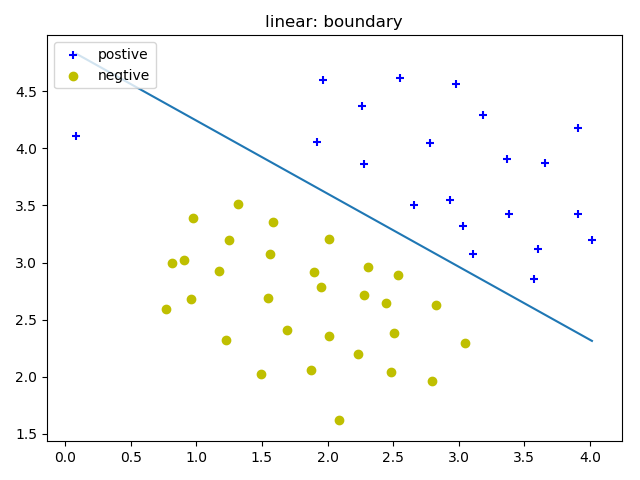
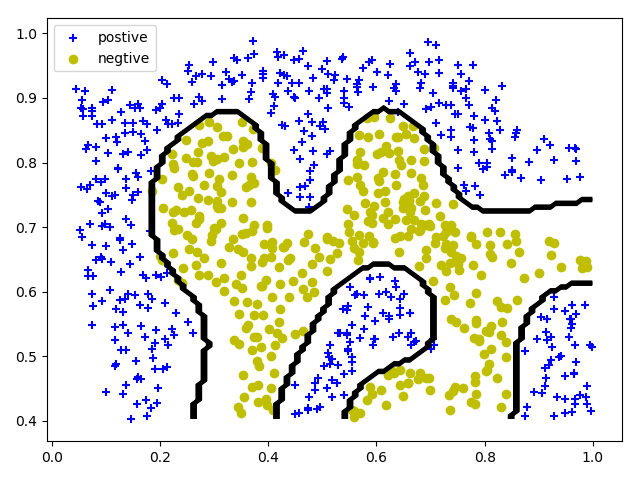
**支持向量机实验报告**

U20613382电信1606 畅嘉宇

1. **Task1** **用线性SVM和高斯核SVM预测对数据进行分类**
   1. 执行步骤
      1. Loaddata函数读入数据
      2. 由gaussianKernel函数计算高斯核矩阵
      3. 由svmTrain\_SMO函数分别训练线性和高斯svm，其中上一步算出的gaussianKernel作为高斯svm的参数
      4. 使用çLinear/Gaussian分别划出边界
   2. 超参数
      1. 线性核：C = 1
      2. 高斯核子：C = 1， sigma = 0.1
   3. 实验结果
      1. 决策边界
         1. 线性核

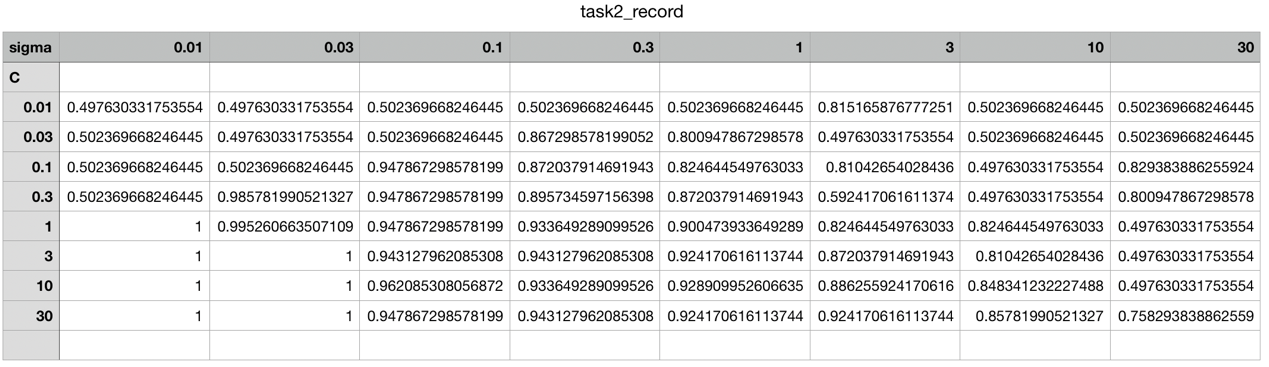


* + - 1. 高斯核



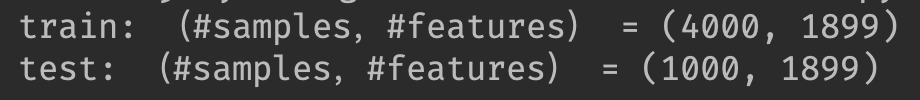
* 1. 结果分析
     1. 从决策边界可看出，基于两种核函数的svm较好的完成了分类任务

1. **Task2 使用高斯核SVM对给定数据集进行分类**
   1. 实验设置
      1. 根据要求将训练参数C和高斯核参数sigma设为要求值，遍历所有可能并记录正确率
   2. 编程思路
      1. 使用loaddata读取数据
      2. 初始化一个pandas DataFrame，行标签为C，列标签为sigma
      3. 设置两层循环，遍历所有超参数取值
      4. 由gaussianKernel函数计算高斯核矩阵
      5. 由svmTrain\_SMO函数训练高斯svm，其中上一步算出的gaussianKernel作为高斯svm的参数
      6. SvmPredict得到预测值y\_est
      7. 与实际y对比并计算正确率
      8. 填入DataFrame中
      9. 结束循环后保存DataFrame
   3. 实验结果



* 1. 思考与分析
     1. 随着c的增大，正确率总体在升高
     2. C取不同值时，最优的sigma值不同
     3. C较大时，sigma较小则较优

1. **Task3 使用线性SVM实现对垃圾邮件分类**
   1. 实验结果
      1. 样本数和特征维度



* 1. 实验流程
     1. 读入数据
     2. 将训练集按8:2划分为两部分，分别用作训练和测试
     3. 取c=30，训练线性svm
     4. 在测试集计算acc
     5. 在整个测试集上重新训练
     6. 预测测试集的标签，保存结果
  2. 结果
     1. Validation\_set上正确率为0.94
     2. 模型较好的完成了分类任务