

作业四：基于 SVM 的垃圾邮件分类

问题描述：

数据集：spamTest.mat 和 spamTrain.mat 分别代表垃圾邮件分类的测试集和训练集，训练集包含了 4000 个垃圾邮件和非垃圾邮件训练示例，测试集包含 1000 个测试示例，数据集中的 4 个 txt 文件则表示垃圾邮件和非垃圾邮件示例：

```
> Anyone knows how much it costs to host a web portal ?  
>  
Well, it depends on how many visitors you're expecting.  
This can be anywhere from less than 10 bucks a month to a couple of $100.  
You should checkout http://www.rackspace.com/ or perhaps Amazon EC2  
if youre running something big..
```

```
To unsubscribe yourself from this mailing list, send an email to:  
groupname-unsubscribe@egroups.com
```

图 1:非垃圾邮件示例

（邮件的预处理过程可以参考网上的教程，需要先将邮件内容转换为向量形式）
试构建 SVM 实现对垃圾邮件的分类。

要求：

1. 构造 SVM 用于垃圾邮件分类
2. 展示测试集上的准确率
3. 自己写一封垃圾邮件，输入分类器查看能否正确识别
4. 完成实验报告