:
7      train = hf.get('train')
8      X_tr = train.get('data')[:] # 训练数据
9      y_tr = train.get('target')[:] # 训练标签
10     test = hf.get('test')
11     X_te = test.get('data')[:] # 测试数据
12     y_te = test.get('target')[:] # 测试标签
13
14 # 过滤类别 0 和 1
15 mask_tr = (y_tr == 0) | (y_tr == 1)
16 X_tr = X_tr[mask_tr]
17 y_tr = y_tr[mask_tr]
18 mask_te = (y_te == 0) | (y_te == 1)
19 X_te = X_te[mask_te]
20 y_te = y_te[mask_te]
21
22 # 第一部分：LDA (协方差矩阵相等)
23 # 计算每类均值
24 mu0 = np.mean(X_tr[y_tr == 0], axis=0) # 类别 0 均值
25 mu1 = np.mean(X_tr[y_tr == 1], axis=0) # 类别 1 均值
26
27 # 计算共享协方差矩阵并正则化
28 epsilon = 1e-5
29 d = X_tr.shape[1]
```

```
30 Sigma = np.cov(X_tr.T) + epsilon * np.eye(d)
31
32 # 计算线性判别参数
33 w = pinv(Sigma) @ (mu1 - mu0)
34 w0 = -0.5 * (mu0 + mu1).T @ pinv(Sigma) @ (mu1 - mu0)
35
36 # 测试集预测
37 scores = X_te @ w + w0
38 y_pred_lda = (scores > 0).astype(int)
39 accuracy_lda = np.mean(y_pred_lda == y_te)
40 print(f'LDA 准确率: {accuracy_lda:.4f}')
41
42 # Fisher 判别: 与 LDA 等价, 省略
43
44 # 第二部分: QDA (分别估计协方差)
45 # 计算每类协方差矩阵并正则化
46 Sigma0 = np.cov(X_tr[y_tr == 0].T) + epsilon * np.eye(d)
47 Sigma1 = np.cov(X_tr[y_tr == 1].T) + epsilon * np.eye(d)
48
49 # 计算逆矩阵和行列式
50 inv_Sigma0 = pinv(Sigma0)
51 inv_Sigma1 = pinv(Sigma1)
52 log_det_Sigma0 = np.log(det(Sigma0))
53 log_det_Sigma1 = np.log(det(Sigma1))
54
55 # 定义二次判别函数
56 def g(x, mu, inv_Sigma, log_det):
57     diff = x - mu
58     return -0.5 * log_det - 0.5 * diff.T @ inv_Sigma @ diff
59
60 # 测试集预测
61 y_pred_qda = []
62 for x in X_te:
63     g0 = ['$x = np.linspace(0, 10, 100)']
64     y = np.sin(x)
65
66     plt.figure(figsize=(8, 6))
67     plt.plot(x, y, 'b-', label='Sine wave')
68     plt.title('Simple Sine Wave')
69     plt.xlabel('X axis')
70     plt.ylabel('Y axis')
71     plt.grid(True)
72     plt.legend()
73
74     plt.savefig('sine_wave.png') g(x, mu0, inv_Sigma0, log_det_Sigma0)
75     g1 = g(x, mu1, inv_Sigma1, log_det_Sigma1)
76     y_pred_qda.append(1 if g1 > g0 else 0)
77
78 accuracy_qda = np.mean(np.array(y_pred_qda) == y_te)
```



```
79 | print(f'QDA 准确率: {accuracy_qda:.4f}')
```

代码说明

- **数据加载与过滤**: 使用 `h5py` 读取数据, 过滤类别 0 和 1。
- **LDA 实现**: 计算均值和共享协方差, 加入正则化, 执行线性分类。
- **QDA 实现**: 计算每类协方差, 计算二次判别函数。
- **Fisher 判别**: 与 LDA 等价, 未单独实现。

结果分析与讨论

- **数据特性**:
 - 训练集约 1400-1500 个样本, 每类约 700 个。
 - 特征维度 256, 图像特征相关性高, 需正则化。
- **LDA 与 Fisher 判别**:
 - LDA 使用线性边界, 适合 0 和 1 的明显差异。
 - Fisher 判别在此等价。
- **QDA 性能**:
 - 优势: 捕捉协方差差异。
 - 挑战: 高维小样本导致过拟合。
- **预期结果**:
 - 准确率可能 >90%。
 - LDA 因参数少更稳定, QDA 可能因过拟合略逊。
- **改进建议**:
 - 使用 PCA 降维。
 - 优化正则化参数 ϵ 。
 - 考虑数据增强。

总结

LDA 和 QDA 分类器成功分类 USPS 数据集的数字 0 和 1。LDA 因简单性更鲁棒, QDA 因高维限制可能过拟合。两者对 0 和 1 均表现良好, 具体准确率需运行验证。

附录 A. 中英文对照表

A.1 中英文对照表

表 A.1: 中英文对照表

英文	中文
Bayes classification	贝叶斯分类
decision rule	决策规则
minimum error rate	最小错误率
minimum risk	最小风险
rejection option	拒识选项
Gaussian distribution	高斯分布
covariance matrix	协方差矩阵
discriminant function	判别函数
decision boundary	决策边界
Hidden Markov Model	隐马尔可夫模型
hidden state	隐状态
observation sequence	观测序列
maximum likelihood estimation	最大似然估计
Bayesian estimation	贝叶斯估计
k-Nearest Neighbor	k 近邻
Parzen window	Parzen 窗
linear discriminant	线性判别
quadratic discriminant	二次判别
prior probability	先验概率
posterior probability	后验概率