

实验五 支持向量机

一、说明

- 实验采用 jupyter notebook, 请填写完代码后提交完整的 ipynb 文件
- 文件命名规则: 班级_姓名_ML2018_HW5.ipynb, 如计科 1701_张三_ML2018_HW5.ipynb
- 提交方式: 采用在线提交至:
<http://pan.csu.edu.cn:80/invitation/6000b76f-8e40-4f8f-98d4-fdae74c00d0f>
- 实验提交截止日期: 2018.12.2 24: 00

二、实验内容

本次实验我们将利用 DataSet.txt 中的数据, 来训练一个 SVM 模型。该数据集包含两个特征以及一个标签, 在训练模型过程中, 我们对 SVM 对偶形式利用 SMO 算法进行求解, 获取最优的拉格朗日乘子, 并根据最终的求解答案绘制出决策边界。

大部分数据都不是线性可分的, 所以核函数就发挥出了作用, 我们利用径向基函数对原始数据进行映射, 使得 SVM 能够解决非线性可分的数据的问题。我们在 KernelTrainData.txt 数据集上进行训练, 并对 KernelTestData.txt 数据进行预测, 获得最终的错误率。以上两个数据集中特征数均为二, 且标签数均为一。

三、实验目标

- 掌握 SVM 基本原理
- 理解对偶问题的求解形式
- 理解 SMO 算法的求解过程及实现
- 理解核函数以及其实现

四、实验操作步骤

1. 启动 jupyter notebook

参考实验一，打开文件 SVM.ipynb。

2. 完成实验任务

任务 1 从 **DataSet.txt** 中导入数据，获得训练集以及标签

```
TrainData = [[3.542485, 1.977398], [3.018896, 2.556416], [7.55151, -1.58003]]  
TrainLabel = [-1.0, -1.0, 1.0]
```

任务 2 调整 **alpha** 的值

```
b = 11  
c = 8
```

任务 3 计算真实值与预测值之间的误差

```
Ek1 = -1.0  
Ek2 = 1.0
```

任务 4 选取最大 $|E_i - E_j|$ 最大的 j ，并返回 j 以及 E_j

```
j = 2  
Ej = -1.0
```

任务 5 计算误差值并存入缓存

```
TestOs.Cache[0] = [[1. 1.]]  
TestOs.Cache[1] = [[ 1. -1.]]  
TestOs.Cache[2] = [[1. 1.]]
```

任务 6 实现 SMO 算法内循环

任务 7 根据计算出的拉格朗日乘子计算出权重向量 **W**

```
Testw = [[-0.02568303] [ 0.04319313]]
```

任务 8 画出 **SVM** 的决策边界

任务 9 实现径向基核函数

```
Result = [[6.21201706e-01]  
[1.00000000e+00]  
[1.67499988e-13]  
[3.14534050e-03]]
```

[3.88031058e-10]]

任务 10 画出线性不可分数据图