Computational Social Intelligence

Coursework 1

Name: Yuge Wang

Student ID: xxxxxxxx

Contents

Part 1

Introduction

In this part of the coursework, it is investigated that whether the gender or the role during phone conversations affect the laughter events of both sides. In order to collect data, 60 phone calls are conducted between 120 unacquainted speakers, out of which, 57 are male and 63 are females; and 60 are callers, 60 are receivers. The data of every laughter event with duration in seconds, gender of the person that laughs and role of the person is collected and recorded. The software used in this coursework is MATLAB.

The following questions are answered using statistical analysis:

1. Is the number of laughter events higher for women than for men?
2. Is the number of laughter events higher for callers than for receivers?
3. Are laughter events longer for women?
4. Are laughter events longer for callers?

Results

Question 1

Research hypothesis:
The number of laughter events tend to be higher for women than for men.

$$H_{a1}:\ \mu_{FN} > \mu_{MN}$$

Null hypothesis:
The number of laughter events do not tend to be higher for women than for men.

$$H_{01}: \mu_{FN} = \mu_{MN}$$

From the hypothesis built, it is indicated that one-tailed test should be conducted. However, because that the variance of the sample data can not be obtained, the test can only be conducted with the mean and total number of laughter events for women and for men. In this case, the Chi square method is used to test the null hypothesis.
The basic parameters obtained from the data are as follow:

Symbol	Magnitude	Description
F	63	Female number
M	57	Male number
FN	496	Female laughter event numbers
MN	346	Male laughter event numbers
N	842	Total laughter event number
$\bar{h}_N$	7.0167	Mean of laughter event number
$\bar{h}_{FN}$	7.8730	Mean of laughter event number of female
$\bar{h}_{MN}$	6.0702	Mean of laughter event number of male
O_{FN}	496	Observation of female laughter event number
O_{MN}	346	Observation of male laughter event number
E_{FN}	442.05	Expectation of female laughter event number
E_{MN}	399.95	Expectation of male laughter event number
k_1	1	Degree of freedom

$$E_{FN} = \bar{h}_N * F = 442.05$$

$$E_{FN} = \bar{h}_N * M = 399.95$$

$$\chi_1^2 = \frac{(O_{FN} - E_{FN})^2}{E_{FN}} + \frac{(O_{MN} - E_{MN})^2}{E_{MN}} = 13.8617$$

$$k_1 = 2 - 1 = 1$$

$$p(\chi_1^2\ ) = \frac{1}{2^{\frac{k_1}{2}}\Gamma\left(\frac{k_1}{2}\right)}(\chi_1^2)^{\frac{k_1}{2}-1}e^{-\frac{1}{2}\chi_1^2} = 1.047 \times 10^{-4}$$

Because that the confidence level is very small, the null hypothesis can be rejected. However, because that the Chi square test is two tailed, it can only be proved that there is difference between the means of male and female laughter events. Because that the sample mean of female laughter events is higher than that is for male, it can be proved that the number of laughter events tend to be higher for women than for men.

Question 2

Research hypothesis:

The number of laughter events tend to be higher for callers than for receivers.

$$H_{a2}: \mu_{CN} > \mu_{RN}$$

Null hypothesis:

The number of laughter events do not tend to be higher for callers than for receivers.

$$H_{02}: \mu_{CN} = \mu_{RN}$$

From the hypothesis built, it is indicated that one-tailed test should be conducted. Because that the variance of the sample data can not be obtained, the test can only be conducted with the mean and total number of laughter events for women and for men. In this case, the Chi square method is used to test the null hypothesis.

The basic parameters obtained from the data are as follow:

Symbol	**Magnitude**	**Description**
C	60	Caller number
R	60	Receiver number
CN	505	Caller laughter event numbers
RN	337	Receiver laughter event numbers
N	842	Total laughter event number
$\bar{h}_N$	7.0167	Mean of laughter event number
$\bar{h}_{CN}$	8.4167	Mean of laughter event number of caller
$\bar{h}_{RN}$	5.6167	Mean of laughter event number of receiver
O_{CN}	505	Observation of caller laughter event number
O_{RN}	337	Observation of receiver laughter event number
E_{CN}	421	Expectation of caller laughter event number
E_{RN}	421	Expectation of receiver laughter event number
k_2	1	Degree of freedom

$$E_{FN} = \bar{h}_N * C = 421$$

$$E_{FN} = \bar{h}_N * R = 421$$

$$\chi_2^2 = \frac{(O_{CN} - E_{CN})^2}{E_{CN}} + \frac{(O_{RN} - E_{RN})^2}{E_{RN}} = 33.5202$$

$$k = 2 - 1 = 1$$

$$p(\chi_2^2) = \frac{1}{2^{\frac{k_2}{2}}\Gamma\left(\frac{k_2}{2}\right)}(\chi_2^2)^{\frac{k_2}{2}-1}e^{-\frac{1}{2}\chi_2^2} = 3.6261 \times 10^{-9}$$

Because that the confidence level is very small, the null hypothesis can be rejected. However, because that the Chi square test is two tailed, it can only be proved that there is difference between the means of caller and receiver laughter events. Because that the sample mean of callers laughter events is higher than that is for receivers, it can be proved that the number of laughter events tend to be higher for callers than for receivers.

Question 3

Research hypothesis:

Laughter events tend to be longer for women than for men.

$$H_{a3}:\ \mu_{FD} > \mu_{MD}$$

Null hypothesis:

Laughter events do not tend to be longer for women than for men.

$$H_{03}: \mu_{FD} = \mu_{MD}$$

From the hypothesis built, it is indicated that one-tailed test should be conducted. Because that the population mean is unknown, the student's t test can be conducted in this case to test the null hypothesis.

The basic parameters obtained from the data are as follow:

Symbol	Magnitude	Description
FN	496	Female laughter event numbers
MN	346	Male laughter event numbers
N	842	Total laughter event number
$\bar{h}_{FD}$	0.7097	Mean of laughter event duration of female
$\bar{h}_{MD}$	0.6062	Mean of laughter event duration of male
s_{FD}	0.4412	Standard deviation of female duration
s_{MD}	0.4041	Standard deviation of male duration
k_3	840	Degree of freedom

$$t_1 = \frac{\bar{h}_{FD} - \bar{h}_{MD}}{\sqrt{\frac{S_{FD}^2}{FN} + \frac{S_{MD}^2}{MN}}} = 3.519$$

$$k_3 = FN - 1 + MN - 1 = 840$$

From the student's t table (see appendix 1), it can be seen that the critical value for confidence level 0.05 and one-tailed should be within 1.646 to 1.660. which is obviously less than t_1. Thus, the null hypothesis can be rejected. It can be proved that the number of laughter events tend to be longer for female than for male.

Question 4

Research hypothesis:

Laughter events tend to be longer for caller than for receiver

$$H_{a4}: \mu_{CD} > \mu_{RD}$$

Null hypothesis:

Laughter events do not tend to be longer for caller than for receiver.

$$H_{04}: \mu_{CD} = \mu_{RD}$$

From the hypothesis built, it is indicated that one-tailed test should be conducted. Because that the population mean is unknown, the student's t test can be conducted in this case to test the null hypothesis.

The basic parameters obtained from the data are as follow:

Symbol	Magnitude	Description
FN	496	Female laughter event numbers
MN	346	Male laughter event numbers
N	842	Total laughter event number
$\bar{h}_{FD}$	0.7097	Mean of laughter event duration of female
$\bar{h}_{MD}$	0.6062	Mean of laughter event duration of male
s_{FD}	0.4412	Standard deviation of female duration
s_{MD}	0.4041	Standard deviation of male duration
k_3	840	Degree of freedom

$$t_1 = \frac{\bar{h}_{FD} - \bar{h}_{MD}}{\sqrt{\frac{S_{FD}^2}{FN} + \frac{S_{MD}^2}{MN}}} = 3.519$$

$$k_3 = FN - 1 + MN - 1 = 840$$

From the student's t table(see appendix 1), it can be seen that the critical value for confidence level 0.05 and one-tailed should be within 1.646 to 1.660. which is obviously less than t_1. Thus, the null hypothesis can be rejected. It can be proved that the number of laughter events tend to be longer for female than for male.

Part 2

Introduction

In this part of the coursework, a classifier is built to classify whether the facial expression detected is frown or smile, with 17 action units forming the feature vector. In this research, 52 face images were investigated and the feature vectors extracted from the pictures were split into training and test set. The training set was used to develop the classifier and the test set was used to test the behavior of the classifier.

Methodology

In this research, the algorithm of the classifier is Bayes' Theorem:

$$p(C_i|\vec{x}) = \frac{p(\vec{x}|C_i)p(C_i)}{p(\vec{x})}$$

In which, $p(C_i|\vec{x})$ represents the posterior of class i, $p(\vec{x}|C_i)$ represents the likelihood of class i with respect to the feature vector, $p(C_i)$ represents the prior of class i, and $p(\vec{x})$ represents the evidence probability.

The class of feature is determined by the highest posterior:

$$C^* = \arg\max p(C_k|\vec{x}) \ = \arg\max\ p(\vec{x}|C_k)p(C_k)\ ,\ C_k \in C$$

Because that the data of feature in this case is continuous and can be assumed to act normally of each action unit, the Gaussian Discriminant Function is used to model the likelihood of the classifier:

$$p(x_i|C_k) = \frac{1}{\sqrt{2\pi}\sigma_{ik}} e^{-\frac{(x_i-\mu_{ik})^2}{2\sigma_{ik}^2}}$$

Because that the probability of the feature vector is the product of the probability of every unit in feature vector, when the components can be assumed to be independent to each other, it can be expressed that:

$$p(\vec{x}|C_k) = \prod_{i=1}^{D} p(x_i|C_k)$$

For the purpose of easier calculation, the logarithm calculation can be taken in both sides. The above equations can be combined, and in this case, the natural logarithm can be taken, the combined equation can be shown as follow:

$$\ln p(\vec{x}|C_k) = \sum_{i=1}^{D} \left[\ln \frac{1}{\sqrt{2\pi}\sigma_{ik}} - \frac{(x_i-\mu_{ik})^2}{2\sigma_{ik}^2} \right]$$

The decision of classification through this classifier can be expressed as follow:

$$\hat{k} = \arg\max \ln p(\vec{x}|C_k) + \ln p(C_k)$$

The software used to conduct calculation and analysis in this research is MATLAB.

Results

In the develop of classifier, the first step is to train the classifier with training data. Because that in the determination equation, the parameter to be ensured is the standard deviation (std) and the mean of the feature vector, the training process can be conducted by calculating the matrices of σ and μ of each class using the training data.

Because that the file contains the classification at the last column, it should be removed in the extraction of feature vector. The probability of each class can be calculated using the data of the last column. It can be obtained that there are 36 sets of feature data, in which the class 0 and the class 1 have the same probability. There are 17 units in the feature, thus, the mean and std matrices of this case should be 1×17. In this research, the class 'frown' is defined as class 0, and the class 'smile' is defined as class 1. The training results can be shown below:

$$\mu_0 = [.504\ .251\ 2.172\ .576\ .404\ 1.290\ .728\ .728$$
$$.120\ .286\ .945\ .844\ .504\ .167\ .323\ .140\ .206]$$
$$\mu_1 = [.419\ \ .109\ \ .239\ \ .334\ \ 2.435\ \ 1.739\ \ .482\ \ 1.989\ \ 3.063$$
$$1.684\ \ .027\ \ .003\ \ .328\ \ .062\ 2.564\ \ .383\ \ .081]$$
$$\sigma_0 = [.666\ .597\ 1.015\ .761\ .552\ 1.183\ 1.063\ .719$$
$$.221\ .473\ .675\ .840\ .590\ .281\ .703\ .305\ .362]$$
$$\sigma_1 = [.493\ .323\ .535\ .427\ .958\ .789\ .605\ 1.008$$
$$.760\ .896\ .105\ .009\ .398\ .111\ .986\ .435\ .137]$$

In the test part, the test set is input into the equation of classification. It can be observed that there are 16 sets of feature vector in the test data, also, the data in the last column is gathered separately. The classification of the data is expressed in a vector c_test with size 16, and the classification through the classifier can be expressed in a vector c. The results are as follow:

c_test	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
c	0	0	0	0	1	0	0	0	1	1	1	1	1	1	0	1

It can be observed that in this test, the classifier gives the wrong answer two times, the error rate can be calculated as follow:

$$err = \frac{2}{16} = 0.125$$

This indicate that in this research, the classifier developed has a probability of 87.5% to give the right classification output.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

```
clc;
clear;
```

```
%read table
filename='laughter-corpus.csv';
data=readtable(filename);
gender=data.Gender;
role=data.Role;
duration=data.Duration;
```

```
F=63; %female number
M=57; %male number
C=60; %caller number
R=60; %receiver number

N=size(gender);
N=N(1); %total laughter event number
D=sum(duration); %total laughter event duration
hD=D/N; %mean of laughter duration in every event
hN=N/(F+M); %mean of laughter event number of one person

FN=0;%female laughter event numbers
MN=0;%male laughter event numbers
CN=0;%caller laughter event numbers
RN=0;%receiver laughter event numbers

FD=0;%female laughter duration
MD=0;%male laughter duration
CD=0;%caller laughter duration
RD=0;%receiver laughter duration

Dss=0;%duration variance
FDss=0;%female duration variance
MDss=0;%male duration variance
CDss=0;%caller duration variance
RDss=0;%receiver duration variance

%store duration data seperately
durationF=[];
durationM=[];
durationC=[];
durationR=[];
```

```
for i=1:N
    if strcmp(gender(i),'Female')==1
        FD=FD+duration(i);
        FN=FN+1;
        durationF=[durationF duration(i)];
```

```
    elseif strcmp(gender(i),'Male')==1
        MD=MD+duration(i);
        MN=MN+1;
        durationM=[durationM duration(i)];
    end
    Dss=Dss+(duration(i)-hD)^2/N;
    i=i+1;
end

for j=1:N
    if strcmp(role(j),'Caller')==1
        CD=CD+duration(j);
        CN=CN+1;
        durationC=[durationC duration(j)];
    elseif strcmp(role(j),'Receiver')==1
        RD=RD+duration(j);
        RN=RN+1;
        durationR=[durationR duration(j)];
    end
    j=j+1;
end

hFN=FN/F; %mean of laughter number of female
hMN=MN/M; %mean of laughter number of male
hCN=CN/C; %mean of laughter number of caller
hRN=RN/R; %mean of laughter number of receiver
hFD=FD/FN; %mean of laughter duration of female laughter events
hMD=MD/MN; %mean of laughter duration of male laughter events
hCD=CD/CN; %mean of laughter duration of caller laughter events
hRD=RD/RN; %mean of laughter duration of receiver laughter events
```

```
i=0;
j=0;

for i=1:N
    if strcmp(gender(i),'Female')==1
        FDss=FDss+(duration(i)-hFD)^2/FN;
    elseif strcmp(gender(i),'Male')==1
        MDss=MDss+(duration(i)-hMD)^2/MN;
    end
    i=i+1;
end

for j=1:N
    if strcmp(role(j),'Caller')==1
        CDss=CDss+(duration(j)-hCD)^2/CN;
    elseif strcmp(role(j),'Receiver')==1
        RDss=RDss+(duration(j)-hRD)^2/RN;
    end
```

```
    j=j+1;
end

Ds=sqrt(Dss); %duration standard deviation
FDs=sqrt(FDss); %female duration standard deviation
MDs=sqrt(MDss); %male duration standard deviation
CDs=sqrt(CDss); %caller duration standard deviation
RDs=sqrt(RDss); %receiver duration standard deviation
```

```
%observations
OFN=FN;
OMN=MN;
OCN=CN;
ORN=RN;
OFD=FD;
OMD=MD;
OCD=CD;
ORD=RD;

%expectations
EFN=hN*F;
EMN=hN*M;
ECN=hN*C;
ERN=hN*R;
EFD=hD*FN;
EMD=hD*MN;
ECD=hD*CN;
ERD=hD*RN;
```

```
%Chi square of Q1 2
CD1=(OFN-EFN)^2/EFN+(OMN-EMN)^2/EMN;%chi square of male and female laughter number
CD2=(OCN-ECN)^2/ECN+(ORN-ERN)^2/ERN;%chi square of caller and receiver laughter number
pCD1=(CD1^(-0.5)*exp(-0.5*CD1))/(2^0.5*gamma(0.5));%The probability density function of CD1
pCD2=(CD2^(-0.5)*exp(-0.5*CD2))/(2^0.5*gamma(0.5));%The probability density function of CD2
```

```
%student t of Q3 4
t1=(hFD-hMD)/sqrt(FDss/FN+MDss/MN);
t2=(hCD-hRD)/sqrt(CDss/CN+RDss/RN);
k34=N-2;
pt1=gamma((j+1)/2)/(sqrt(j*pi)*gamma(j/2))*(1+t1^2/j)^(-(j+1)/2);
pt2=gamma((j+1)/2)/(sqrt(j*pi)*gamma(j/2))*(1+t2^2/j)^(-(j+1)/2);
```

```
clc
clear
```

```
%read training part
filename1='training-part-2.csv';
data1=readtable(filename1);
```

```
s1=size(data1);
rows1=s1(1);%row number of training part
cols1=s1(2);%column number of training part
x_train=table2array(data1(:,1:end-1));% training data part
```

```
%calculate number of class frown(c0) and class smile(c1) in training data
nc0_train=0;
nc1_train=0;

for i=1:rows1
    if strcmp(data1.Class{i},'frown')==1
        nc0_train=nc0_train+1;
    elseif strcmp(data1.Class{i},'smile')==1
        nc1_train=nc1_train+1;
    end
    i=i+1;
end

%possibility of c0 and c1
pc0_train=nc0_train/rows1;
pc1_train=1-pc0_train;

%c0 data part and c1 data part
xc0_train=x_train(1:nc0_train,1:cols1-1);
xc1_train=x_train(nc0_train+1:end,1:cols1-1);
```

```
%calculate mean with training data
for k = 1:cols1-1
    mu0(k)=mean(xc0_train(:,k));%mean of x in c0 part
    mu1(k)=mean(xc1_train(:,k));%mean of x in c1 part
    k=k+1;
end
```

```
%calculate std with training data
sigma0=zeros(1,cols1-1);
sigma1=zeros(1,cols1-1);

%c0 part
for j = 1:nc0_train
```

```
    for i = 1:cols1-1
        sigma0(i)=sigma0(i)+(xc0_train(j,i)-mu0(i))^2;
        i=i+1;
    end
    j=j+1;
end

%c1 part
for j = 1:nc1_train
    for i = 1:cols1-1
        sigma1(i)=sigma1(i)+(xc1_train(j,i)-mu1(i))^2;
        i=i+1;
    end
    j=j+1;
end

sigma0=sqrt(sigma0/(nc0_train));%std c0
sigma1=sqrt(sigma1/(nc1_train));%std c1
```

```
%read test data
filename2='test-part-2.csv';
data2=readtable(filename2);

s2=size(data2);
rows2=s2(1);
cols2=s2(2);
x_test=table2array(data2(:,1:end-1));
```

```
%read classes of test part
Class_test=data2.Class;
c_test=[];

nc0_test=0;
nc1_test=0;

for i=1:size(Class_test)
    if strcmp(Class_test{i},'frown')==1
        c_test=[c_test 0];
        nc0_test=nc0_test+1;
    elseif strcmp(Class_test{i},'smile')==1
        c_test=[c_test 1];
        nc1_test=nc1_test+1;
    end
    i=i+1;
end
```

```
%calculate likelihood
pxc0=zeros(rows2,1);
```

```
pxc1=zeros(rows2,1);

for i = 1 : cols2-1
    for j = 1 : rows2
        pxc0(j)=pxc0(j)+log(1/(sqrt(2*pi)*sigma0(i)))-(x_test(j,i)-mu0(i))^2/(2*sigma0(i)^2);
        pxc1(j)=pxc1(j)+log(1/(sqrt(2*pi)*sigma1(i)))-(x_test(j,i)-mu1(i))^2/(2*sigma1(i)^2);
        j=j+1;
    end
    i=i+1;
end
```

```
%record classification
c=zeros(1,rows2);
for j = 1:rows2
    if pxc0(j)*pc0_train>pxc1(j)*pc1_train
        c(j)=0;
    else
        c(j)=1;
    end
end
```

```
%calculate error rate
err=0;
c_dif=c_test-c;
sc=size(c_dif);
sc=sc(2);
for i = 1:sc
    if c_dif(i)~=0
        err=err+1;
    end
    i=i+1;
end
err=err/sc;
```