编程作业说明：SVM

**任务一：分别用线性SVM和高斯核SVM预测对数据进行分类**

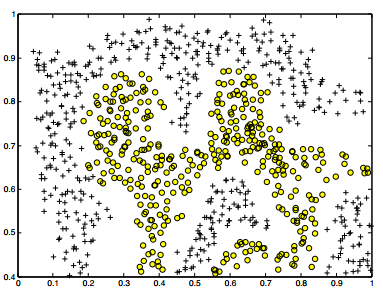
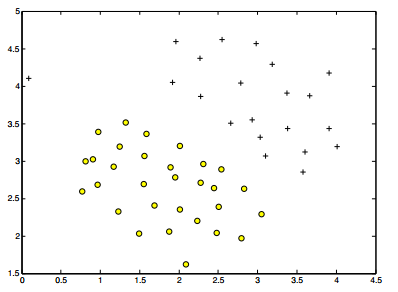
**（1）问题**：task1\_linear.mat中有一批数据点，试用线性SVM对他们进行分类，并在图中画分出决策边界。task1\_gaussian中也有一批数据点，试用高斯核SVM对他们进行分类，并在图中画出决策边界。

**（2）步骤提示：**

1.实验所需函数的代码在SVM\_Functions.py中已经给出，请尝试读懂代码，理解算法思想与步骤，并自行读入数据，执行所需函数，观察实验结果。（也可以自己编写实验代码）

2.附件PPT中给出了SMO训练算法的详细分析。

3.task1\_linear.mat和task1\_gaussian.mat图像如下所示：

 4.训练过程：

（a）加载与可视化数据loadData()，plotData()

（b）训练模型

model = SVM\_Functions.svmTrain\_SMO(X, y, C=1, max\_iter=20)

model=SVM\_Functions.svmTrain\_SMO(X,y,C=1,kernelFunction='gaussian',K\_matrix=s.gaussianKernel(X,sigma=0.1))

（c）决策边界可视化 SVM\_Functions.visualizeBoundaryLinear(X, y, model)

**（3）提交要求：**

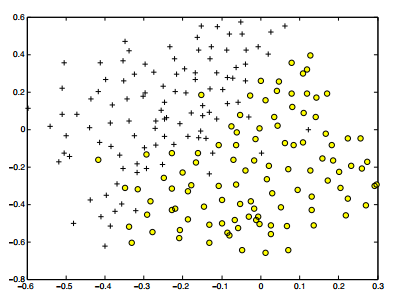
任务一需要编写实验报告进行简述其原理，代码步骤，并根据训练结果在图中画出决策边界。

**任务二：使用高斯核SVM对给定数据集进行分类**

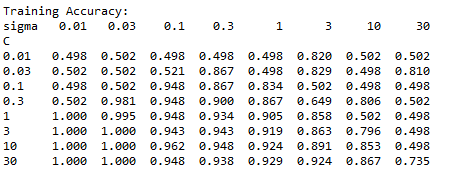
**（1）数据集讲解**：给定数据集（文件task2.mat）, 参考task1的代码, 编程实现一个高斯核SVM进行分类。输出训练参数C, sigma分别取0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30时(共64组参数组合)的训练集上的准确率。（程序运行时间8mins左右，准确率 = 预测正确样本数/样本总数 ）

**（2）提示：**

1. 数据集可视化：

****

2. 程序运行结果举例（由于SMO算法的随机性，你的结果应该跟下面的例子不完全相同）：



**（3）提交要求：**

编写实验报告，报告中需要包含：实验设置，编程思路，实验结果展示以及自己的思考等。其中实验结果展示按照上图的格式在实验报告中给出。

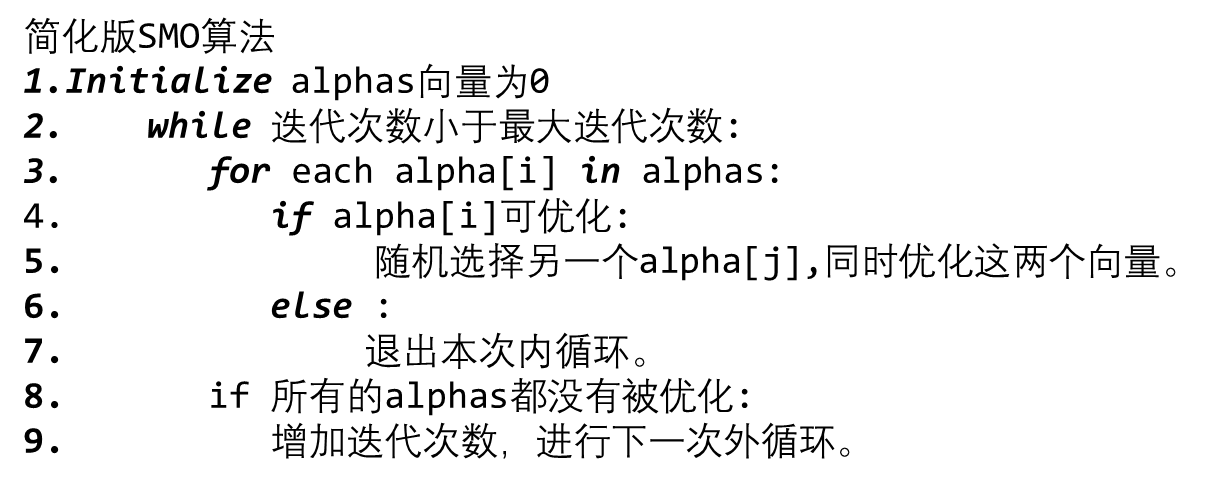
**任务三：使用线性SVM实现对垃圾邮件分类**

**（1）数据集讲解**：

编程实现一个垃圾邮件SVM线性分类器，分别在训练集和测试集上计算准确率。其中训练数据文件：task3\_train.mat，要求导入数据时输出样本数和特征维度。测试数据文件：task3\_test.mat，要求导入数据时输出样本数和特征维度，测试数据标签未给出。（程序运行时间10mins左右）

**（2）步骤提示：**

1. 对SMO算法的实现进行举例说明:

****

**（3）提交要求**：

将测试数据预测结果按顺序存储为txt文件，每一行为一个样本的标签。实验报告中需要写明具体实验流程，思路。