BIOMETRICS
CSE190
Fall 2006

Assignment 2
Due: October 31, 2006

**1.** Duda, Hart, Stork 3.35.
Let the sample mean $\mu_n$ and the sample covariance matrix $C_n$ for a set of $n$ samples $\mathbf{x_1 \ldots x_n}$ (each of which is d-dimensional) be defined by

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$$

$$C_n = \frac{1}{n-1}\sum (x_i - \mu_n)(x_i - \mu_n)^t$$

We call these the "nonrecursive" formulae.

(a) What is the computational complexity of calculating $\mu_n$ and $C_n$ by these formulae?
(b) Show that the alternative" recursive techniques based on successive addition of new samples $\mathbf{x_{n+1}}$ can be derived using the recursive relations

$$\mu_{n+1} = \mu_n + \frac{1}{n+1}(\mathbf{x}_{n+1} - \mu_n)$$

$$C_n = \frac{n-1}{n} C_n + \frac{1}{n+1}(x_{n+1} - \mu_n)(x_{n+1} - \mu_n)^t$$

(c) What is the computational complexity of finding $\mu_n$ and $C_n$ by these recursive methods?
(d) Describe situations where you might prefer to use the recursive method for computing $\mu_n$ and $C_{n,}$, and ones where you might prefer the nonrecursive method?

**2.** Consider a normal $p(x)=N(\mu,\sigma^2)$ and Parzen window function $\varphi(x) = N(\mu,1)$. Show that the Parzen window estimate

$$p_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{h_n}\right)$$

has the following property

$E[p_n(x)] = N(\mu,\sigma^2+h_n^2)$

**3.** Consider the following set of two dimensional vectors from three categories:

| $\omega_1$ | | $\omega_2$ | | $\omega_3$ | |
|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_1$ | $X_2$ | $X_1$ | $X_2$ |
| 10 | 0 | 5 | 10 | 2 | 8 |
| 0 | -10 | 0 | 5 | -5 | 2 |
| 5 | -2 | 5 | 5 | 10 | -4 |

(a) Plot the decision boundary resulting from the nearest neighbor rule just for categorizing $\omega_1$ and $\omega_2$. Find the sample mean $\mathbf{m_1}$ and $\mathbf{m_2}$ and on the same figure sketch the decision boundary corresponding to classifying x by assigning it to the category of the nearest sample mean.
(b) Repeat part (a) for categorizing only $\omega_1$ and $\omega_3$.
(c) Repeat part (a) for categorizing only $\omega_2$ and $\omega_3$.
(d) Repeat part (a) for three-category classifier, classifying $\omega_1$, $\omega_2$ and $\omega_3$.