# 学生实验报告

|  |  |  |  |
| --- | --- | --- | --- |
| 学号 | 1120182525 | 学院 | 徐特立学院 |
| 姓名 | 梁瑛平 | 专业 | 计算机科学与技术 |

## 支持向量机

## 实验简介

本实验采用支持向量机，对iris数据集进行分类预测，其目的是提升学生应用支持向量机的实际解决问题能力。

## 实验目的

（1）帮助学生理解支持向量机在数据科学中的应用。

（2）帮助学生掌握支持向量机的原理。

（3）帮助学生掌握利用sklearn实现支持向量机的方法。

## 相关理论与知识点

（1）支持向量机的原理

（2）sklearn中支持向量机的应用

## 实验条件与环境

|  |  |  |  |
| --- | --- | --- | --- |
| 要求 | 名称 | 版本要求 | 备注 |
| **编程语言** | python | 3.6以上 |  |
| **开发环境** | Jupter | 无要求 |  |
| **第三方工具包/库/插件** | sklearn | 0.23.1 |  |
| **第三方工具包/库/插件** | numpy | 1.16.2 |  |
| **其他工具** | 无 | 无要求 |  |
| **硬件环境** | 台式机、笔记本均可 | 无要求 |  |

## 实验任务

|  |  |  |
| --- | --- | --- |
| **序号** | **任务名称** | **任务具体要求** |
| **1** | 数据理解 | 理解数据集背景以及数据含义。 |
| **2** | 数据读入 | 可使用sklearn中自带的iris数据集：datasets.load\_iris() |
| **3** | 训练集和测试集划分 | 随机抽取总数据集70%作为训练集，测试集占30% |
| **4** | 支持向量机 | 构建SVM模型，并进行训练 |
| **5** | 预测 | 对测试集进行分类预测，输出预测值及准确率 |
| **6** | 分析 | 调整参数，对比分析实验结果，体现在实验报告中 |

## 实验结果及分析

SVM原理：

在机器学习中，支持向量机（英语：support vector machine，常简称为SVM，又名支持向量网络）是在分类与回归分析中分析数据的监督式学习模型与相关的学习算法。给定一组训练实例，每个训练实例被标记为属于两个类别中的一个或另一个，SVM训练算法创建一个将新的实例分配给两个类别之一的模型，使其成为非概率二元线性分类器。SVM模型是将实例表示为空间中的点，这样映射就使得单独类别的实例被尽可能宽的明显的间隔分开。然后，将新的实例映射到同一空间，并基于它们落在间隔的哪一侧来预测所属类别。

除了进行线性分类之外，SVM还可以使用所谓的核技巧有效地进行非线性分类，将其输入隐式映射到高维特征空间中。

当数据未被标记时，不能进行监督式学习，需要用非监督式学习，它会尝试找出数据到簇的自然聚类，并将新数据映射到这些已形成的簇。将支持向量机改进的聚类算法被称为支持向量聚类，当数据未被标记或者仅一些数据被标记时，支持向量聚类经常在工业应用中用作分类步骤的预处理。

支持向量聚类：支持向量聚类是一种创建在核函数上的类似方法，同适用于非监督学习和数据挖掘。它被认为是数据科学中的一种基本方法。

转导支持向量机

多元分类支持向量机：SVM算法最初是为二值分类问题设计的，实现多分类的主要方法是将一个多分类问题转化为多个二分类问题。常见方法包括“一对多法”和“一对一法”，一对多法是将某个类别的样本归为一类,其他剩余的样本归为另一类，这样k个类别的样本就构造出了k个二分类SVM；一对一法则是在任意两类样本之间设计一个SVM。

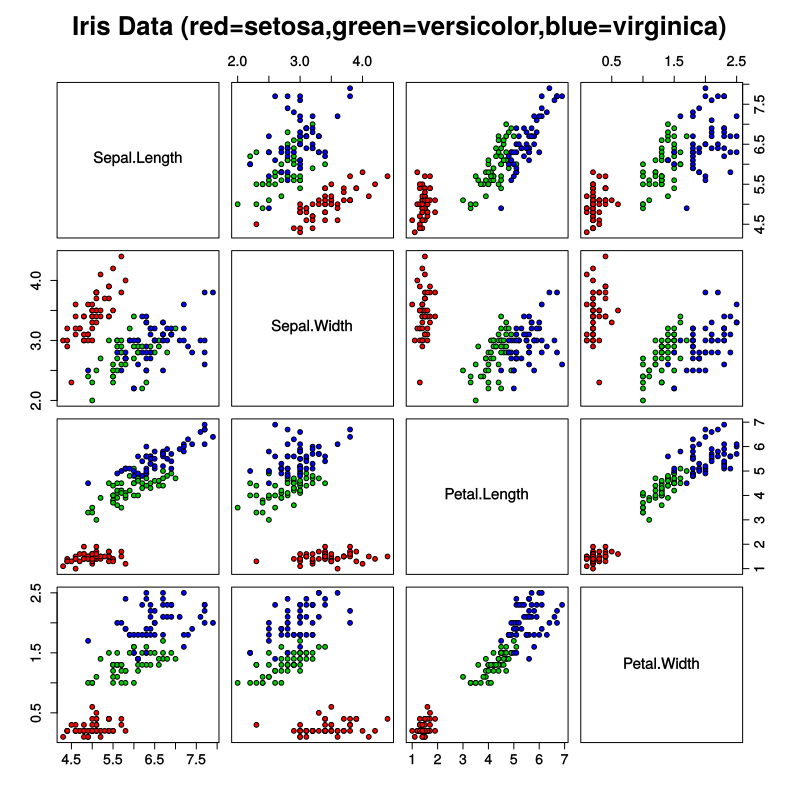
支持向量回归

结构化支持向量机：支持向量机可以被推广为结构化的支持向量机，推广后标签空间是结构化的并且可能具有无限的大小。

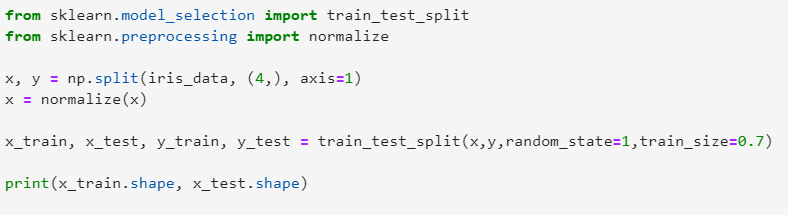
数据分析：

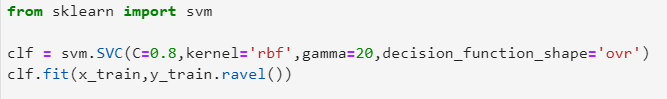
安德森鸢尾花卉数据集（英文：Anderson's Iris data set），也称鸢尾花卉数据集（英文：Iris flower data set）或费雪鸢尾花卉数据集（英文：Fisher's Iris data set），是一类多重变量分析的数据集。它最初是埃德加·安德森从加拿大加斯帕半岛上的鸢尾属花朵中提取的形态学变异数据，后由罗纳德·费雪作为判别分析的一个例子，运用到统计学中。

其数据集包含了150个样本，都属于鸢尾属下的三个亚属，分别是山鸢尾、变色鸢尾和维吉尼亚鸢尾。四个特征被用作样本的定量分析，它们分别是花萼和花瓣的长度和宽度。基于这四个特征的集合，费雪发展了一个线性判别分析以确定其属种。

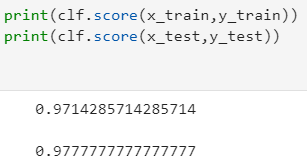


对于特征数据我进行normalize处理。





实验结果：



## 收获与体会

1. 学习了SVM的基本原理；
2. 在代码实践中掌握了算法的使用；

（3）提高了综合能力；

## 备注及其他

无