

作业5：支持向量机

助教邮箱:wanghc17@mails.tsinghua.edu.cn

1 分析图片并回答问题

图1是对同一组数据使用不同核函数的支持向量机进行训练的结果。使用的核函数包括：A.线性核；B.二次多项式核；C.三次多项式核；D.径向基($\sigma = 1$);E.径向基($\sigma = 0.5$);F.径向基($\sigma = 0.1$);但它们的次序是打乱的，请分析散点图回答以下问题：

- (1) 请问以下各图分别是使用哪种核函数的结果？为什么
- (2) 对于这个数据集，选择哪种核函数比较合理？说明你的理由。

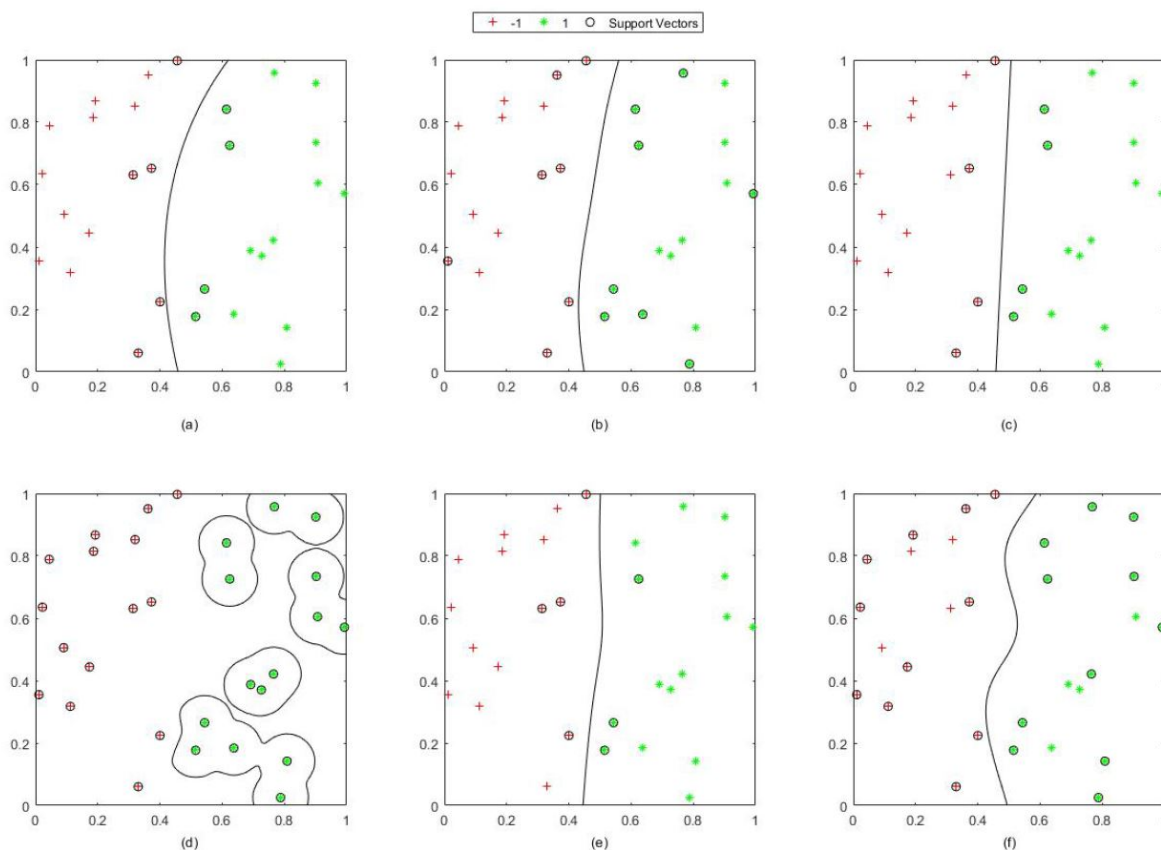


图 1: 第一题图

2 多项式核支持向量机简单分析

现有一拥有两个一维样本的数据集，两样本分别是： $(x_1 = 0, y_1 = -1)$ 以及 $(x_2 = \sqrt{2}, y_2 = 1)$ 。使用特征向量 $\phi = [1, \sqrt{x}, x^2]^T$ 将样本映射到三维空间(即使用二阶多项式核)，分类器的优化目标如下：

$$\begin{aligned} \min ||\omega^2|| \quad s.t. \\ y_1(\omega^T \phi(x_1) + \omega_0) \geq 1 \\ y_2(\omega^T \phi(x_2) + \omega_0) \geq 1 \end{aligned}$$

请回答下列问题：

- (1) 请写出一个与最优向量 ω 平行的向量。
- (2) 请计算在该 ω 情况下，SVM的margin是多少？（提示：margin是指支持向量到决策面之间的距离）
- (3) 请求出 ω （提示：margin的大小为 $1/||\omega||$ ）。
- (4) 请写出判别面方程 $f(x) = \omega_0 + \omega^T \phi(x)$ 关于x的显式表达式。（即求出 ω 以及 ω_0 ）

3 利用支持向量机对MNIST数据集进行分类

本任务将使用SVM对MNIST数据集中的两类手写体数字进行分类，请按照以下步骤完成该二分类任务：

(1) 预处理

请按照作业4中的手写体数据的要求做数据的预处理，并筛选出你在作业4中你所挑选出的混淆最严重的两个数字，用于分类。

(2) 训练

请使用sklearn.svm包中的SVC函数完成MNIST数据的二分类任务，请查阅官方文档，学习主要的几类核函数的参数设置方式（包括：线性核、多项式核、径向基核、sigmoid核等）；使用超过三种核函数并调整参数，完成对手写体数据的训练，并记录训练正确率。

（官方文档：<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>）

(3) 测试

请在测试集上测试训练模型的正确率，记录测试结果。

(4) 比较分析

请使用已经学习过的两个模型（神经网络，Logistic Regression）对相同数据集进行分类，比较分析包括支持向量机在内的三种模型的分类结果。请注意比较方法应该公平合理。

本题允许使用之前作业的代码和结论。