**《模式识别实验指导手册（2017版）》**

（信息工程专业适用）

**学时： 8 课程编号： 0424200**

**一、目的与任务**

将基本原理与应用直接联系起来，使学生巩固和加深对《模式识别》理论教学中重要概念和理论的理解。使学生能逐步由基本理论过渡到实际应用，指导学生基本的实验技能和测试方法。本实验以MATLAB仿真平台为工具，以标准的文本、图像等数据为实验样本。配合模式识别理论教学，使学生了解和掌握最小错误率贝叶斯决策、感知器准则算法、Fisher线性判别，以及基于K-L变换的人脸识别等内容。

**二、项目、要求与安排方式**

1、实验项目

实验一：最小错误率贝叶斯决策实验

实验二：感知器准则算法实验

实验三：Fisher线性判别实验

实验四：基于K-L变换的人脸识别实验

2、基本要求

掌握模式识别的经典算法：最小错误率贝叶斯决策方法、感知器准则算法、Fisher线性判别方法，并实现基于K-L变换的人脸识别。

3、实验安排方式

本实验以MATLAB仿真平台为工具，配合模式识别理论教学，使学生了解和掌握统计模式识别的重要方法，并与理论分析所获得的结果进行比较。学生每人一组，根据实验要求，在预习实验指导手册与教材相应原理的基础上，独立地编写实验程序，并分析实验结果，完成实验报告。

**三、综合成绩的评定方法**

通过实验态度、操作技能、实验分析能力，以及实验报告的文案情况综合评分，用百分制给出成绩。

**起草人：蒋斌**

**实验一 最小错误率贝叶斯决策实验**

**实验学时：2 实验类型：验证**

**一、目的与任务**

目的：使学生掌握基本的统计决策方法，熟悉基于概率的分类方法，为复杂模型（例如：高斯混合模型、隐马尔科夫模型等）的学习打下基础。

任务：完成最小错误率贝叶斯决策方法的代码编写，实现对文本数据的分类。

**二、内容、要求与安排方式**

1、实验内容与要求：完成最小错误率贝叶斯决策方法，实现对文本数据的分类。完成实验报告。

2、实验安排方式：本实验以MATLAB仿真平台为工具，配合模式识别理论教学，使学生了解和掌握统计模式识别的重要方法，并与理论分析所获得的结果进行比较。学生每人一组，根据实验要求，在预习实验指导手册与教材相应原理的基础上，独立地编写实验程序，并分析实验结果，完成实验报告。

**三、实验设备**

1、所用设备：PC机、MATLAB软件。

2、消耗性器材：无。

**实验二 感知器准则算法实验**

**实验学时：2 实验类型：验证**

**一、目的与任务**

目的：实现模式识别中经典的线性判别方法，从基于样本的分类角度，掌握感知器准则函数。进而熟悉BP神经网络方法的基本内容。

任务：编制程序实现感知器准则算法。

**二、内容、要求与安排方式**

1、实验内容与要求：通过编制程序，完成感知器准则算法，并实现线性可分样本的分类。完成实验报告。

2、实验安排方式：本实验以MATLAB仿真平台为工具，配合模式识别理论教学，使学生了解和掌握统计模式识别的重要方法，并与理论分析所获得的结果进行比较。学生每人一组，根据实验要求，在预习实验指导手册与教材相应原理的基础上，独立地编写实验程序，并分析实验结果，完成实验报告。

**三、实验设备**

1、所用设备：PC机、MATLAB软件。

2、消耗性器材：无。

**实验三 Fisher线性判别实验**

**实验学时：2 实验类型：验证**

**一、目的与任务**

目的：应用统计方法解决模式识别问题的困难之一是维数问题，低维特征空间的分类问题一般比高维空间分类问题简单。Fisher线性判别方法是从数据降维的角度，找到线性可分样本的投影方向，使得降维后的样本不但线性可分，而且可分性更好（即不同类别的样本之间的距离尽可能远，同一类别的样本尽可能集中分布）。本实验通过编制程序让初学者能够体会 Fisher线性判别的基本思路，理解线性判别的基本思想，掌握Fisher线性判别问题的实质。

任务：编制程序实现Fisher线性判别算法。

**二、内容、要求与安排方式**

1、实验内容与要求：通过编制程序，完成Fisher线性判别算法，并实现对线性可分样本的降维分类。完成实验报告。

2、实验安排方式：本实验以MATLAB仿真平台为工具，配合模式识别理论教学，使学生了解和掌握统计模式识别的重要方法，并与理论分析所获得的结果进行比较。学生每人一组，根据实验要求，在预习实验指导手册与教材相应原理的基础上，独立地编写实验程序，并分析实验结果，完成实验报告。

**三、实验设备**

1、所用设备：PC机、MATLAB软件。

2、消耗性器材：无。

**实验四 基于K-L变换的人脸识别实验**

**实验学时：2 实验类型：综合**

**一、目的与任务**

目的：掌握特征提取的基本方法，培养学生灵活使用KL变换进行模式识别的能力。

任务：编制程序基于K-L变换的人脸识别。

**二、内容、要求与安排方式**

1、实验内容与要求：选取人脸数据库（例如：ORL数据库）中的部分样本（每个人的前N张图片）作为训练样本，其余为未知的测试样本。从训练样本中得到K-L变换矩阵；然后对训练样本和测试样本都进行变换；最后用变换后的数据作最近邻识别，统计识别率，并基于MATLAB GUI设计系统界面，实现基于K-L变换的人脸识别系统。完成实验报告。

2、实验安排方式：本实验以MATLAB仿真平台为工具，配合模式识别理论教学，使学生了解和掌握统计模式识别的重要方法，并与理论分析所获得的结果进行比较。学生每人一组，根据实验要求，在预习实验指导手册与教材相应原理的基础上，独立地编写实验程序，并分析实验结果，完成实验报告。

**三、实验设备**

1、所用设备：PC机、MATLAB软件。

2、消耗性器材：无。

**实验一 最小错误率贝叶斯决策**

1. **实验原理**

对于具有多个特征参数的样本（如本实验的iris数据样本有4个参数），其正态分布的概率密度函数可定义为

 (1)

式中，和都是d维行向量，是d×d维协方差矩阵，是的逆矩阵，是的行列式。本实验我们采用最小错误率的贝叶斯决策，使用如下的函数作为判别函数

 (2)

其中为类别发生的先验概率，为类别的类条件概率密度函数。由其判决规则可知，如果使对一切成立，则将归为类。如果服从正态分布，公式(1)写成

 (3)

对公式(3)右端取对数，可得

(4)

公式(4)中的第二项与样本所属类别无关，将其从判别函数中消去，不会改变分类结果。则判别函数可简化为以下形式

 (5)

**二、实验步骤**

（1）从iris.txt 文件（课程邮箱-文件中心）中读取估计参数用的样本，每一类样本抽出前40个，分别求其均值；

（2）求每类样本的协方差矩阵、逆矩阵以及协方差矩阵的行列式；

（3）对三个类别，分别取每组剩下的 10个样本，每两组进行分类。由于每类样本都相等，且每类选取用作训练的样本也相等，在每两组进行分类时，待分类样本的类先验概率为0.5。将各个样本代入判别函数既公式(5)，进行分类。

**三、实验要求**

（1）复习最小错误率贝叶斯决策原理，写出实验代码，实现对三类样本的分类；

（2）计算分类的正确率，画出三维空间的样本分类图；

（3）分析实验结果，完成实验报告。

**实验二 Fisher线性判别实验**

**一、实验原理**

低维特征空间的分类问题一般比高维空间分类问题简单，Fisher线性判别方法就是将高维样本投影到低维特征空间，以便于实现样本分类的一种方法。具体思想是构建一个数据的投影方向，使得投影后的同类样本之间的距离最小，不同类样本之间的距离最大。换言之，就是不同类别的样本分布尽可能稀疏，同一类别的样本分布尽可能密集。

**二、实验步骤**

（1）给出两类实验数据如表2-1所示，计算两类样本的投影方向；

（2）画出投影方向W的直线，并标记出投影后的点在直线上的位置；

**表2-1 实验数据**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **类别** | *W1* | | | *W2* | | |
| **样本** | *X1* | *X2* | *X3* | *X1* | *X2* | *X3* |
| 1 | -0.4 | 0.58 | 0.089 | 0.83 | 1.6 | -0.014 |
| 2 | -0.31 | 0.27 | -0.04 | 1.1 | 1.6 | 0.48 |
| 3 | -0.38 | 0.055 | -0.035 | -0.44 | -0.41 | 0.32 |
| 4 | -0.15 | 0.53 | 0.011 | 0.047 | -0.45 | 1.4 |
| 5 | -0.35 | 0.47 | 0.034 | 0.28 | 0.35 | 3.1 |
| 6 | 0.17 | 0.69 | 0.1 | -0.39 | -0.48 | 0.11 |
| 7 | -0.011 | 0.55 | -0.18 | 0.34 | -0.079 | 0.14 |
| 8 | -0.27 | 0.61 | 0.12 | -0.3 | -0.22 | 2.2 |
| 9 | -0.065 | 0.49 | 0.0012 | 1.1 | 1.2 | -0.46 |
| 10 | -0.12 | 0.054 | -0.063 | 0.18 | -0.11 | -0.49 |

（3）选择决策边界，实现新样本 *XX1 =* ( -0.7, 0.58, 0.089 )和 *XX2 =* ( 0.047, -0.4, 1.04 ) 的分类；

（4）再给出一类新实验数据如表2-2所示，求新类别*W3*分别和*W1、W2*分类的投影方向和分类阈值；

**表2-2 新实验数据**

|  |  |  |  |
| --- | --- | --- | --- |
| **类别** | *W3* | | |
| **样本** | *X1* | *X2* | *X3* |
| 1 | 1.58 | 2.32 | -5.8 |
| 2 | 0.67 | 1.58 | -4.78 |
| 3 | 1.04 | 1.01 | -3.63 |
| 4 | -1.49 | 2.18 | -3.39 |
| 5 | -0.41 | 1.21 | -4.73 |
| 6 | 1.39 | 3.61 | 2.87 |
| 7 | 1.2 | 1.4 | -1.89 |
| 8 | -0.92 | 1.44 | -3.22 |
| 9 | 0.45 | 1.33 | -4.38 |
| 10 | -0.76 | 0.84 | -1.96 |

**三、实验要求**

（1）复习Fisher线性判别方法的原理，写出实验代码，实现实验步骤中规定的分类任务；

（2）分析实验结果，完成实验报告。

**实验三 感知器准则算法实验**

**一、实验原理**

与实验二相同，感知器也属于线性分类方法。单层感知器是一个具有一层神经元、采用阈值激活函数的前向网络。通过对网络权值的训练，可以使感知器对一组输入矢量的响应达到元素为0或1的目标输出，从而实现对输入矢量分类的目的。

**二、实验步骤**

（1）给出四类实验数据如表3-1所示，其中每个样本数据为两维，X1表示第一维的值、X2表示第二维的值. 实现两类样本（例如：*W1、W2*类，*W2、W3*类）的分类，并记录感知器的分类性能；

**表3-1 实验数据**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类别 | *W1* | | *W2* | | *W3* | | *W4* | |
| 样本 | *X1* | *X2* | *X1* | *X2* | *X1* | *X2* | *X1* | *X2* |
| 1 | 0.1 | 1.1 | 7.1 | 4.2 | -3.0 | -2.9 | -2.0 | -8.4 |
| 2 | 6.8 | 7.1 | -1.4 | -4.3 | 0.5 | 8.7 | -8.9 | 0.2 |
| 3 | -3.5 | -4.1 | 4.5 | 0.0 | 2.9 | 2.1 | -4.2 | -7.7 |
| 4 | 2.0 | 2.7 | 6.3 | 1.6 | -0.1 | 5.2 | -8.5 | -3.2 |
| 5 | 4.1 | 2.8 | 4.2 | 1.9 | -4.0 | 2.2 | -6.7 | -4.0 |
| 6 | 3.1 | 5.0 | 1.4 | -3.2 | -1.3 | 3.7 | -0.5 | -9.2 |
| 7 | -0.8 | -1.3 | 2.4 | -4.0 | -3.4 | 6.2 | -5.3 | -6.7 |
| 8 | 0.9 | 1.2 | 2.5 | -6.1 | -4.1 | 3.4 | -8.7 | -6.4 |
| 9 | 5.0 | 6.4 | 8.4 | 3.7 | -5.1 | 1.6 | -7.1 | -9.7 |
| 10 | 3.9 | 4.0 | 4.1 | -2.2 | 1.9 | 5.1 | -8.0 | -6.3 |

（2）分别记录感知器，针对不同类别样本分类时的收敛步数，解释它们收敛步数的差别；

（3）*W3、W4*类的前5个样本不是线性可分的，请问如何构造这两类样本的感知器？

**三、实验要求**

（1）复习感知器准则方法的原理，写出实验代码，实现实验步骤中规定的分类任务；

（2）分析实验结果，完成实验报告。

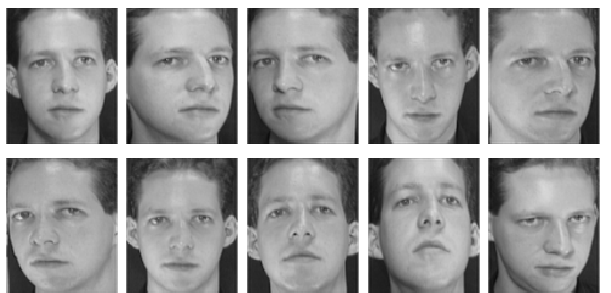
**实验四 基于K-L变换的人脸识别实验**

**一、实验原理**

基于K-L变换的方法在人脸识别中也称为PCA（Principal Component Analysis）方法，中文译作“主元分析”或“主分量分析”。PCA通过线性变换将原始数据变换为一组各维度线性无关的表示，可用于提取数据的主要特征分量。本次实验不是单纯的验证原理，而是一个综合性能的实验。大家已经了解了模式识别系统，知道模式识别系统的构成，在给定实验数据以后，希望同学们采用PCA方法实现特征提取，而分类方法则由同学们自主选择，比如：近邻分类法、神经网络分类器、支持向量机，都可以尝试。

**二、实验步骤**

（1）给大家提供一个常用的人脸数据库ORL数据库（课程邮箱-文件中心）。



**图4-1 ORL数据库人脸图像示例**

（2）编写PCA代码，构建训练集与测试集，标注人脸图像的类别标记，实现特征提取；

（3）选择分类器，实现人脸分类，并计算分类正确率。

**三、实验要求**

（1）复习K-L方法和已学分类器的原理，写出实验代码，实现实验步骤中规定的分类任务；

（2）分析实验结果，完成实验报告。