

# 程序报告

2013011570 自 31 唐静娴

## 一、程序说明

- 程序来源：自行编写
- 包含文件：

文件名称	备注
dataset1.txt	给定的数据
mydataset.txt	采集的数据
main.m	主函数
preprocess.m	数据预处理函数，分离出男性、女性两组数据
gender_classify.m	贝叶斯决策分类器，并计算分类错误率
roc.m	ROC 曲线参数计算函数

- 各函数详细说明：

**function [male,female] = preprocess(file\_name)**

	参数名称	参数说明	参数类型
输入	file_name	即将进行预处理的文件名称	string
输出	male	分离出的男性身高、体重数据	M*2 double
	female	分离出的女性身高、体重数据	N*2 double

**function [err\_m,err\_f] = gender\_classify(thre,train\_m,train\_f,test\_m,test\_f,minRisk)**

	参数名称	参数说明	参数类型
输入	thre	阈值，绘制 ROC 曲线时使用	double
	train_m	训练集男性数据	M1*2 double
	train_f	训练集女性数据	N1*2 double
	test_m	测试集男性数据	M2*2 double
	test_f	测试集女性数据	N2*2 double
	minRisk	值为 true 时采用最小风险贝叶斯决策；值为 false 时采用最小错误率贝叶斯决策	bool
输出	err_m	男性分类错误率	double
	err_f	女性分类错误率	double

**function [Sp,Sn] = roc(train\_m,train\_f,test\_m,test\_f,minRisk)**

	参数名称	参数说明	参数类型
输入	train_m	训练集男性数据	M1*2 double
	train_f	训练集女性数据	N1*2 double
	test_m	测试集男性数据	M2*2 double
	test_f	测试集女性数据	N2*2 double
	minRisk	判断采用最小错误率或最小风险贝叶斯决策	bool
输出	Sp	特异度	1*100 double
	Sn	灵敏度	1*100 double

### main.m

主函数，调用上述三个函数，输出各个情况下的分类错误率，并根据 Sp, Sn 数组绘制 ROC 曲线。

## 二、运行步骤

直接运行 main.m 即可完成题目要求的所有任务。

## 三、源代码

### 1. main.m

```
clear all;
% 读文件
[data_male,data_female] = preprocess('dataset1.txt');
[mydata_male,mydata_female] = preprocess('mydataset.txt');

% 最小错误率贝叶斯决策
disp('*****最小错误率贝叶斯决策结果*****');
[err_m,err_f] = gender_classify(1,mydata_male,mydata_female,data_male,data_female,false);
disp('  A.训练集为采集数据，测试集为给定数据');
disp(['  男性错误率：',num2str(err_m,3),'  女性错误率：',num2str(err_f,3)]);

[err_m,err_f] = gender_classify(1,mydata_male,mydata_female,mydata_male,mydata_female,false);
disp('  B.训练集为采集数据，测试集为采集数据');
disp(['  男性错误率：',num2str(err_m,3),'  女性错误率：',num2str(err_f,3)]);

[err_m,err_f] = gender_classify(1,data_male,data_female,mydata_male,mydata_female,false);
disp('  C.训练集为给定数据，测试集为采集数据');
disp(['  男性错误率：',num2str(err_m,3),'  女性错误率：',num2str(err_f,3)]);

[err_m,err_f] = gender_classify(1,data_male,data_female,mydata_male,mydata_female,false);
disp('  D.训练集为给定数据，测试集为给定数据');
disp(['  男性错误率：',num2str(err_m,3),'  女性错误率：',num2str(err_f,3)]);
disp('');
```

```

% 最小风险贝叶斯决策
disp('*****最小风险贝叶斯决策结果*****');
[err_m,err_f] = gender_classify(1,mydata_male,mydata_female,data_male,data_female,true);
disp('  A.训练集为采集数据,测试集为给定数据');
disp(['  男性错误率:',num2str(err_m,3),'  女性错误率:',num2str(err_f,3)]);

[err_m,err_f] = gender_classify(1,mydata_male,mydata_female,mydata_male,mydata_female,true);
disp('  B.训练集为采集数据,测试集为采集数据');
disp(['  男性错误率:',num2str(err_m,3),'  女性错误率:',num2str(err_f,3)]);

[err_m,err_f] = gender_classify(1,data_male,data_female,mydata_male,mydata_female,true);
disp('  C.训练集为给定数据,测试集为采集数据');
disp(['  男性错误率:',num2str(err_m,3),'  女性错误率:',num2str(err_f,3)]);

[err_m,err_f] = gender_classify(1,data_male,data_female,mydata_male,mydata_female,true);
disp('  D.训练集为给定数据,测试集为给定数据');
disp(['  男性错误率:',num2str(err_m,3),'  女性错误率:',num2str(err_f,3)]);

% ROC曲线
[Sp1,Sn1] = roc(mydata_male,mydata_female,data_male,data_female,false);
[Sp2,Sn2] = roc(mydata_male,mydata_female,data_male,data_female,true);
figure,plot(1-Sp1,Sn1,'bx-'),title('ROC曲线(训练集为采集数据,测试集为给定数据)');
hold on,plot(1-Sp2,Sn2,'r*-');
legend('最小错误率','最小风险');

[Sp1,Sn1] = roc(data_male,data_female,mydata_male,mydata_female,false);
[Sp2,Sn2] = roc(data_male,data_female,mydata_male,mydata_female,true);
figure,plot(1-Sp1,Sn1,'bx-'),title('ROC曲线(训练集为给定数据,测试集为采集数据)');
hold on,plot(1-Sp2,Sn2,'r*-');
legend('最小错误率','最小风险');

```

## 2. preprocess.m

```

function [male,female] = preprocess(file_name)
% 对数据进行预处理,将男性和女性数据分别导出
file = fopen(file_name);
predataset = textscan(file,'%f %f %c');
fclose(file);
dataset = [predataset{1},predataset{2}];
fseq = find(predataset{3)=='F' | predataset{3)=='f';
mseq = find(predataset{3)=='M' | predataset{3)=='m';
female(:,1) = dataset(fseq,1);
female(:,2) = dataset(fseq,2);
male(:,1) = dataset(mseq,1);
male(:,2) = dataset(mseq,2);
end

```

## 3. gender\_classify.m

```

function [err_m,err_f] = gender_classify(thre,train_m,train_f,test_m,test_f,minRisk)
%% 男性训练集
ave_m = mean(train_m);
cov_m = cov(train_m);
W_m = -0.5 * inv(cov_m);
w_m = inv(cov_m) * ave_m';
w0_m = -0.5*ave_m'*inv(cov_m)*ave_m' - 0.5*log(det(cov_m)) + log(0.5);

%% 女性训练集
ave_f = mean(train_f);
cov_f = cov(train_f);
W_f = -0.5 * inv(cov_f);
w_f = inv(cov_f) * ave_f';
w0_f = -0.5*ave_f'*inv(cov_f)*ave_f' - 0.5*log(det(cov_f)) + log(0.5);

```

```

%% 根据 最小错误/最小风险 选择不同的lamda矩阵
if(minRisk == true)
    lamda = [0,5;4,0];
else
    lamda = [0,1;1,0];
end

```

```

%% 男性性别决策错误率
err_m = 0;
for i = 1:size(test_m,1)
    g_m = test_m(i,:)*W_m*test_m(i,:)' + test_m(i,:)*w_m + w0_m;
    g_f = test_m(i,:)*W_f*test_m(i,:)' + test_m(i,:)*w_f + w0_f;
    if(g_m*lamda(2,1) > thre*g_f*lamda(1,2));
    else
        err_m = err_m + 1;
    end
end
err_m = err_m/size(test_m,1);

```

```

%% 女性性别决策错误率
err_f = 0;
for i = 1:size(test_f,1)
    g_m = test_f(i,:)*W_m*test_f(i,:)' + test_f(i,:)*w_m + w0_m;
    g_f = test_f(i,:)*W_f*test_f(i,:)' + test_f(i,:)*w_f + w0_f;
    if(g_m*lamda(2,1) > thre*g_f*lamda(1,2))
        err_f = err_f + 1;
    else
    end
end
err_f = err_f/size(test_f,1);
end

```

#### 4. roc.m

```

function [Sp,Sn] = roc(train_m,train_f,test_m,test_f,minRisk)
i = 1;
for thre = 0.1:0.1:10
    [err_m,err_f] = gender_classify(thre,train_m,train_f,test_m,test_f,minRisk);
    % 设结果为男性时是阳性
    Sn(i)=1-err_m;
    Sp(i)=1-err_f;
    i = i + 1;
end
end

```