

# lab3. fisher LDA

[多分类问题](#)

[代码原理与实现](#)

[项目代码](#)

[原理介绍](#)

[测试过程和运行结果](#)

[实验总结和尚未解决的问题](#)

[数据预处理问题](#)

[模型泛化性较差](#)

[REFERENCE](#)

## 多分类问题

该问题是一个多分类问题。Slides中给出的方法适用于二分类问题。**参考的博客**（见报告结尾）中给出了计算多个类别的类间离散度和类内离散矩阵的方法。方法如下：

计算多类别的类间离散度和类内离散度

$$S_b = \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^T$$
$$S_w = \sum_{i=1}^c \sum_{x_k \in \text{class}_i} (\mu_i - x_k)(\mu_i - x_k)^T$$

目标函数为

$$J(w) = \frac{J_B}{J_W} = \frac{w^T S_B w}{w^T S_W w}$$

假设分母为1，那么问题转化为

$$\begin{cases} \max_w w^T S_B w \\ 1 - w^T S_W w = 0 \end{cases}$$

使用拉格朗日定理，得

$$F(w) = w^T S_B w + \lambda(1 - w^T S_W w)$$

求导求极值可解得

$$S_W^{-1} S_B w = \lambda w$$

接着求特征向量即可。求得投影方向后再将结果输入sklearn得Perceptron即可。

## 代码原理与实现

### 项目代码

```
.
├── main.py          # project entry
└── LDA.py           # class LDA definition
```

### 原理介绍

class LDA中提供了一下成员变量和成员函数：

`n_components`, `data`, `labels`, `all_labels`, `label_nums`, `projection_direction`，成员函数在代码中均有注释说明。

`mean_of_class`：计算某类的均值

`between_class_scatter`：计算类间离散度

`within_class_scatter`：计算类内离散度

`fit`：根据数据集进行拟合

`transform`：根据拟合得到的模型对数据进行降维变换

注释中有更详细的说明。

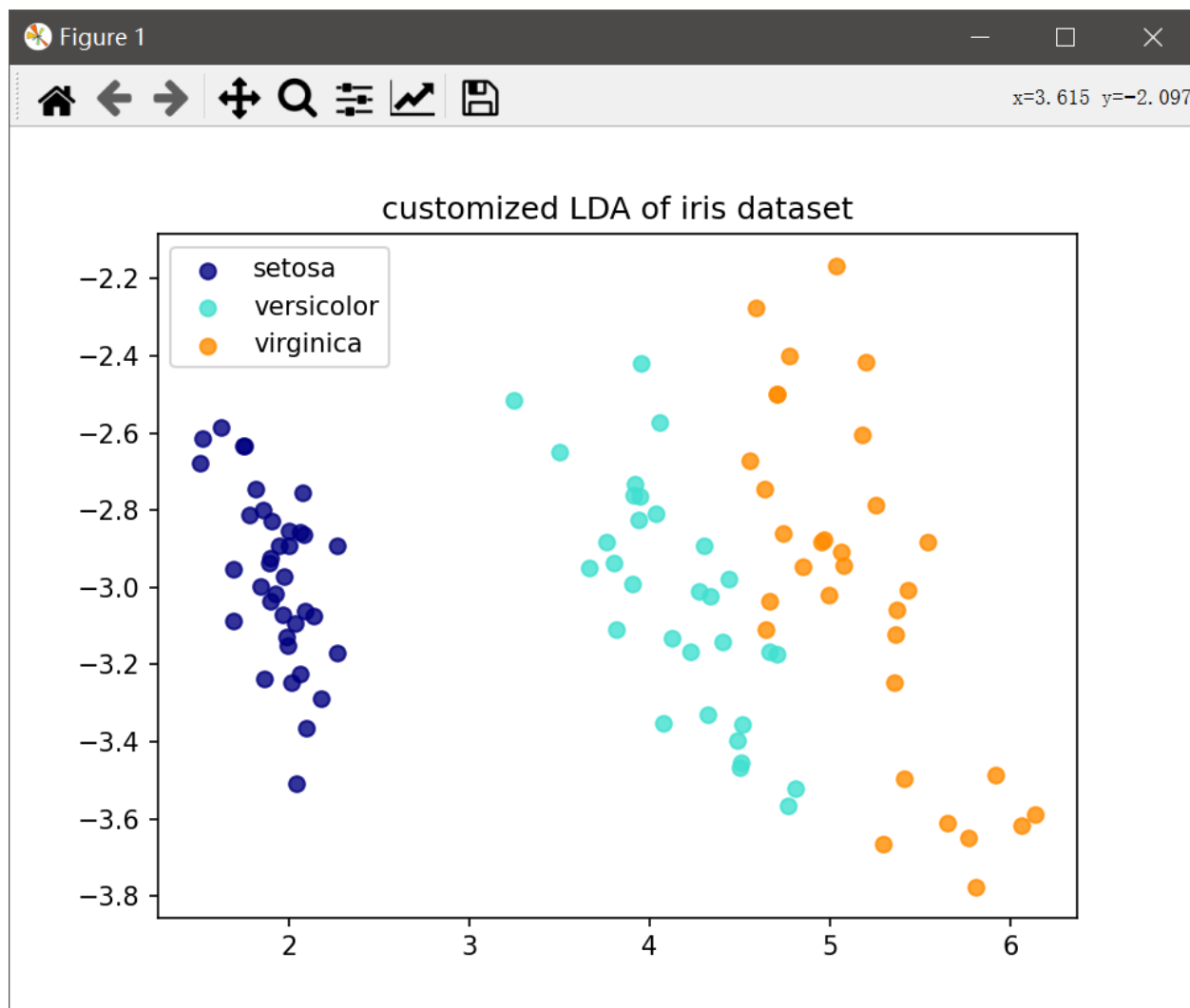
除此外，在main文件中还实现了

`split_dataset`：划分训练集和测试集

`standardize`：对降维后的数据进行Z-score标准化

## 测试过程和运行结果

`main.py` 为项目的入口，运行该文件，等待程序结果。会弹出降维后的结果的图像以及在终端输出准确率。



```
D:\anaconda\envs\pr\python.exe "E:\23FA\Pattern Recognition\lab\lab3\main.py"
train accuracy is 0.94
test accuracy is 0.9

Process finished with exit code 0
```

## 实验总结和尚未解决的问题

### 数据预处理问题

实验最初通过实现的LDA类进行降维后将降维后的数据输入sklearn的Perception中后，得到的准确率较低，甚至是训练准确率。接着将降维后的数据进行一定处理后，尝试改进。

结果：发现即使对数据进行Z-score标准化后效果仍然一般。


### 模型泛化性较差

多次运行项目，发现准确率结果方差较大。有时效果很好，测试准确率能达到0.98，有时效果很差，测试准确率仅0.6左右。猜测可能是每次运行时对数据集进行训练集和测试集的随机划分不一样。。但是如果模型泛化性较好不应有此问题。。**所以应该模型还有优化的地方。那么如何优化呢？**

## REFERENCE


### LDA 线性判别分析 - liuwu265 - 博客园

LDA, Linear Discriminant Analysis, 线性判别分析。注意与LDA (Latent Dirichlet Allocation, 主题生成模型) 的区别。1、引入 上文介绍的PCA方法对提取样本数据的主要变化信息非常有效, 而忽略了次要变化的信息。在有些情况下, 次要信息可能正是把。

 <https://www.cnblogs.com/liuwu265/p/4724758.html>

### 线性判别分析(Linear Discriminant Analysis, LDA) 算法分析 - 迈克老狼2012 - 博客园

原文来自: <http://blog.csdn.net/xiazhaoliang/article/details/6585537> LDA算法入门 一. LDA算法概述: 线性判别式分析(Linear Discriminant Analysis, LDA), 也叫做Fisher线性判别(Fisher Lin

 <https://www.cnblogs.com/mikewolf2002/p/3435750.html>