ECE-597: Probability, Random Processes, and Estimation
Computer Project #3

Due: April 3, 2015

In this project, we will generate multivariate Gaussian random vectors with a given mean and covariance, and then use this 'training data' to determine the optimal linear discriminant and guess a threshold. We will then generate new data to test with. You should do all of this in Matlab, you do not need to do anything by hand.

1) Determine the matrix B such that if $\underline{\mathbf{X}}$ is a random vector with zero mean, and uncorrelated components with

$$K_{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

then the random variable $\underline{\mathbf{Y}}_1 = A\underline{\mathbf{X}} + \underline{\mu}_{\mathbf{Y}_1}$ will have covariance matrix

$$K_{\mathbf{Y_1}} = A * K_{\mathbf{X}} * A^T = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

and mean $\underline{\mu}_{\mathbf{Y}_1}$

2) Using Matlab, generate 1000 2-dimensional random vectors ($\underline{X}$'s) with uncorrelated components, each component with an $N(0,1)$ distribution using the command

```
X=randn(2,1000)
```

and construct the random variables $\underline{\mathbf{Y}}_1$ so they will have covariance $K_{\mathbf{Y1}}$ and mean

$$\underline{\mu}_{\mathbf{Y_1}} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

3) Now construct random variables $\underline{\mathbf{Y}}_2$ with 1000 samples and with covariance and mean

$$K_{\mathbf{Y2}} = \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix}, \underline{\mu}_{\mathbf{Y}_2} = \begin{bmatrix} -3 \\ -5 \end{bmatrix}$$

4) To determine if your transformation is correct, estimate the means and covariances using the (unbiased) estimators

$$\begin{aligned} \underline{\hat{\mu}} &= \frac{1}{N} \sum_{i=1}^{i=N} \underline{\mathbf{Y}}(i) \\ \hat{K} &= \frac{1}{N-1} \sum_{i=1}^{i=N} (\underline{\mathbf{Y}}(i) - \underline{\hat{\mu}})(\underline{\mathbf{Y}}(i) - \underline{\hat{\mu}})^T \end{aligned}$$

You may want to use a larger sample size at this point to be sure your estimators are converging to the correct values.

5) Now, your variables $\underline{\mathbf{Y}}_1$ and $\underline{\mathbf{Y}}_2$ are your "training" data. That is, you have measured these parameters and know that data from $\underline{\mathbf{Y}}_1$ supports hypothesis one and the data from $\underline{\mathbf{Y}}_2$ supports another hypothesis two. Now construct the optimal linear discriminant

$$\underline{a} = (\hat{K}_{\mathbf{Y}_1} + \hat{K}_{\mathbf{Y}_2})^{-1}(\underline{\hat{\mu}}_{\mathbf{Y}_1} - \underline{\hat{\mu}}_{\mathbf{Y}_2})$$

Note that we need to use the estimates generated by the "training" data, since in reality we don't know the real underlying distributions. Now form $\mathbf{Z}_1 = \underline{a}^T\underline{\mathbf{Y}}_1$ and $\mathbf{Z}_2 = \underline{a}^T\underline{\mathbf{Y}}_2$. Note that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are *scalars*.

6) Now plot some of the results. There are a variety of ways to look at what is going on, but some more useful ways are the following (feel free to do what works for you):

```
%
%  plot Y1 and Y2
%
  figure;
  plot(Y1(1,:),Y1(2,:),'kx',Y2(1,:),Y2(2,:),'ro'); grid;
  legend('Y_1','Y_2');
%
% put both histograms on one plot
%
  [n1,z1out] = hist(Z1,20);
  [n2,z2out] = hist(Z2,30);
  figure;
  hold on
  bar(z1out,n1,'k'); axis([-6 4 -max(n2) max(n1)]);
  bar(z2out,-n2,'r');
  legend('Z_1', '-Z_2');
  ylabel('occurrences'); xlabel('Z value');
  grid;
  hold off
%
%  plot values of Z versus sample
%
  S = [1:1:N];
  figure;
  plot(S,Z1,'kx',S,Z2,'ro'); grid; legend('Z_1','Z_2');
  xlabel('Sample'); ylabel('Z values');
```

Include some of these figures in your write-up.

7) Now, based on your "training" data, determine what you think an "appropriate" discriminat point ν_o will be (i.e., if $\mathbf{Z}_1 > \nu_o$ classify it as hypothesis 1, if $\mathbf{Z}_2 < \nu_o$ classify it as hypothesis 2.) Be sure to indicate to me why you think this is an "appropriate" discriminant point based only on the "training" data. (Note: there are no "wrong" answers, as long as

you choose ν_o in some reasonable way...)

8) Now compute a new set of data $\underline{\mathbf{Y}}_1$ and $\underline{\mathbf{Y}}_2$. Determine how many bad decisions you get, i.e, how many times $\mathbf{Z}_2 > \nu_o$ and how many times $\mathbf{Z}_1 < \nu_o$. Include this in your write-up. DO NOT change the value of the threshold based on the "real data". You are doing ok if you are above 80% correct for this part. Be sure to indicate your results in your write-up.

9) Now we're going to do the whole thing again in 3 dimensions. Assume we want

$$K_{\mathbf{Y}_1} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 3 & 2 \\ 3 & 2 & 5 \end{bmatrix}, \underline{\mu}_{\mathbf{Y}_1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, K_{\mathbf{Y}_2} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 1 \\ 1 & 1 & 5 \end{bmatrix}, \underline{\mu}_{\mathbf{Y}_2} = \begin{bmatrix} 4 \\ 8 \\ 10 \end{bmatrix}$$

Go through the whole thing again, with 1000 training vectors and 1000 test vectors. It is best to copy your program to a different file (one for 2D and one for 3D) so you do not mess up what already works. We will also use these programs again. You will have to change the *axis* parameters to get a nice plot of the histograms.

Hints:

(1) To plot in three dimensions, use

```
plot3(Y1(1,:),Y1(2,:),Y1(3,:),'x'....)
```

(2) To add the mean (a vector) to the matrix of samples, use

```
[M,N] = size(X)     %
Y = B*X + mY*ones(1,N) %
```

(3) To determine how many times the components of a vector are grater than a prescribed amount, use

```
sum(Z1>v)
```

(4) Your write-up should be short and neat, but not just computer listing!!