中国地质大学(武汉)

机器学习第四次作业

姓 名：叶宇涛

专 业：计算机科学与技术

学 号：20191000595

指导老师：刘超

目录

[中国地质大学(武汉) 1](#_Toc101029596)

[机器学习第四次作业 1](#_Toc101029597)

[6 支持向量机 2](#_Toc101029598)

[6.2 在西瓜数据集3.0上分别用线性核和高斯核训练一个SVM. 2](#_Toc101029599)

[6.2.1 编程题目理解 2](#_Toc101029600)

[6.2.2 原理阐述 3](#_Toc101029601)

[6.2.3 算法设计思路 3](#_Toc101029602)

[6.2.4 实验流程、测试结果及分析 3](#_Toc101029603)

[6.2.5 代码结构，核心代码简要分析 3](#_Toc101029604)

[6.2.6 本次实验解决的主要问题，主要收获 3](#_Toc101029605)

[6.2.7 编码及内容撰写中的参考来源 3](#_Toc101029606)

# 支持向量机

## 在西瓜数据集3.0上分别用线性核和高斯核训练一个SVM.

### 编程题目理解

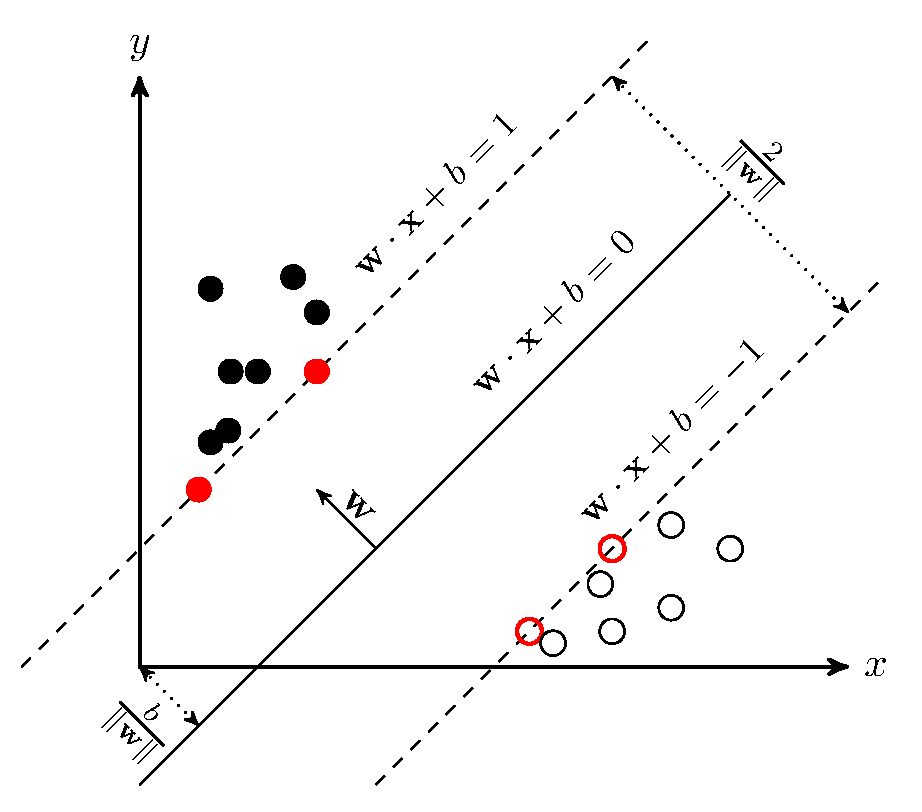
题目要求我们利用西瓜数据集3.0,分别利用不同的核函数：线性核以及高斯核，训练一个支持向量机。其中，给出西瓜数据集的内容如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 密度 | 含糖率 | 好瓜 |
| 1 | 0.697 | 0.46 | 1 |
| 2 | 0.774 | 0.376 | 1 |
| 3 | 0.634 | 0.264 | 1 |
| 4 | 0.608 | 0.318 | 1 |
| 5 | 0.556 | 0.215 | 1 |
| 6 | 0.403 | 0.237 | 1 |
| 7 | 0.481 | 0.149 | 1 |
| 8 | 0.437 | 0.211 | 1 |
| 9 | 0.666 | 0.091 | 0 |
| 10 | 0.243 | 0.267 | 0 |
| 11 | 0.245 | 0.057 | 0 |
| 12 | 0.343 | 0.099 | 0 |
| 13 | 0.639 | 0.161 | 0 |
| 14 | 0.657 | 0.198 | 0 |
| 15 | 0.36 | 0.37 | 0 |
| 16 | 0.593 | 0.042 | 0 |
| 17 | 0.719 | 0.103 | 0 |

表格 6-1 西瓜数据集3.0

而对于支持向量机SVM，基本的想法就是求解能够正确划分训练数据集，并且使几何间隔最大的分离超平面，是一种二分类模型。本题需要通过两个核函数，将原本线性不可分的点投影至高维空间，转化为线性可分，来区分密度、含糖率不同的好瓜以及坏瓜。

### 原理阐述

SVM学习的基本原理是求解能够正确划分训练数据集并且几何间隔最大的分离超平面。如下图所示，即为分离超平面，对于线性可分的数据集来说，这样的超平面有无穷多个（即感知机），但是几何间隔最大的分离超平面却是唯一的。

因此，求解最大超平面分割问题就可以转化为约束优化问题：

此时，用有约束的拉格朗日乘子法求解上式：

其中，为拉格朗日乘子，并且。因此，需要求出带入中就可以求出最大值。

而求最大值，通常采用求偏导的方式。即对求偏导为0，可分别得出下式：

带入到中，消去有：

在满足条件之后，用SMO算法求解凸二次规划问题：

太难了，这里看不懂，

### 算法设计思路

### 实验流程、测试结果及分析

### 代码结构，核心代码简要分析

### 本次实验解决的主要问题，主要收获

### 编码及内容撰写中的参考来源