

模式识别 (Pattern Recognition)

期末复习笔记

整理人: Your Name

2026 年 1 月 8 日

目录

1 绪论 (3 学时向世明)	4
1.1 模式识别基础	4
1.2 模式识别系统流程	4
1.3 模式识别系统设计	4
1.4 模式识别方法分类	5
1.5 本课程内容体系	5
2 贝叶斯决策理论 (3 学时向世明)	6
2.1 核心公式	6
2.2 最小错误率贝叶斯决策	6
2.3 最小风险贝叶斯决策	6
2.4 分类器设计	7
2.5 高斯密度下的判别函数	7
2.5.1 情形一: $\Sigma_i = \sigma^2 I$ (各特征独立且方差相等)	7
2.5.2 情形二: $\Sigma_i = \Sigma$ (协方差矩阵相等)	8
2.5.3 情形三: $\Sigma_i \neq \Sigma_j$ (协方差矩阵任意)	8
2.6 错误率分析	8
2.7 离散变量的贝叶斯决策	9
2.8 复合模式分类	9
3 概率密度函数估计 (6 学时张燕明)	10
3.1 基本概念	10
3.2 最大似然估计	10
3.3 贝叶斯估计	10
3.4 正态分布下的贝叶斯估计	10
3.5 贝叶斯学习	10
3.6 特征维数问题	10
3.7 期望最大法 (EM)	10

3.8 隐马尔可夫模型	10
4 非参数法 (3 学时张燕明)	11
4.1 密度估计	11
4.2 Parzen 窗方法	11
4.3 K 近邻估计	11
4.4 最近邻分类器	11
4.5 K 近邻分类器的改进	11
4.6 距离度量	11
5 线性分类器设计 (3 学时张燕明)	12
5.1 引言	12
5.2 线性判别函数与决策面	12
5.3 广义线性判别函数	12
5.4 感知准则函数	12
5.5 松驰方法	12
5.6 线性最小二乘方法	12
5.7 多类线性判别函数	12
6 神经网络和深度学习 (9 学时张燕明)	13
6.1 人工神经网络发展历程	13
6.2 人工神经网络基础	13
6.3 单层前馈神经网络	13
6.4 多层感知器与误差反向传播算法	13
6.5 BP 算法讨论	13
6.6 反馈神经网络	13
6.7 径向基函数网络	13
6.8 自组织映射	13
6.9 深度学习简介	13
6.10 波尔兹曼机简介	13
6.11 自编码器	13
6.12 卷积神经网络	13
6.13 Recurrent Neural Network	13
7 特征提取与选择 (6 学时张燕明)	14
7.1 维数灾难	14
7.2 特征提取	14
7.3 线性方法	14
7.4 非线性方法	14
7.5 特征选择	14

8 模型选择 (3 学时张燕明)	15
8.1 模型选择原则	15
8.2 模型评价标准	15
8.3 基于偏差和方差的分类器设计准则	15
8.4 分类器集成	15
8.5 Adaboost 学习方法及其应用	15
9 聚类分析 (6 学时张燕明)	16
9.1 K 均值聚类	16
9.2 模糊 K 均值聚类	16
9.3 层次聚类	16
9.4 谱聚类	16
9.5 在线聚类	16
9.6 集成聚类	16
10 支持向量机与核方法 (6 学时向世明)	17
10.1 结构风险、经验风险与 VC 维	17
10.2 线性可分支持向量机与硬间隔最大化	17
10.3 线性支持向量机与软间隔最大化	17
10.4 非线性支持向量机	17
10.5 支持向量机的优化与扩展	17
11 决策树方法 (3 学时向世明)	18
11.1 决策树模型、学习与特征选择	18
11.2 决策树生成与剪枝 (ID3, C4.5, CART 等算法)	18
11.3 随机森林算法	18
12 模式识别前沿趋势 (3 学时向世明)	19
12.1 课程内容总体回顾	19
12.2 模式识别发展现状与新动态	19
12.3 模式识别中的挑战性问题	19
13 考核 (6 学时向世明)	20
13.1 考试 + 答疑	20
附录：常用 LaTeX 模板	21

1 緒論 (3 學時向世明)

1.1 模式识别基础

- 基本定义:
 - 模式 (Pattern): 对客体 (Object) 或事件 (Event) 的描述, 通常是向量形式。
 - 模式类 (Pattern Class): 具有共同属性的模式集合。
 - 模式识别: 利用机器 (计算机) 自动将特定模式映射到特定类别的过程。
- 经典实例:
 - 硬币分类: 利用尺寸、重量等特征区分不同面值的硬币。
 - 鱼的分类 (Duda 教材经典例子): 区分 “三文鱼 (Salmon)” 和 “鲈鱼 (Sea Bass)”。单一特征 (如长度) 往往重叠严重, 需要多特征 (如长度 + 光泽度) 组合以寻找最佳决策边界。

1.2 模式识别系统流程

一个典型的模式识别系统包含以下五个核心环节:

1. 数据获取 (Data Acquisition): 通过传感器将物理变量转换为数字信号 (如摄像头采集图像、麦克风采集声音)。
2. 预处理 (Preprocessing): 去噪、分割、平滑、归一化, 旨在提高信噪比, 简化后续处理。
3. 特征提取与选择 (Feature Extraction & Selection):
 - 提取: 从原始数据中计算新的特征 (变换)。
 - 选择: 从一组特征中挑选出最具区分力的子集 (降维), 避免 “维数灾难”。
4. 分类决策 (Classification/Decision): 利用训练好的模型 (分类器) 对输入特征进行类别判定。
5. 后处理 (Post-processing): 利用上下文信息 (Context)、先验知识或代价函数对分类结果进行修正 (例如在 OCR 中利用语言模型纠错)。

1.3 模式识别系统设计

设计周期通常是一个不断迭代的反馈过程:

- 数据收集: 划分训练集 (Training Set)、验证集 (Validation Set) 和测试集 (Test Set)。
- 特征选择: 寻找具有不变性 (Invariance) 的特征 (如对平移、旋转、缩放不敏感)。
- 模型选择: 决定使用统计模型、神经网络模型还是句法模型; 决定模型的复杂度。
- 训练 (Training): 利用样本进行学习, 估计模型参数。

- **评价 (Evaluation)**: 核心指标是泛化能力 (**Generalization**), 即在未见过的测试数据上的表现, 需警惕过拟合 (**Overfitting**)。

1.4 模式识别方法分类

基于数据形式 :

- **统计模式识别 (Statistical PR)**: 基于特征向量, 利用概率论和统计学理论 (本课程重点)。
- **句法/结构模式识别 (Syntactic/Structural PR)**: 将模式分解为基元 (Primitives), 利用形式语言和语法规则描述结构关系 (适合汉字、染色体等结构性强的对象)。

基于学习方式 :

- **监督学习 (Supervised Learning)**: 训练数据带有类别标签 (Label)。
- **无监督学习 (Unsupervised Learning)**: 训练数据无标签, 旨在发现数据内部结构 (如聚类)。
- **强化学习 (Reinforcement Learning)**: 通过与环境交互, 基于奖励/惩罚机制进行学习。

1.5 本课程内容体系

- **基础**: 贝叶斯决策理论 (统计模式识别的基石)。
- **估计**: 概率密度函数估计 (参数法与非参数法)。
- **分类器**: 线性分类器、非线性分类器 (神经网络、SVM、决策树)。
- **工程**: 特征提取与选择、聚类分析。

2 贝叶斯决策理论 (3 学时向世明)

2.1 核心公式

首先，我们需要充分理解先验、似然、后验的意义。在贝叶斯决策理论中，最重要的公式是贝叶斯公式：

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)} \quad (1)$$

其中， $P(\omega_i)$ 是类别 ω_i 的先验概率， $p(x|\omega_i)$ 是在类别 ω_i 下观测到样本 x 的似然概率密度， $P(\omega_i|x)$ 是在观测到样本 x 后类别 ω_i 的后验概率。

一个典型案例是 HIV 检测，假设某种检测方法对 HIV 阳性患者的检测准确率为 99.9%，对 HIV 阴性患者的检测准确率为 99.5%。如果一个人被检测为阳性，我们需要计算他实际患有 HIV 的概率。假设总体中 HIV 阳性率为 0.1%，则可以利用贝叶斯公式计算后验概率。概率计算为：

$$P(\text{HIV}^+|\text{Test}^+) = \frac{P(\text{Test}^+|\text{HIV}^+)P(\text{HIV}^+)}{P(\text{Test}^+)} \quad (2)$$

2.2 最小错误率贝叶斯决策

等价于最大后验概率决策：

$$\text{Decide } \omega_i \text{ if } P(\omega_i|x) > P(\omega_j|x), \forall j \neq i \quad (3)$$

这比较容易理解，因为大多数情况下我们总是希望在当前特征下做出最有可能的决策。

2.3 最小风险贝叶斯决策

在实际应用中，不同类型的错误造成的后果严重程度不同（例如误诊癌症），因此引入损失函数（Loss Function）的概念。定义 $\lambda(\alpha_i|\omega_j)$ 为当真实类别为 ω_j 时采取决策 α_i 所带来的损失。

我们的目标是最小化期望风险（Expected Risk）。对于观测样本 x ，采取决策 α_i 的条件风险定义为：

$$E_{\omega|x}[\lambda(\alpha_i|\omega)] = \sum_{j=1}^c \lambda(\alpha_i|\omega_j) \underbrace{P(\omega_j|x)}_{\text{在该条件下 } \omega \text{ 发生的概率}} \quad (4)$$

决策规则：选择风险最小的决策 a ：

$$a = \arg \min_{j=1, \dots, c} R(\alpha_j|x) \quad (5)$$

特例：0-1 损失函数如果损失函数定义为“对则无损，错则罚 1”：

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \quad (6)$$

此时，最小风险决策完全等价于最小错误率贝叶斯决策（即最大后验概率决策 MAP）。

2.4 分类器设计

分类器可以看作是一个计算一组判别函数 (**Discriminant Functions**) $g_i(x)$ 并选择最大值的机器。

决策规则: 如果 $g_i(x) > g_j(x), \forall j \neq i$, 则将 x 归类为 ω_i 。

判别函数的形式:

1. **基础形式:** 常用的判别函数形式有:

- $g_i(x) = P(\omega_i|x)$ (直接使用后验概率)
- $g_i(x) = p(x|\omega_i)P(\omega_i)$ (去掉分母 $p(x)$)
- $g_i(x) = \ln p(x|\omega_i) + \ln P(\omega_i)$ (取对数, 变乘法为加法)

2. **更一般形式 (General Case):** PPT 中给出了判别函数的更一般定义:

$$g_i(x) = f(p(x|\omega_i)) + h(x, \omega_i) \quad (7)$$

其中:

- $f(\cdot)$ 是单调函数 (通常取 \ln), 用于变换似然度。
- $h(x, \omega_i)$ 是包含先验概率 $P(\omega_i)$ 和损失 λ 的项。

3. **注:** 与神经网络的联系这与现代神经网络最后的判别器 (**Discriminator**) 设计本质一致。神经网络最后一层 (Logits) 输出的值 $z_i = w_i^T x + b_i$, 实际上就是在拟合这个 $g_i(x)$ 。其中 $w_i^T x$ 对应特征的似然度部分, 而偏置 b_i 往往隐含了先验概率 $P(\omega_i)$ 的信息。

2.5 高斯密度下的判别函数

当类条件概率密度 $p(x|\omega_i)$ 服从多元正态分布 $N(\mu_i, \Sigma_i)$ 时, 判别函数具有解析解。

多元正态分布概率密度公式:

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (8)$$

取对数形式的判别函数 $g_i(x)$ 为:

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i) \quad (9)$$

根据协方差矩阵 Σ_i 的不同, 分为三种情形:

2.5.1 情形一: $\Sigma_i = \sigma^2 I$ (各特征独立且方差相等)

此时各类协方差矩阵相同且为对角阵, 几何上数据分布呈正球体, 大小相同。判别函数可简化。

1. **先验概率相等 ($P(\omega_i) = P(\omega_j)$):**

• **判别依据:** 此时 $\ln P(\omega_i)$ 项为常数可忽略, 决策规则简化为最小化欧氏距离平方:

$$g_i(x) \propto -\|x - \mu_i\|^2$$

- 分类器名称：最小距离分类器 (Minimum Distance Classifier)。
- 决策边界：连接两类均值向量 μ_i 和 μ_j 线段的垂直平分面，经过中点 $x_0 = \frac{1}{2}(\mu_i + \mu_j)$ 。

2. 先验概率不相等 ($P(\omega_i) \neq P(\omega_j)$)：

- 判别依据：依然是线性判别函数，但包含偏置项：

$$g_i(x) = \frac{1}{\sigma^2} \mu_i^T x - \frac{1}{2\sigma^2} \mu_i^T \mu_i + \ln P(\omega_i)$$

- 决策边界：决策面依然与 $\mu_i - \mu_j$ 正交，但不再经过连线中点。
- 偏移规律：决策面会向先验概率较小的一侧偏移（即先验概率大的类别占据更大的决策区域）。偏移量为：

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\mu_i - \mu_j)$$

2.5.2 情形二： $\Sigma_i = \Sigma$ (协方差矩阵相等)

此时各类数据分布呈超椭球体，形状和方向相同（平行），但中心不同。

- 1. 先验概率相等 ($P(\omega_i) = P(\omega_j)$)：
- 判别依据：最小化马氏距离 (Mahalanobis Distance) 平方：

$$r^2 = (x - \mu_i)^T \Sigma^{-1} (x - \mu_i) \quad (10)$$

- 决策边界：经过均值连线中点的超平面。但由于 Σ 不再是单位阵，决策面通常不与均值连线正交。

- 2. 先验概率不相等 ($P(\omega_i) \neq P(\omega_j)$)：
- 判别依据：线性判别函数 $g_i(x) = w_i^T x + w_{i0}$ ，其中 $w_i = \Sigma^{-1} \mu_i$ 。
- 决策边界：决策面 x_0 同样会偏离中点，向先验概率较小的一方偏移。

2.5.3 情形三： $\Sigma_i \neq \Sigma_j$ (协方差矩阵任意)

- 几何意义：各类数据的分布形状、方向、大小都不一样。
- 判别函数：无法消除二次项，形式为 $x^T W_i x + w_i^T x + w_{i0}$ ，是二次判别函数。
- 决策边界：超二次曲面。根据协方差矩阵的差异，可能表现为超球面、超椭球面、超双曲面或超平面等多种形态。

2.6 错误率分析

贝叶斯错误率 (Bayes Error) 是理论上分类器能达到的最低错误率。

$$P(\text{error}) = \int P(\text{error}|x)p(x)dx \quad (11)$$

对于两类问题，错误率等于两个类条件概率密度曲线重叠部分的面积。特征的分离度越好，重叠越小，贝叶斯错误率越低。

3 概率密度函数估计 (6 学时张燕明)

- 3.1 基本概念
- 3.2 最大似然估计
- 3.3 贝叶斯估计
- 3.4 正态分布下的贝叶斯估计
- 3.5 贝叶斯学习
- 3.6 特征维数问题
- 3.7 期望最大法 (EM)
- 3.8 隐马尔可夫模型

4 非参数法 (3 学时张燕明)

4.1 密度估计

4.2 Parzen 窗方法

4.3 K 近邻估计

4.4 最近邻分类器

4.5 K 近邻分类器的改进

4.6 距离度量

5 线性分类器设计 (3 学时张燕明)

- 5.1 引言
- 5.2 线性判别函数与决策面
- 5.3 广义线性判别函数
- 5.4 感知准则函数
- 5.5 松驰方法
- 5.6 线性最小二乘方法
- 5.7 多类线性判别函数

6 神经网络和深度学习 (9 学时张燕明)

- 6.1 人工神经网络发展历程
- 6.2 人工神经网络基础
- 6.3 单层前馈神经网络
- 6.4 多层感知器与误差反向传播算法
- 6.5 BP 算法讨论
- 6.6 反馈神经网络
- 6.7 径向基函数网络
- 6.8 自组织映射
- 6.9 深度学习简介
- 6.10 波尔兹曼机简介
- 6.11 自编码器
- 6.12 卷积神经网络
- 6.13 Recurrent Neural Network

7 特征提取与选择 (6 学时张燕明)

7.1 维数灾难

7.2 特征提取

7.3 线性方法

7.4 非线性方法

7.5 特征选择

8 模型选择 (3 学时张燕明)

- 8.1 模型选择原则
- 8.2 模型评价标准
- 8.3 基于偏差和方差的分类器设计准则
- 8.4 分类器集成
- 8.5 Adaboost 学习方法及其应用

9 聚类分析 (6 学时张燕明)

9.1 K 均值聚类

9.2 模糊 K 均值聚类

9.3 层次聚类

9.4 谱聚类

9.5 在线聚类

9.6 集成聚类

10 支持向量机与核方法 (6 学时向世明)

- 10.1 结构风险、经验风险与 VC 维
- 10.2 线性可分支持向量机与硬间隔最大化
- 10.3 线性支持向量机与软间隔最大化
- 10.4 非线性支持向量机
- 10.5 支持向量机的优化与扩展

11 决策树方法 (3 学时向世明)

11.1 决策树模型、学习与特征选择

11.2 决策树生成与剪枝 (ID3, C4.5, CART 等算法)

11.3 随机森林算法

12 模式识别前沿趋势 (3 学时向世明)

12.1 课程内容总体回顾

12.2 模式识别发展现状与新动态

12.3 模式识别中的挑战性问题

13 考核 (6 学时向世明)

13.1 考试 + 答疑

附录：模式识别常用 LaTeX 模板

1. 学术三线表模板

表 1: 不同分类器在 Iris 数据集上的错误率对比

算法	训练集错误率	测试集错误率	计算耗时 (ms)
KNN ($K = 3$)	0.02	0.04	12
Fisher LDA	0.05	0.06	2
SVM (RBF Kernel)	0.01	0.03	45

2. 图片插入模板 (已注释)

3. 模式识别核心公式模板

贝叶斯决策规则： 后验概率计算：

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}$$

最小错误率决策：若 $P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x})$, 则判定 $\mathbf{x} \in \omega_i$ 。

多元正态分布 (Gaussian Density)：

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

线性判别函数 (Linear Discriminant)：

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

最大似然估计 (MLE)：

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \prod_{k=1}^N p(\mathbf{x}_k|\boldsymbol{\theta}) \quad \Rightarrow \quad \hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \sum_{k=1}^N \ln p(\mathbf{x}_k|\boldsymbol{\theta})$$

SVM 优化目标 (硬间隔)：

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{s.t.} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, N$$

特征值分解 (PCA)： 协方差矩阵 \mathbf{S} 的特征方程：

$$\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

K-Means 目标函数：

$$J = \sum_{j=1}^K \sum_{\mathbf{x} \in C_j} \|\mathbf{x} - \mathbf{m}_j\|^2$$