

机器学习第七章实验内容

详细内容见第六周实验内的 jupyter notebook。

一、实验内容

1.1 使用 sklearn 的 GaussianNB 完成 spambase 垃圾邮件分类任务

1. 使用 GaussianNB 完成 spambase 邮件分类
2. 计算十折交叉验证的精度、查准率、查全率、F1 值

1.2 使用 sklearn 的 BernoulliNB 完成 spambase 垃圾邮件分类任务

1. 使用 BernoulliNB 完成 spambase 邮件分类
2. 计算十折交叉验证的精度、查准率、查全率、F1 值

1.3 使用 sklearn 的 MultinomialNB 完成 spambase 垃圾邮件分类任务

1. 使用 MultinomialNB 完成 spambase 邮件分类
2. 计算十折交叉验证的精度、查准率、查全率、F1 值

1.4 实现一个高斯朴素贝叶斯分类器

1. 实现高斯朴素贝叶斯分类器
2. 计算模型的查准率，查全率，F1 值

1.5 实现带有拉普拉斯修正的朴素贝叶斯分类器（选做）

1. 叙述拉普拉斯修正的作用
2. 给出使用的数据集
3. 给出实现的代码，要有详细的注释
4. 给出模型评价指标的结果

二、数据介绍

2.1 spambase 垃圾邮件数据集

文件名：

1. 原始数据：spambase.data
2. 字段说明：spambase 数据说明.txt

数据来源：<http://archive.ics.uci.edu/ml/datasets/spambase>

数据无需预处理，读取即可使用

```
import numpy as np  
data = np.loadtxt('data/spambase/spambase.data', delimiter = ",")
```