EECE 5698

Homework 1: Due Wednesday, Jan. 15, 2014

Reading:

- Appendices A.1, A.2, A.3, A.4, A.5
- Notes, Chapter 2.1-2.5

Problems:

1. Let x be a real-valued random variable.
   a. Prove that the variance of $x = \sigma^2 = E[(x-\mu)^2] = E[x^2] - \mu^2$.
   b. Let $\mathbf{x}$ be a real-valued random vector. Prove that the covariance matrix of $\mathbf{x} = \Sigma = E[\mathbf{x}\mathbf{x}^t] - \boldsymbol{\mu}\boldsymbol{\mu}^t$.

2. Consider a two-class problem, with classes c_1 and c_2 where $P(c_1) = P(c_2) = 0.5$. There is a one-dimensional feature variable x. Assume that the x data for class one is uniformly distributed between a and b, and the x data for class two is uniformly distributed between r and t. Assume that $a < r < b < t$. Derive a general expression for the Bayes error rate for this problem. (Hint: a sketch may help you think about the solution.)

3. Consider a two-class, one-dimensional problem where $P(\omega_1) = P(\omega_2) = 0.5$, and

   $p(x|\omega_i) \sim N(\mu_i, \sigma_i^2), \quad i = 1, 2$

Let $\mu_1 = 0$, $\sigma_1^2 = 1$, $\mu_2 = \mu$, $\sigma_2^2 = \sigma^2$.

   a. Derive a general expression for the location of the Bayes optimal decision boundary as a function of μ and σ^2.
   b. Let $\mu = 1$, $\sigma^2 = 2$.
      i. Determine the location of the optimal decision boundaries.
      ii. Sketch the densities $p(x|\omega_i)$, the posterior probabilities for $p(\omega_i|x)$, and the location of the optimal decision regions.
      iii. Estimate the Bayes error rate p_e^* for the problem. (Hint: Use can use Table A.1 to perform integration.)
   c. Comment on the case where $\mu = 0$, and σ^2 is much greater than 1. Describe a practical example of a pattern classification problem where such a situation might arise.

Computer Assignment:

Turn in documented print outs of your code, plots and write-ups as requested by the assignment.

1. Write a simple MATLAB function called *twogaussian.m*. The purpose of this exercise is primarily to get you working with MATLAB and to write a simple function to be used in later homeworks. A template, which contains 90% of the code, is shown in template.m. You can download template.m and mvnrnd.m from blackboard. You are to fill in the few lines (denoted with "dots") that are missing in template.m. Use "help plot" to find out about various options in plotting your own options to the code beyond what is required.

This is the header for the MATLAB function:

```
function [data1, data2] = twogaussian(n1,mu1,cov1,n2,mu2,cov2);
%
%  [data1, data2] = twogaussian(n1,mu1,sigma1,n2,mu2,sigma2);
%
%  Function to simulate data from 2 Gaussian densities in d dimensions
%       and to plot the data in the first 2 dimensions
%
%  INPUTS:
%       n1, n2: two integers, size of data set 1 and 2 respectively
%       mu1, mu2: two vectors of dimension 1 x d, means
%                       for data set 1 and 2
%       cov1, cov2: two matrices of dimension d x d, covariance
%                       matrices for data set 1 and 2 respectively
%
%  OUTPUTS:
%       data1: n1 x d matrix of data for data set 1
%       data2: n2 x d matrix of data for data set 2
```

Be sure to test your function with some simple cases to validate that it is working correctly.

2. Produce two-dimensional "scatter plots" for the following three cases.

- n1 = 100, mu1 = (0,0), cov1 = I, where I is the identity matrix, and n2 = 100, mu2=(2,2), cov2 = I.
- Same as above, but now cov1 = (3, 0; 0, 3).
- n1 = 100, mu1 = (0,0), cov1 = (1 0.9; 0.9 1), where I is the identity matrix, and n2 =100, mu2 = (1,1), cov2 = (1 -0.9; -0.9 1).

Note color plots are not necessary for your printouts that you hand it with your homework, but if your plots are not in color then you need to modify your code so that the data points from "data1" are plotted with one symbol (e.g., "+") and data points from "data2" are plotted with a different symbol (e.g.,"o").