Homework 8 - Nicholas Hardy U97871602

```
%%MATLAB Likelihood Script for Homework 8
```

ANSWERS

dataset0 = syntheticH0.csv

dataset1 = syntheticH1.csv

Probability of error for identity covariance is **0.068**

Probability of error for the same covariance is **0.05**

Probability of error for different covariances is **0.045**

dataset0 = benignfull.csv

dataset1 = maliggnantfull.csv

Probability of error for identity covariance is **0.48**

Probability of error for the same covariance is **0.065**

Probability of error for different covariances is **0.015**

Observations: The probability of error for the identity covariance is much higher than I had expected for the second data used.

Math explanation: Because of the dot product will be larger, we get that the quantity is maximized when Y is closer to mu1 than to mu0.

```
%  Choose which data set you will read

%dataset0 = readmatrix("syntheticH0.csv");
%dataset1 = readmatrix("syntheticH1.csv");
dataset0 = readmatrix("benignfull.csv");
dataset1 = readmatrix("malignantfull.csv");
%[dataset0 dataset1] = read_cats_dogs;


[n0 d0] = size(dataset0);
[n1 d1] = size(dataset1);
if (d0 == d1)
    d = d0;
else
    error("dataset0 and dataset1 have a different number of columns.")
```

```
end

%Split dataset into training and test data.
train0 = dataset0(1:floor(n0/2),:);
test0 = dataset0(floor(n0/2)+1:n0,:);
train1 = dataset1(1:floor(n1/2),:);
test1 = dataset1(floor(n1/2)+1:n1,:);
n0train = size(train0,1);
n1train = size(train1,1);
n0test = size(test0,1);
n1test = size(test1,1);

%Estimate mean vectors and covariance matrices from training data.
mu0 = mean(train0);
mu1 = mean(train1);
sigma0 = cov(train0);
sigma1 = cov(train1);

%Apply decision rules.
H0guesses_idcov = zeros(n0test,1);
H1guesses_idcov = zeros(n1test,1);

for i = 1:n0test
    currentdata= test0(i,:);
    %%%Problem 8.4(a) code goes here. Compute the pdf of the gaussian model
    %%% at current data under hypothesis H0 (Gaussian(mu0,I)) and
    %%% under hypothesis H1 (Gaussian(mu1,I), where I is the identity
    %%% Use I = eye(d0) to generate an identity matrix with the same
    %%% dimension as the number of features in the data.
    %%% Pick the one that is largest, and store the decision as 1 if H1
    %%% is more likely, 0 if H0 is more likely in the array below.

   pdf_H0 = mvnpdf(currentdata, mu0, eye(d0));
   pdf_H1 = mvnpdf(currentdata, mu1, eye(d0));

   if pdf_H1 > pdf_H0
       H0guesses_idcov(i) = 1;
   else
       H0guesses_idcov(i) = 0;
   end

end

for i = 1:n1test
    currentdata = test1(i,:);
    %%%Problem 8.4(a) code goes here. Do the same thing as above for the
    %%%test data from class 1.

    pdf_H0 = mvnpdf(currentdata,mu0,eye(d1));
    pdf_H1 = mvnpdf(currentdata,mu1,eye(d1));

    if pdf_H1 > pdf_H0
        H1guesses_idcov(i) = 1;
    else
```

```
        H1guesses_idcov(i) = 0;
    end
end

Pe_idcov = proberror(H0guesses_idcov,H1guesses_idcov);
fprintf('Probability of error for identity covariance is %.2g.\n',Pe_idcov);
```

```
Probability of error for identity covariance is 0.48.
```

```
H0guesses_samecov = zeros(n0test,1);
H1guesses_samecov = zeros(n1test,1);

%%%Problem 8.4(b) code goes here %%%
%%% First, compute the pooled covariance matrix as indicated in the homework
%%% sheet.
S0 = cov(train0);
S1 = cov(train1);
n = n0train + n1train;
Sp = ((n0train-1)/n)*S0 + ((n1train-1)/n)*S1;

PooledVar = Sp;

for i = 1:n0test
    currentdata = test0(i,:);
    %%%Problem 8.3(b) code goes here. Compute the pdf of the gaussian model
    %%% at current data under hypothesis H0 (Gaussian(mu0,PooledVar)) and
    %%% under hypothesis H1 (Gaussian(mu1,PooledVar).
    %%% Pick the one that is largest, and store the decision as 1 if H1
    %%% is more likely, 0 if H0 is more likely in the array below.
    pdf_H0 = mvnpdf(currentdata,mu0,PooledVar);
    pdf_H1 = mvnpdf(currentdata,mu1,PooledVar);

    if pdf_H1 > pdf_H0
        H0guesses_samecov(i) = 1;
    else
        H0guesses_samecov(i) = 0;
    end

end

for i = 1:n1test
    currentdata = test1(i,:);
        %%%Problem 8.4(b) code goes here. Repeat the above for data from
        %%%test1
    pdf_H0 = mvnpdf(currentdata,mu0,PooledVar);
    pdf_H1 = mvnpdf(currentdata,mu1,PooledVar);

    if pdf_H1 > pdf_H0
        H1guesses_samecov(i) = 1;
    else
        H1guesses_samecov(i) = 0;
    end
end
```

```
Pe_samecov = proberror(H0guesses_samecov,H1guesses_samecov);
fprintf('Probability of error for same covariance is %.2g.\n',Pe_samecov);
```

```
Probability of error for same covariance is 0.065.
```

```
H0guesses_diffcov = zeros(n0test,1);
H1guesses_diffcov = zeros(n1test,1);

for i = 1:n0test
    currentdata = test0(i,:);
    %%%Problem 8.4(c) code goes here. Compute the pdf of the gaussian model
    %%% at current data under hypothesis H0 (Gaussian(mu0,sigma0)) and
    %%% under hypothesis H1 (Gaussian(mu1,sigma1), where sigma0 and sigma1
    %%% are the covariances computed from the training data.
    %%% Pick the one that is largest, and store the decision as 1 if H1
    %%% is more likely, 0 if H0 is more likely in the array below.

    pdf_H0 = mvnpdf(currentdata, mu0, sigma0);
    pdf_H1 = mvnpdf(currentdata, mu1, sigma1);

    if pdf_H1 > pdf_H0
        H0guesses_diffcov(i) = 1;
    else
        H0guesses_diffcov(i) = 0;
    end
end

for i = 1:n1test
    currentdata = test1(i,:);

    %%%Problem 8.4(c) code goes here.  Repeat the
    %%% above for data from test1 %%%

    pdf_H0 = mvnpdf(currentdata, mu0, sigma0);
    pdf_H1 = mvnpdf(currentdata, mu1, sigma1);

    if pdf_H1 > pdf_H0
        H1guesses_diffcov(i) = 1;
    else
        H1guesses_diffcov(i) = 0;
    end
end
Pe_samecov = proberror(H0guesses_diffcov,H1guesses_diffcov);
fprintf('Probability of error for different covariance is %.2g.\n',Pe_samecov);
```

```
Probability of error for different covariance is 0.015.
```