**班 级 1402052**

**学 号 14020520013**

****

本科课程设计报告



**题 目** Fisher辨别分析用于

人脸数据维数约简的实现

**学 院** 电子工程学院

**专 业**  智能科学与技术

**学生姓名**  董龙锐

**导师姓名** 冯婕

**西 安 电 子 科 技 大 学**

**电 子 工 程 学 院**

**课 程 设 计（报告）任 务 书**

学生姓名 董龙锐 指导教师 冯婕 职称 副教授

学生学号 14020520013 专业 智能科学与技术

题目 Fisher辨别分析用于人脸数据维数约简的实现

相关专业课程 《模式识别》等

任务与要求

|  |
| --- |
| 1、利用网络和图书馆查阅Fisher辨别分析的相关知识；  2、学习掌握Python编程语言；  3、掌握Fisher辨别分析的理论知识以及了解数据集的维数约简，投影等相关知识；  4、利用Python语言编程实现Fisher辨别分析用于人脸数据集的维数约简以及分类的验证，分析实验结果；  5、通过编程实验练习让学生了解Fisher辨别分析的知识并且认真撰写课程设计报告。 |

开始日期 2017年 1 月 2 日 完成日期 2017年 1 月 26 日

课程设计所在单位 智能科学与技术系 2017 年 1 月 2 日

Fisher辨别分析用于人脸数据维数约简的实现

**摘要**：Fisher辨别分析是模式识别领域的经典算法之一，它又称线性判别分析（Linear Discriminant Analysis, LDA）。本文基于Python语言，选取YaleB、AT&T (Olivettti)作为人脸数据集，利用Fisher辨别分析算法对两个数据集进行维数约简；然后对降维后的人脸数据进行分类，并对分类结果进行定量分析，最后做出讨论和总结。

**关键词**：Fisher辨别分析，维数约简，分类，人脸数据

**Abstract**: Fisher discriminant analysis is one of the classical algorithms in the field of Pattern Recognition, which is also called Linear Discriminant Analysis (LDA). Based on the Python programing language, this paper selects YaleB, AT&T (Olivettti) as the face data set and implements Fisher discriminant analysis algorithm to reduce the dimensionality of those two data sets; After that, the transformed face data are classified by LDA classifier, and the result of the classification are quantitatively analyzed. Finally, this paper makes a discussion and conclusion.

**Keyword**: Fisher discriminant analysis, Dimensionality reduction, Classification, Face data

1. **引言**

在如今这个互联网+的新时代，可能最不稀有的就是数据了。因为数据的获取变得越来越容易且经济。但随之产生的负面后果是数据呈现出高维度、大规模的特点。举一个简单的例子[]，在科学研究中，我们常常要对数据进行处理，而这些数据通常位于一个高维空间中，例如当处理一个256\*256 的图像序列时，我们需要将其拉成一个向量，这样，我们就得到了4096 维的数据，如果直接对这些数据进行处理，会有以下问题：首先，会出现所谓的“维数灾难”问题，巨大的计算量将使我们无法忍受；其次，这些数据通常没有反映出数据的本质特征，如果直接对他们进行处理，不会得到理想的结果。为了高效的分析这些数据，一个主要的途径是维数约简，然后对约简后的数据进行处理。关键是要保证约简后的数据特征能反映甚至更能揭示原数据的本质特征。作为机器学习中的经典问题之一，维数约简主要用于处理维数灾难问题、帮助加速算法的计算效率和提高可解释性以及数据可视化[]。该方向的研究分为无监督维数约简和有监督维数约简等。

那么维数约简方法都有哪些呢？数据维数约简的方法可以分为线性维数约简和非线性维数约简，而非线性维数约简又分为基于核函数的方法和基于特征值的方法。线性维数约简的方法主要有主成分分析（PCA）、[独立成分分析](http://www.baike.com/sowiki/%E7%8B%AC%E7%AB%8B%E6%88%90%E5%88%86%E5%88%86%E6%9E%90?prd=content_doc_search" \o "独立成分分析)（ICA）、线性判别分析（LDA）、局部特征分析（LFA）等等。

本文用到的就是经典的Fisher辨别分析（LDA）。线性判别分析是对费舍尔的线性鉴别方法的归纳[]，这种方法使用统计学，模式识别和机器学习方法，试图找到两类物体或事件的特征的一个线性组合，以能够特征化或区分它们。所得的组合可用来作为一个线性分类器，或者，更常见的是，为后续的分类做维数约简。为了展示该算法，本文选择YaleB、AT&T (Olivettti)两个经典的人脸数据集作为实验数据集。因为人脸数据的维数相对较高，可以检验实际降维效果。我们先用LDA算法的降维功能对人脸数据进行维数约简，然后再使用LDA的分类器功能对降维后的数据进行分类。最后对分类结果进行定量分析。分类的效果直接反映出维数约简的好坏：若维数约简后的数据仍能被准确的分类，则说明这样的降维揭示了数据的本质特征；反之则降维丢失了原始数据的重要信息。

1. **基础原理**
2. **方案设计**
3. **结果与总结**
4. **参考文献**
5. **代码附录**