

SVM实现

----陈汝丹

1.数据来源

Andrew Ng在coursera上的机器学习公开课第七课提供的二维数据集，加号代表正样本，圆圈代表负样本。两个数据文件分别为ex6data1.mat， ex6data2.mat

训练数据集读取，由load函数读取两个mat文件，读出来的数据如右图：

```
X =
```

1.964300	4.595700
2.275300	3.858900
2.978100	4.565100
2.932000	3.551900
3.577200	2.856000
4.015000	3.193700
3.381400	3.429100
3.911300	4.176100
2.782200	4.043100
2.551800	4.616200
3.369800	3.910100
3.104800	3.070900
1.918200	4.053400
2.263800	4.370600
2.655500	3.500800
3.185500	4.288800
3.657900	3.869200
3.911300	3.429100
3.600200	3.122100
3.035700	3.316500
1.584100	3.357500

[illegible]

1.数据来源

dataset1截图如下：

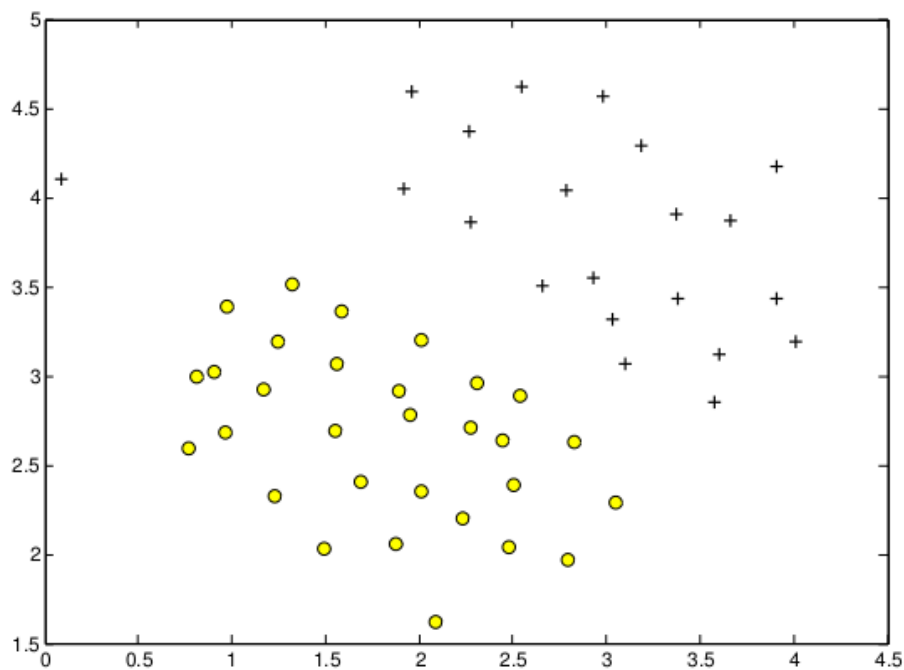


Figure 1: Example Dataset 1

Dataset2截图如下：

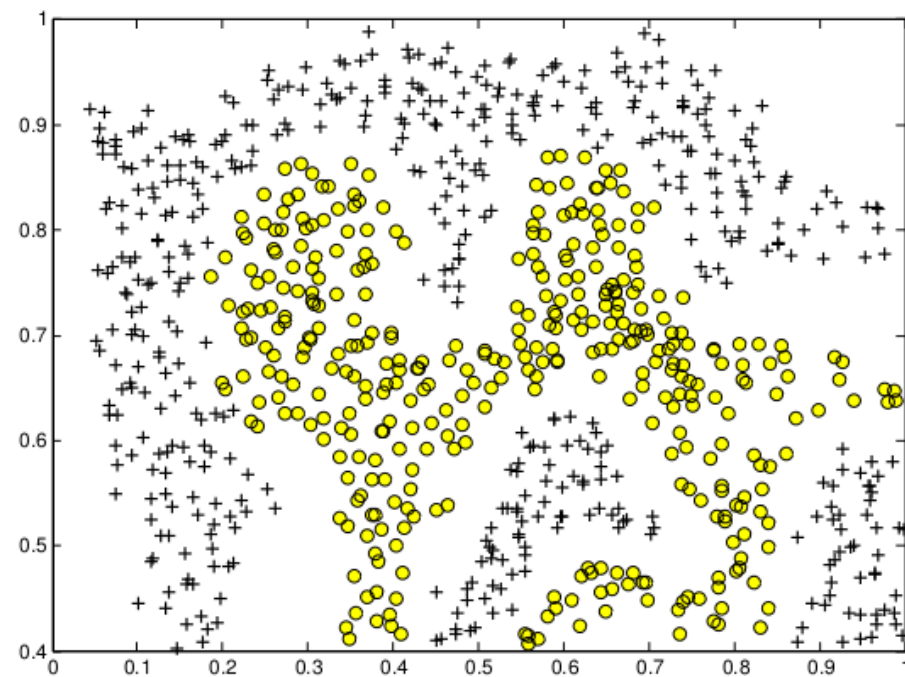


Figure 4: Example Dataset 2

2.SVM实现

SVM是为了在训练数据间找到一个划分不同类的超平面：

$$W^T * x + b = 0$$

通过对目标函数的定义及转化，得到了最终的目标函数和约束条件。

$$\min \frac{1}{2} ||w||^2 \quad s.t., y_i(W^T * x + b) \geq 1, i = 1, \dots, n$$

从而得到了一个凸二次规划问题，在本次作业中采用了SMO优化算法来解决这个二次规划问题。

a.核函数线性

当数据如figure.1所示，明显发现两个类之间的超平面是一条直线，因而在求解时， x 保持在原空间即可。

2.SVM实现

a.核函数非线性

而figure2的数据集可以看出，两个类之间的分界线不是线性的，就无法直接求解，因而需要使用核函数将x变换到一个特征空间中去，这个空间是高维空间，映射后的数据时就变成线性可分了。

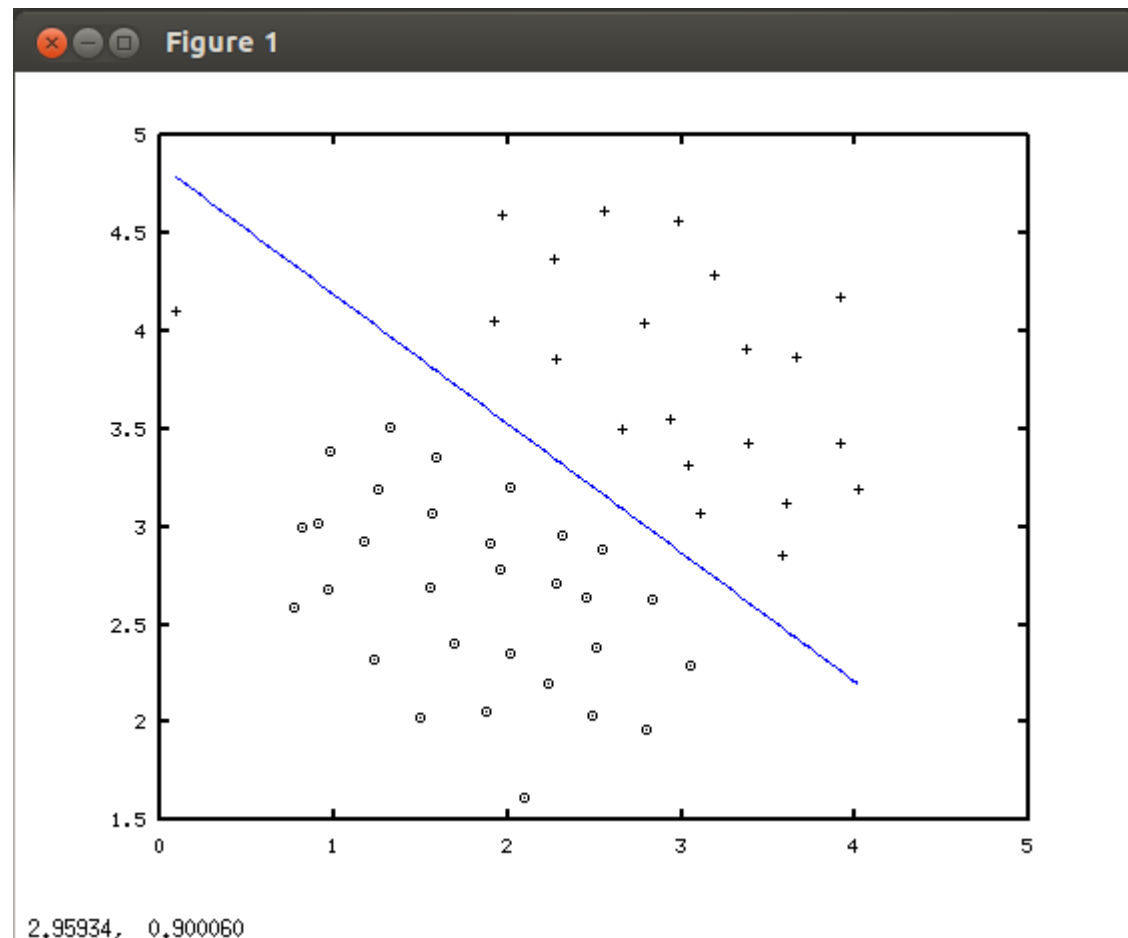
在本次作业中，核函数采用了Gauss Kernel。

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)$$

3.执行过程及实验结果

a.核函数为线性，调用
linearKernel

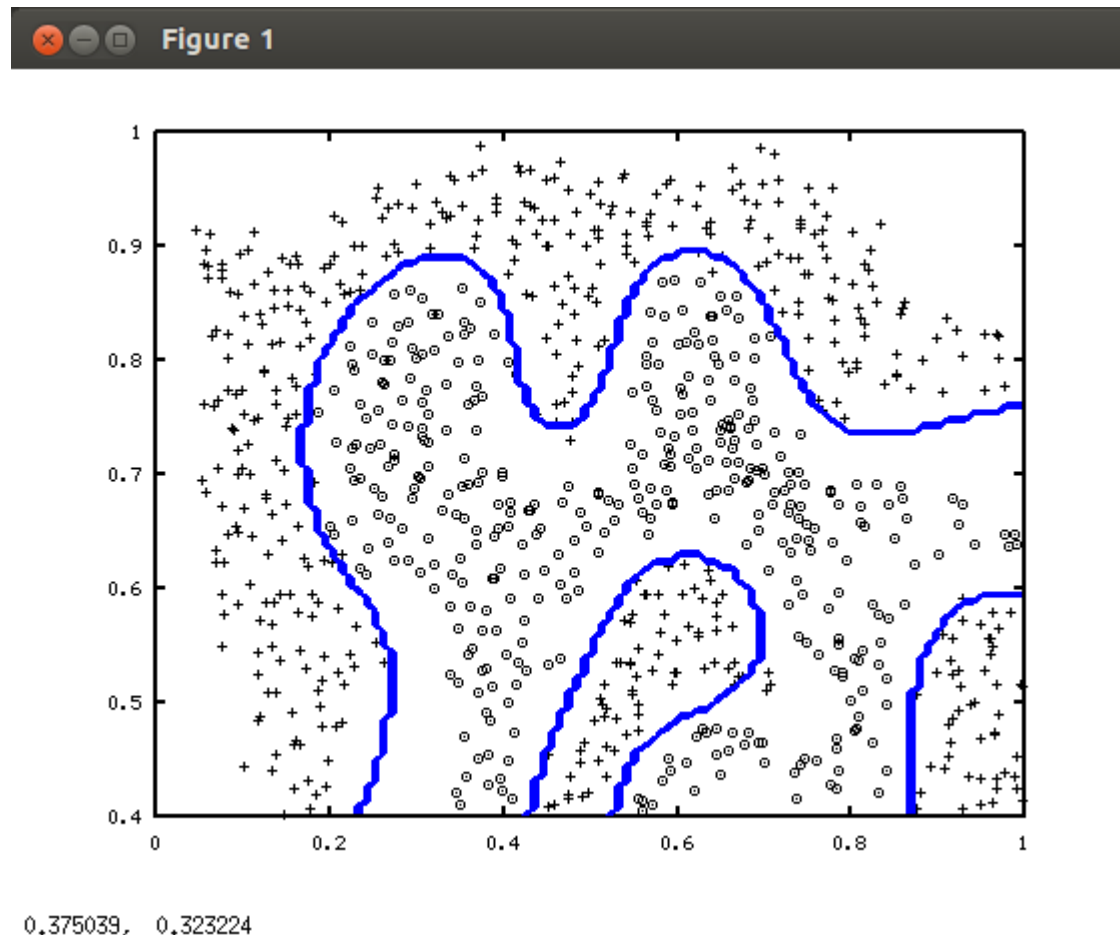
执行ex6，第一部分就是直接显示数据集一没有分类的原始数据分布图，然后进行svm分类后，会出现右图的实验结果



3. 执行过程及实验结果

b. 核函数为高斯函数，调用
gaussianKernel

ex6后面的部分先显示数据集2原始分布图，然后执行svm分类，出现右图结果



4.小结

支持向量机是为了实现线性分类器，从而提出核函数将低维线性不可分的数据映射到高维，映射后便可以进行线性分类。因而选择合适的核函数就比较重要。