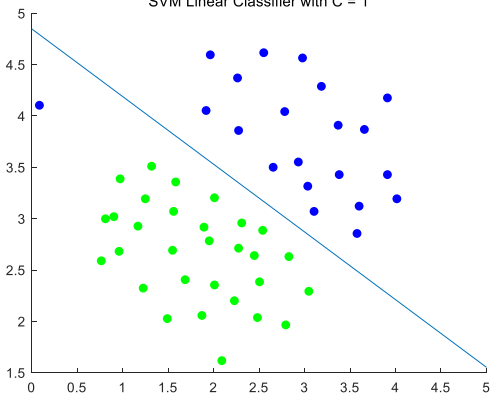
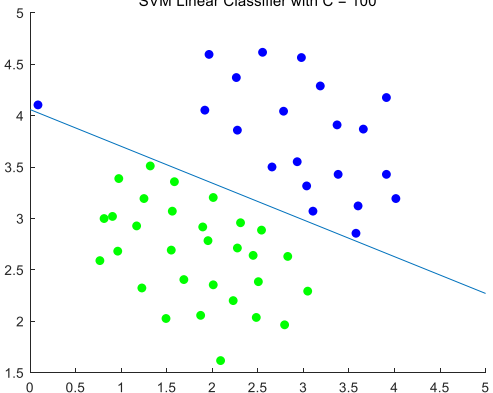


山东大学计算机科学与技术学院

机器学习与模式识别课程实验报告

学号：201600181073	姓名：唐超	班级：智能 16
实验题目：SVM Linear Classification		
实验学时：2	实验日期：2018. 11. 23	
实验目的： 用支持向量机的方法训练线性分类模型，并对测试数据进行分类。		
硬件环境： DELL 台式机		
软件环境： MATLAB R2018b		
实验步骤与内容： 1. 下载并编译 SVM 的一个软件包 LIBSVM，后续实验过程需要调用 LIBSVM 内的函数； 2. 下载解压实验数据集，对于 twofeature.txt, 用 libsvmread 函数并不能成功读入，需要手动更改该文件数据的格式，即将每个 feature 前的标记“1”或“2”去掉，最后直接用 load 函数以 table 格式读取； 3. 将读取的 table 的第一列作为 trainlabels, 第二、三列作为 trainfeatures, 用 svmtrain 函数进行训练，取惩罚项系数 C 分别为 1, 100，并进一步得到 w, b, 最后得到结果如下：		
<div><div><p>SVM Linear Classifier with C = 1</p></div><div><p>SVM Linear Classifier with C = 100</p></div></div>		
<p>可以看出，当 C 取值较小时，分类边界较为合理，但对于异常点容易分错，当 C 的取值较大时，异常点被分进正确的类别，但边界与两类数据的间隔较小。</p> <p>4. 用 libsvmread 函数分别读取文件“email_train-50.txt”，“email_train-100.txt”，“email_train-400.txt”，“email_train-all.txt”，分别用 svmtrain 函数训练（取 C 为默认值）得到 model，并在“email_test.txt”上用 svmpredict 函数测试，准确率依次为：0. 7538, 0. 8846, 0. 9808, 0. 9846.</p>		

结论分析与体会：

通过这次实验，使我对 SVM 的有了进一步的认识。

附录：程序源代码

```
%twofeature
twofeature = load('twofeature.txt');
x = [twofeature(:,2), twofeature(:,3)];
y = twofeature(:,1);
scatter(x(1:21,1), x(1:21,2), 40, 'b', 'filled');
hold on
scatter(x(22:51,1), x(22:51,2), 40, 'g', 'filled');

model = svmtrain(y, x, '-s 0 -t 0 -c 100');
w = model.SVs' * model.sv_coef;
b = -model.rho;
if (model.Label(1) == -1)
    w = -w;
    b = -b;
end

x1 = (0:0.1:5)' ;
x2 = zeros(51,1);
for i = 1:51
    x2(i) = (-b-w(1)*x1(i))/w(2);
end
plot(x1,x2)
title('SVM Linear Classifier with C = 100');

%email classification
[trainlabels, trainfeatures] = libsvmread('email_train-all.txt');
model = svmtrain(trainlabels, trainfeatures, '-s 0 -t 0 ');

w = model.SVs' * model.sv_coef;
b = -model.rho;
if (model.Label(1) == -1)
    w = -w;
    b = -b;
end
```

```
%test  
[testlabels, testfeatures] = libsvmread('email_test.txt');  
predicted_labels = svmpredict(testlabels, testfeatures, model);
```