

作业5:支持向量机

1 分析图片并回答问题

图1是对同一组数据使用带有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？请说明你的理由。

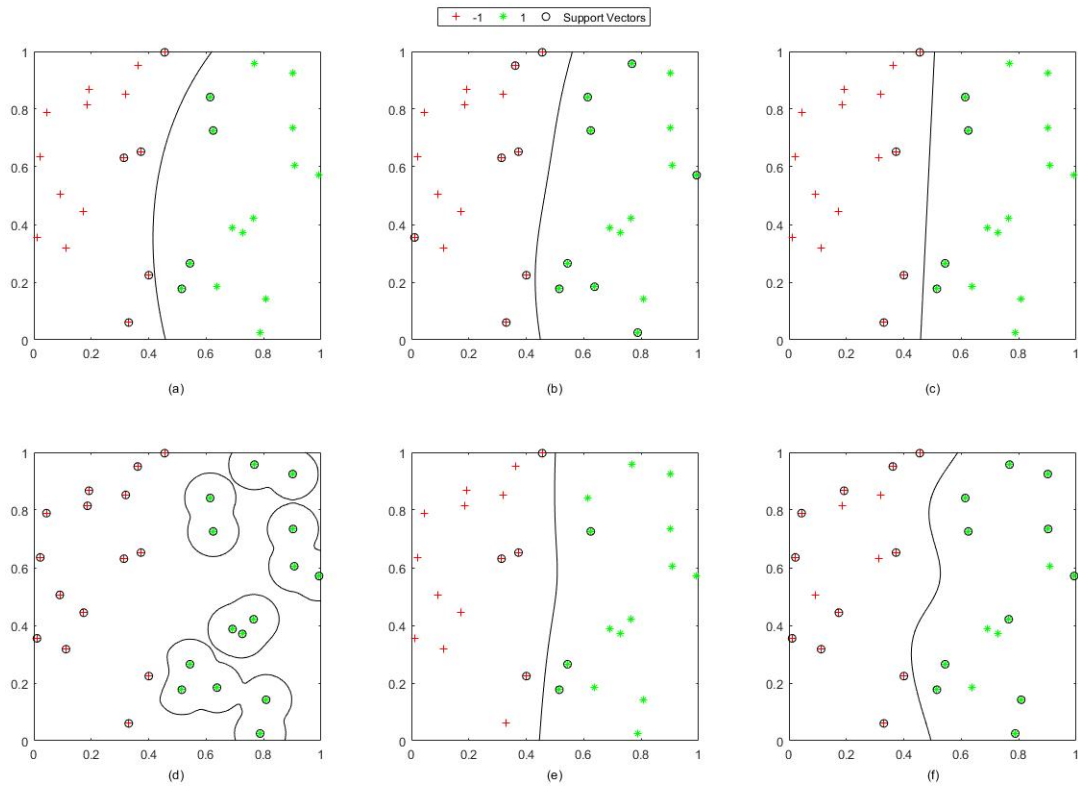


图 1: 第一题图

2 利用支持向量机对MNIST数据集进行分类

提示: LIBSVM是著名的支持向量机实现版本, 其中的功能非常强大。推荐感兴趣的同学使用它的Matlab版本进行下面过程的尝试, 下载网址<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 此项不做要求。

(1) 预处理

首先, 请按照作业4中的要求做数据的预处理, 筛选出作业4中所挑选的混淆最严重的两个数字, 分别设置标签为+1和-1。对于训练样本, 请将标签列向量合并到样本矩阵中, 使得第一列为标签向量, 其余列为特征向量。

(2) 训练

打开Matlab分类器工具箱(例如Matlab 2016a→应用程序→Classification Learner)。选择“New Session From Workspace”, 在“Step 1”中选择训练数据; 在“Step 2”中选择“predictor”和“response”, 在“Step 3”中选择“Holdout Validation”, 验证集合选择预留20%的样本。选择算法为“All SVM”, 点击“start”, 则会自动依次训练工具箱中的6个模型(包括: 线性核, 二次核, 三次核, 精细径向基, 中等径向基, 粗糙径向基), 使用老版本的同学请手动依次选择训练, 应用程序会自动根据验证集的正确率选择相关的参数。等待训练完成, 训练过程总共接近10分钟。请记录训练正确率。

(3) 测试

请对6个模型分别点击“Export Model”, 会保存一个名为“trainedClassifier”的结构体。

调用trainedClassifier.predictFcn(“你的测试矩阵”)会返回预测结果。请记录测试结果。

(4) 比较分析

请使用已经学习过的三个模型(神经网络, Logistic Regression和贝叶斯方法)对相同数据集合进行分类, 请比较分析包括支持向量机在内的四种模型的分类结果。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。