

## 机器学习实习一

掌握基于 Python 语言，完成四个分类器（逻辑回归、支持向量机、线性判别分析、朴素贝叶斯）的使用。

### 一、 具体任务：

1. 掌握 sklearn 库的使用；
2. 对数据集进行可视化，利用课程所学的知识对数据集进行划分，并在实习报告中阐明具体的划分方法和代码；（划分方法可采用简单交叉验证或 K 折交叉验证，K 折交叉验证为额外加分项！）
3. 利用训练集分别采用逻辑回归、支持向量机、线性判别分析、朴素贝叶斯的方法，对分类器进行训练；

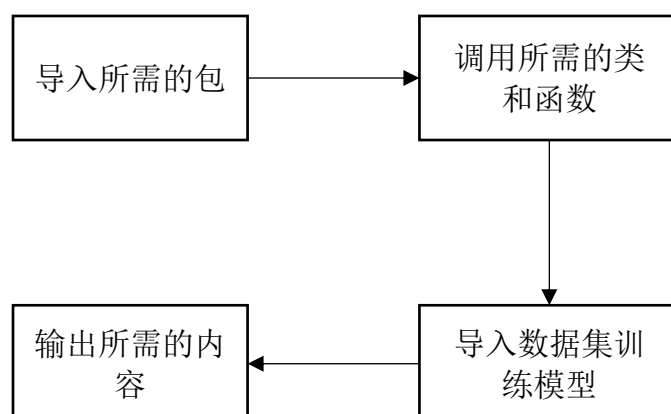


图 1 分类器使用流程

4. 输出各分类器的对应参数（如图 2），对各分类器的决策边界分别在训练集和测试集上进行可视化输出（如图 3）；

```
预测结果 : [0 1]
概率分布 : [[0.76977862 0.23022138]
             [0.02499289 0.97500711]]
特征系数 : [-0.46335005]
截距 : [[ 2.80731317 -0.65994503]]
```

图 2 分类器的参数输出

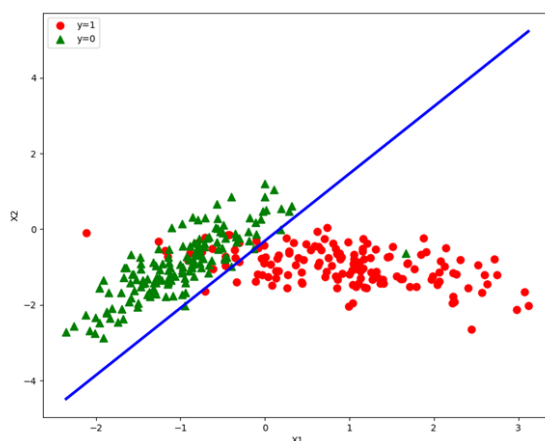


图 3 决策边界的可视化

5. 计算并输出各分类器的混淆矩阵及查准率、查全率；
6. 自己定义不同的阈值，对各分类器绘制 ROC 曲线。输出自定义的不同阈值、阈值对应的 TPR 和 FPR，ROC 曲线的可视化及基于 AUC 的评价结果。

## 二、 实习报告提交内容及排版要求：

1. 实习任务中要求输出的所有结果；
2. 文字总结使用不同分类器的过程，编程过程中遇到的问题，解决方法；
3. 子标题采用宋体小三加粗，正文用宋体 4 号，要求排版美观，图表均需要有图名图号或表名表号。

4. 代码要求进行注释!!!

5. 代码文件命名：学号\_代码.\*

报告提交 PDF 版本，文件名为：学号\_报告.\*

统一打包后提交，文件夹名称：学号\_姓名

（注：如有雷同，一律判无效作业!!!）

课程编号：

课程性质：公共基础必修

## 机器学习

## 实习报告

学院：\_\_\_\_\_

专业：\_\_\_\_\_

地点：\_\_\_\_\_

班级：\_\_\_\_\_

姓名：\_\_\_\_\_

学号：\_\_\_\_\_

教师：\_\_\_\_\_

2021 年 11 月 22 日