

Homework 1

Submitted By-Debatri Mitra

ID-63128512

Problem 1

CODE:

```
clear all;
%old=cd('C:\Users\DEBATRI\Desktop\HW1\data'); % cd to the data
folder
DATA=load('data/iris.txt'); % read data from
text
%cd(old); % cd back to old
directory

x=DATA(:,1:end-1); % features
y=DATA(:,end); % classes

whos; % show current variables
in memory and sizes

no_of_features = size(x,2); % get no of features
i.e. equal to num of columns
no_of_data = size(x,1); % get no of data points
i.e. equal to the no of rows

%size(x)
MEAN =zeros(4,1);
VAR=zeros(4,1);

for i=1:no_of_features
    subplot(2,2,i);
    hist(x(:,i));
    MEAN(i,1)=mean(x(:,i));
    VAR(i,1)=var(x(:,i));
end
STD=VAR.^(1/2);
%NORMAL=zeros(size(x));
ONES=ones(no_of_data,1);

for i=1:no_of_features
    x(:,i)=(x(:,i)-ONES.*MEAN(i,1));
    x(:,i)=x(:,i)./STD(i,1);
end
%STD=zeros(4,1);
```

```
figure;

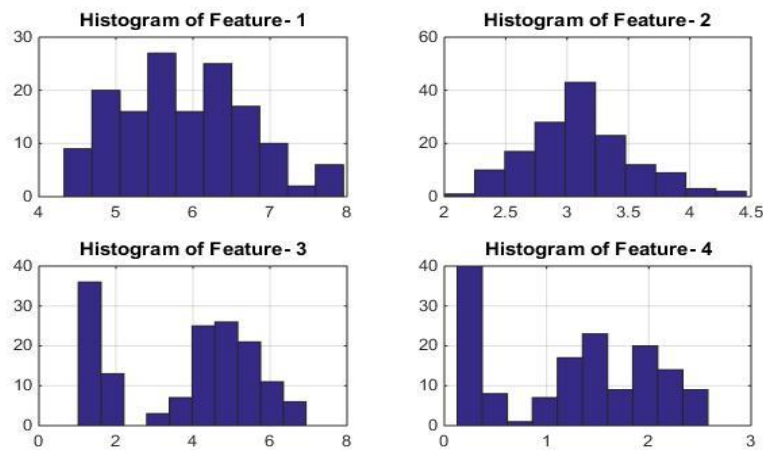
subplot(3,1,1);
gscatter(x(:,1),x(:,2),y(:,1),'bgr','xo+');
xlabel('feature1');
ylabel('feature2');
subplot(3,1,2);
gscatter(x(:,1),x(:,3),y(:,1),'bgr','xo+');
xlabel('feature1');
ylabel('feature3');
subplot(3,1,3)
gscatter(x(:,1),x(:,4),y(:,1),'bgr','xo+');
xlabel('feature1');
ylabel('feature4');
%title('Scatter Plot')
```

(a)

```
no_of_features = size(x,2); (ANS : 4)
i.e. % get no of features
no_of_data = size(x,1); (ANS: 148)
points i.e. % get no of data
rows equal to the no of
```

(b)

Histogram



(c)

Computing Mean, Variance and Standard Deviation

```
MEAN =zeros(4,1);
VAR=zeros(4,1);

for i=1:no_of_features
    subplot(2,2,i);
```

```

        hist(x(:,i));
        MEAN(i,1)=mean(x(:,i));
        VAR(i,1)=var(x(:,i));
    end
    STD=VAR.^(1/2);
    %NORMAL=zeros(size(x));

```

ANS:

```

MEAN =
    5.90010376418919
    3.09893091689189
    3.81955484054054
    1.25255548459459]

```

```

VAR=
    0.699283896094493
    0.191645469728850
    3.09764638598857
    0.579652210531846]

```

```

STD_DEV=
    0.836231963090680
    0.437773308607149
    1.76001317778833
    0.761348941374351]

```

(e)

Normalize Data

```

ONES=ones(no_of_data,1);
for i=1:no_of_features
    x(:,i)=(x(:,i)-ONES.*MEAN(i,1));
    x(:,i)=x(:,i)./STD(i,1);
end

```

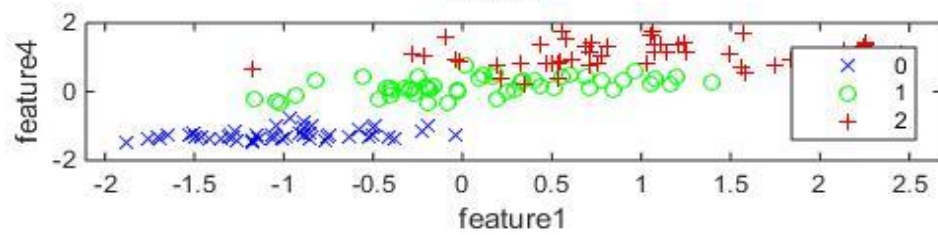
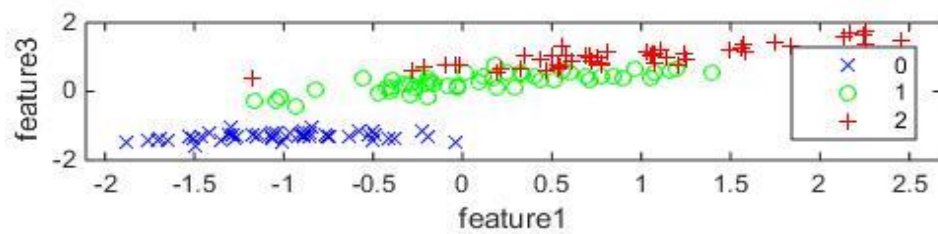
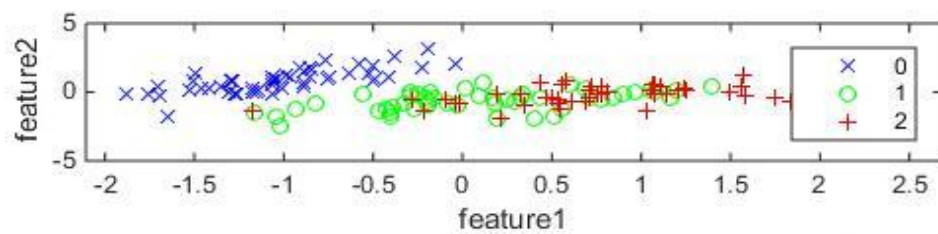
(f)

Scatter Plot

```

subplot(3,1,1);
gscatter(x(:,1),x(:,2),y(:,1),'bgr','xo+');
xlabel('feature1');
ylabel('feature2');
subplot(3,1,2);
gscatter(x(:,1),x(:,3),y(:,1),'bgr','xo+');
xlabel('feature1');
ylabel('feature3');
subplot(3,1,3)
gscatter(x(:,1),x(:,4),y(:,1),'bgr','xo+');
xlabel('feature1');

```

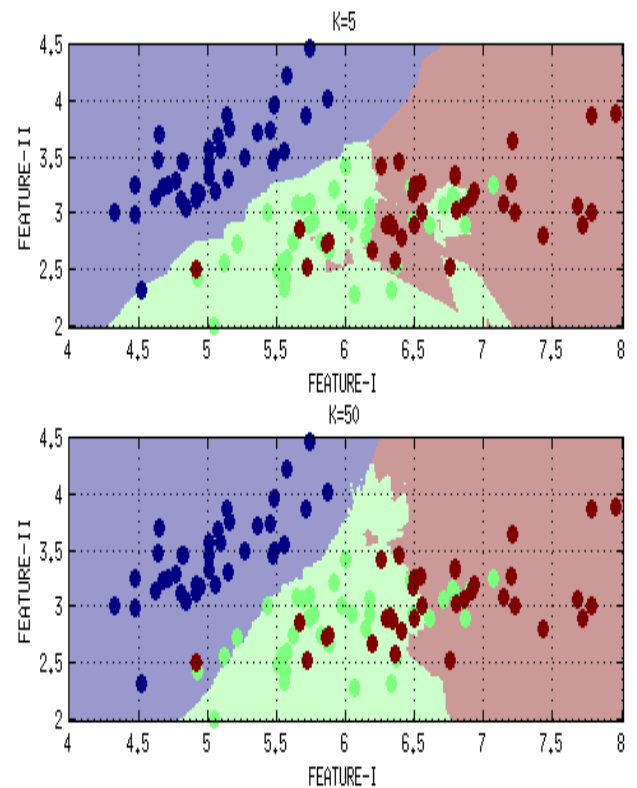
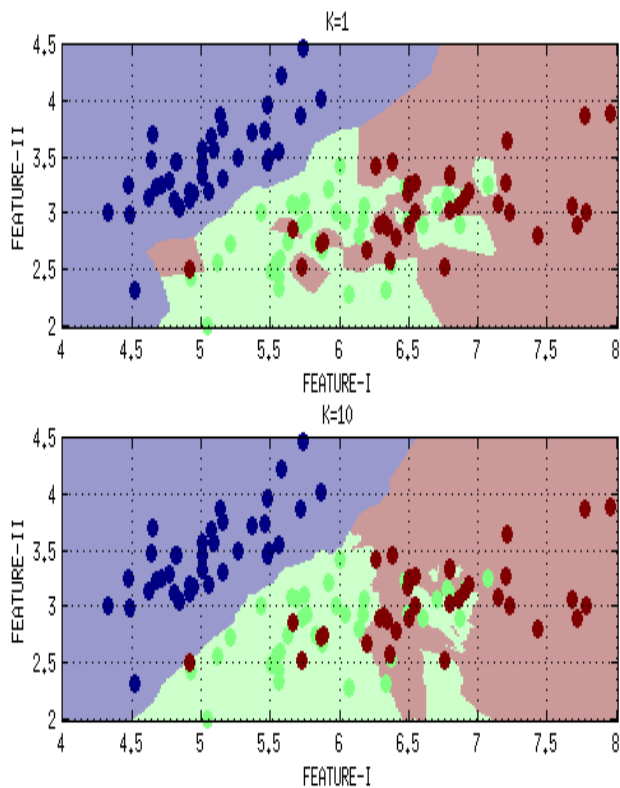


Problem 2

```
%% Problem- 2(a)
j=1;
for K=[10, 50, 100, 200]

knn = knnClassify( Xtr, Ytr, K );
YteHat = predict( knn, Xte );

subplot(2,2,j)
plotClassify2D( knn, Xtr, Ytr );
j=j+1;
end
```



```
%% Problem- 2(b)
%K=[1,2,5,10,50,100,200];
i=1;
for K=[1,2,5,10,50,100,200];
model = knnClassify( Xtr, Ytr, K );
Yhat = predict( model, Xtr );

%model = knnClassify( Xtr, Ytr, K );
Yhat_test = predict( model, Xte );

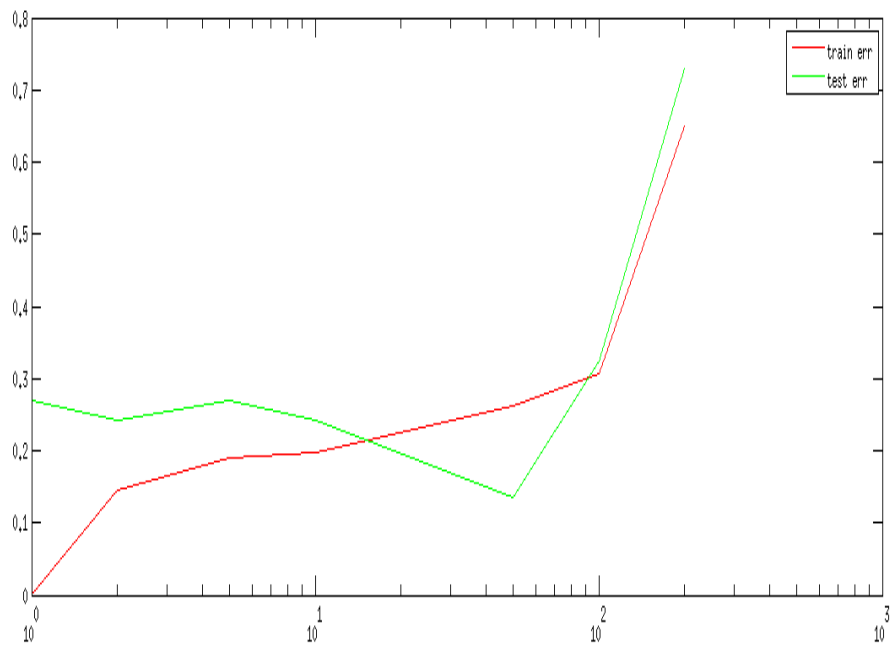
errTrain(i) = mean(Yhat~=Ytr);
errTest(i) = mean(Yhat_test~=Yte);
```

```

i=i+1;
end

K=[1,2,5,10,50,100,200];
semilogx(K,errTrain,'r');
hold on;
semilogx(K,errTest,'g');

```



I would recommend K around 15 as the Test Error is less.

PROBLEM - 4

```

clear all;

iris=load('data/iris.txt');

y=iris(:,end);
x=iris(:,1:2);
[x, y] = shuffleData(x,y);
[Xtr,Xte ,Ytr, Yte] = splitData(x,y, .75);

%% 4(a)
m=1;
n=1;

```

```

p=1;
for i=1:length(Xtr(:,1))
    if(Ytr(i)==0)
        Zero(m,:)=Xtr(i,:);
        m=m+1;
    elseif(Ytr(i)==1)
        One(n,:)= Xtr(i,:);
        n=n+1;
    elseif(Ytr(i)==2)
        Two(p,:)= Xtr(i,:);
        p=p+1;
    end
end
end

```

```

MEAN(:, :, 1)=mean(Zero);
MEAN(:, :, 2)=mean(One);
MEAN(:, :, 3)=mean(Two);
%COV=zeros(2,2,3);
COV(:, :, 1)=cov(Zero);
COV(:, :, 2)=cov(One);
COV(:, :, 3)=cov(Two);

```

ANS:

Mean

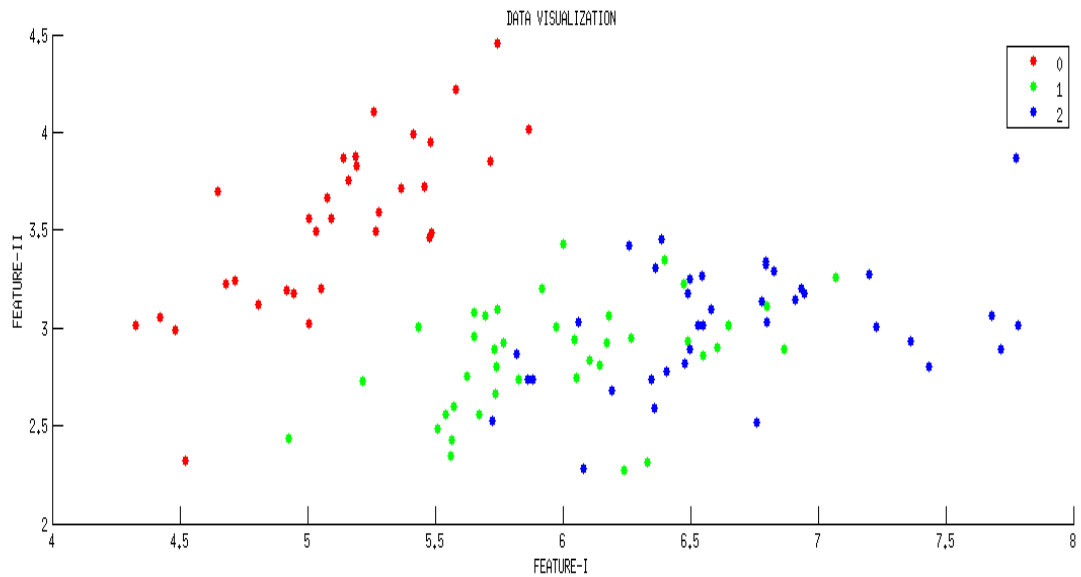
Feature-I	5.0441	3.4936
Feature-II	5.9886	2.8439
Feature-III	6.6531	3.0034

COV MATRIX

Feature-I	0.1460	0.1154
	0.1154	0.1648
Feature-II	0.2939	0.1082
	0.1082	0.1077
Feature-III	0.3364	0.0725
	0.0725	0.0898

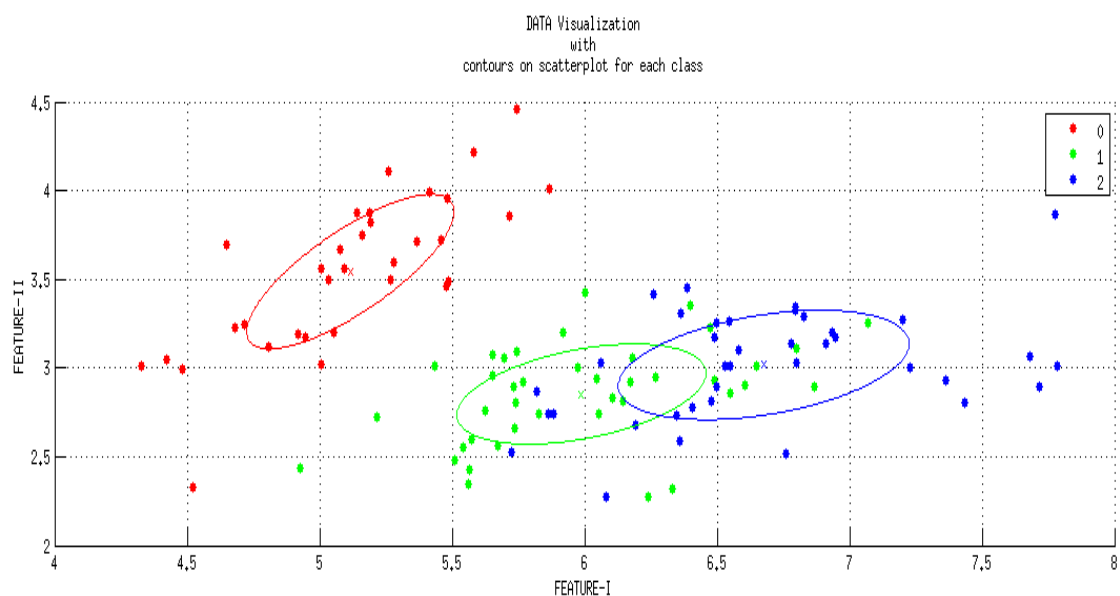
%% 4(b)

```
group=Ytr;
figure;
gscatter(Xtr(:,1),Xtr(:,2),group)
```



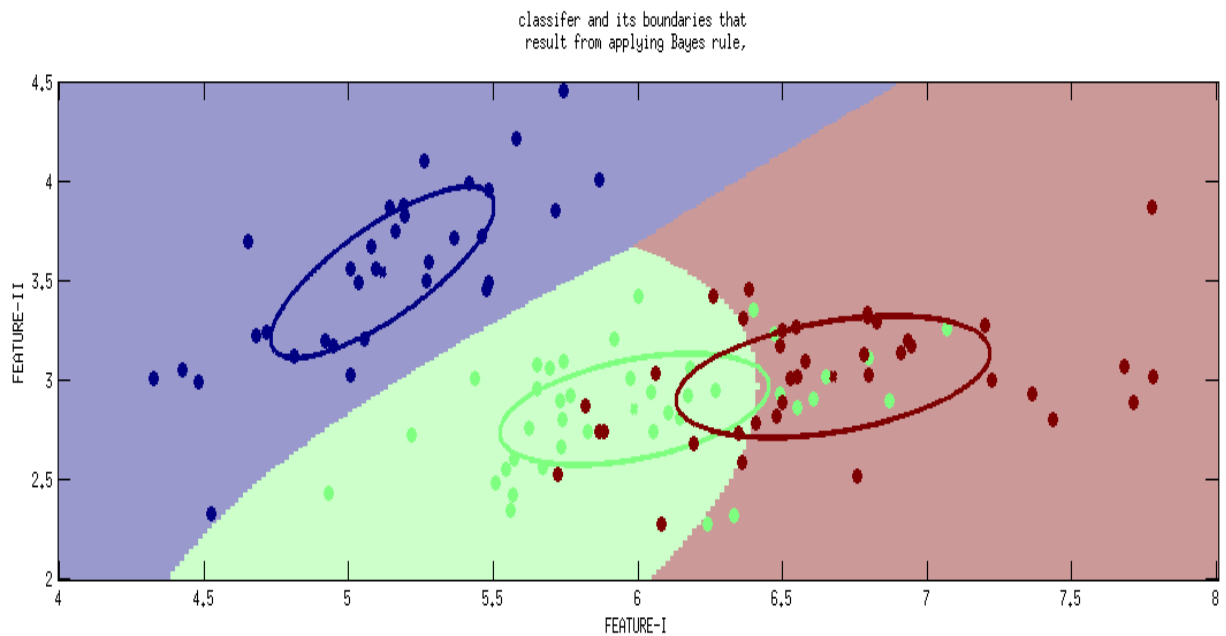
%% 4(c)

```
figure;
gscatter(Xtr(:,1),Xtr(:,2),group)
hold on;
plotGauss2D(MEAN(:, :, 1), COV(:, :, 1), '-r');
plotGauss2D(MEAN(:, :, 2), COV(:, :, 2), '-g');
plotGauss2D(MEAN(:, :, 3), COV(:, :, 3), '-b');
```



%% 4(d)

```
bc = gaussBayesClassify( Xtr, Ytr );  
figure;  
plotClassify2D(bc, Xtr, Ytr);
```



%% 4(e)

```
%old =cd('')  
Yhat = predict( bc, Xtr );  
errTrain = mean(Yhat~=Ytr);  
Yhat_test=predict(bc,Xte);  
errTest=mean(Yhat_test~=Yte);
```

ANS:

```
errTrain = 0.18  
errTest= 0.1892
```

%% 4(f)

```
iris=load('data/iris.txt');  
  
y=iris(:,end);  
x=iris(:,1:end-1);  
[x, y] = shuffleData(x,y);
```

```
[Xtr,Xte ,Ytr, Yte] = splitData(x,y, .60);
```

```
bc = gaussBayesClassify( Xtr, Ytr );
```

```
Yhat = predict( bc, Xtr );
```

```
errTrain_all = mean(Yhat~=Ytr);
```

```
Yhat_test=predict(bc,Xte);
```

```
errTest_all=mean(Yhat_test~=Yte);
```

ANS:

errTrain-using all features= 0.011235955056180

errTest-using all Features = 0.033898305084746