MIT OpenCourseWare
http://ocw.mit.edu

HST.582J / 6.555J / 16.456J Biomedical Signal and Image Processing
Spring 2007

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

Harvard-MIT Division of Health Sciences and Technology
HST.582J: Biomedical Signal and Image Processing, Spring 2007
Course Director: Dr. Julie Greenberg

HST-582J/6.555J/16.456J - Biological Signal and Image Processing - Spring 2007

Problem Set 4
Due April 12, 2007

Problem 1

Bayes' Rule or Am I really sick?

During your visit to the doctor, a test is ordered to determine whether or not you have a certain disease. The test result is positive with probability 0.95 if you have the disease and 0.025 if you do not. The incidence of the disease in the general population is 1 in 10,000.

a) If the test comes back positive what is the probability that you have the disease?

b) If it comes back negative what is the probability that you do not have the disease?

c) If α is the probability of a positive test result when you have the disease and $1 - \alpha$ is probability of a positive test result when you do not have the disease, what is the smallest value of α such that a positive test indicates a 0.95 probability of having the disease?

Problem 2

k-NN Classifiers: Generative vs. Discriminative Approaches

If we have a measurement, x, drawn from one of two class-conditional densities, $p_1(x)$ or $p_0(x)$ where P_1 and $P_0 = (1 - P_1)$ are the *a priori* probabilities of classes 1 and 0, respectively, then the optimal decision rule for minimizing the probability of error can be written:

$$\text{if} \quad \frac{p_1(x)}{p_0(x)} > \frac{P_0}{P_1} \quad \text{assign to class 1, otherwise assign to class 0.}$$

Unless otherwise specified we'll assume that $P_1 = P_0 = \frac{1}{2}$. Additionally, assume that all measurements x are two dimensional (NOTE: this only affects how we measure volume). Suppose we are given N_1 labeled samples of x from class 1 and N_0 (equal to N_1) labeled samples of x from class 0.

Cite as: Julie Greenberg, Bertrand Delgutte, William (Sandy) Wells, John Fisher, and Gari Clifford. Course materials for HST.582J / 6.555J / 16.456J, Biomedical Signal and Image Processing, Spring 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

a) A *generative* k-NN classifier would substitute the k-NN density estimate into the decision rule above. The k-NN density estimate is defined as:

$$\hat{p}_i(x) = \frac{k_g}{N_i V(x,r)}$$

where $V(x,r)$ is the area of the circle with radius r, r is the smallest radius such that a circle centered at x contains at least k_g samples from class i. Express the decision rule using the k-NN density estimate in it's simplest form. What are we comparing in order to make our decision?

b) A *discriminative* k-NN classifier rule, as described in the chapter from Bishop's book, is:

1. Find the radius of the circle centered at x such that it contains at least k_d samples (without regard to their labels).
2. Count the number of samples with label 1 versus label 0.
3. Classify the sample according to the majority.

Assuming both classifier make use of the same labeled samples and given k_g, can one determine k_d such that the generative and discriminative classifiers result in the same classification at *every* point x? If so, what is the relationship between k_d and k_g? If not, why not?

c) Given the exact probability models, suppose we occasionally encounter samples x such that $p_1(x)/p_0(x) = P_0/P_1$, if we always classify such samples as class 1 will we still achieve the minimum classification error (assume we average over a very large number of trials)?

d) Using the k-NN probability density estimates in a generative classifier, suppose we occasionally encounter samples x such that $\hat{p}_1(x)/\hat{p}_0(x) = P_0/P_1$. Explain why *always* classifying such samples as class 1 (or class 0) would not be the preferred decision rule. Suggest an alternate decision rule for such samples.

Problem 3

Receiver Operating Characteristic Curve

A common tool for assessing the quality of a classifier is the receiver operating characteristic curve or **ROC** curve for short. It is a plot of the probability of a "false alarm" versus the probability of a "detection" over the range of some decision rule (e.g. a varying threshold).

Suppose we have one of two conditions which we will refer to as H_0 and H_1, respectively.

Cite as: Julie Greenberg, Bertrand Delgutte, William (Sandy) Wells, John Fisher, and Gari Clifford. Course materials for HST.582J / 6.555J / 16.456J, Biomedical Signal and Image Processing, Spring 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Additionally, we have a related random variable X which we wish to use to determine which condition is currently in effect. We have the following model:

$$H_0 \;:\; x \sim p_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$
$$H_1 \;:\; x \sim p_1(x) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-1.5)^2\right) + \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x+1.5)^2\right)$$

which indicates that if H_0 is in effect the random variable obeys a Gaussian density with $\mu = 0$ and $\sigma = 1$ and that if H_1 is in effect the random variable obeys a density which is a sum of Gaussians with $\mu = \pm 1.5$ and $\sigma = 1$.

In class we learned that the optimal classifier takes the form:

$$R_0(\gamma) = \left\{x : \frac{p_0(x)}{p_1(x)} > \gamma\right\} \qquad R_1(\gamma) = \left\{x : \frac{p_0(x)}{p_1(x)} < \gamma\right\}$$

A *false alarm* occurs when x falls in R_1 when H_0 is correct for a given value of γ. The probabilty of a false alarm is:

$$P_{\mathrm{fa}}(\gamma) = \mathbf{Pr}\left\{x \in R_1(\gamma)\,|H_0\right\}$$

A *detection* is when x falls in R_1 when H_1 is correct. The probabilty of a detection is:

$$P_{\mathrm{d}}(\gamma) = \mathbf{Pr}\left\{x \in R_1(\gamma)\,|H_1\right\}$$

a) Express $P_{\mathrm{fa}}(\gamma)$ and $P_{\mathrm{d}}(\gamma)$ in terms of the CDFs $P_0(x)$ and $P_1(x)$.

b) An error occurs if we declare H_0 to be in effect when H_1 is in effect *or* if we declare H_1 to be in effect when H_0 is in effect. Express the probability of making an error, $P_{\mathrm{err}}(\gamma)$, in terms of $P_{\mathrm{fa}}(\gamma)$, $P_{\mathrm{d}}(\gamma)$, assuming *prior* class probabilities of P_0, and P_1.

c) Plot $P_{\mathrm{fa}}(\gamma)$ vs $P_{\mathrm{d}}(\gamma)$ over the range $\gamma = 0.01$ to 100 (sample γ with a log scale) with false alarm on the horizontal axis and detection on the vertical axis. This plot *is* the **ROC** curve.

d) **This part is strictly optional** Try creating random decision rules (e.g. choose random boundary points for R_0 and R_1) and then computing P_{fa} and P_{d}. Plot these points on the ROC curve from the previous part. Depending on how many you try (this can be automated easily with matlab) you may notice a property of the ROC curve with respect to these points.

Cite as: Julie Greenberg, Bertrand Delgutte, William (Sandy) Wells, John Fisher, and Gari Clifford. Course materials for HST.582J / 6.555J / 16.456J, Biomedical Signal and Image Processing, Spring 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].