CS 4803A/8803A: Pattern Recognition

Problem Set 2
Date: Jan 28, 2003 **Due: start of class Feb 11, 2001**

WARNING: Do not leave this for the last night before the PS is due. It takes some work.... And I will be out of town as of Feb 9.

1. In a particular binary hypothesis testing application, the conditional density for a scalar feature y given class ω_1 is

$$p_{y|\omega_1}(\hat{y}|\omega_1) = k_1 \exp(-\hat{y}^2/8).$$

Given class ω_2, the conditional density is

$$p_{y|\omega_2}(\hat{y}|\omega_2) = k_2 \exp(-(\hat{y}-2)^2/2).$$

   (a) Find k_1 and k_2, and plot the two densities on a single graph using Matlab.

   (b) Assume that the prior probabilities of the two classes are equal, and that the cost for choosing correctly is zero. If the costs for choosing incorrectly are $C_{12} = 1$ and $C_{21} = 2$, what is the expression for the Bayes risk?

   (c) Find the decision regions which minimize the Bayes risk, and indicate them on the plot you made in part (a).

   (d) For the decision regions you found in part (c), what is the numerical value of the Bayes risk? Hint: use Matlab's erf function but be careful - **the erf function is a bit weird. Check help.**

2. Hans and Frans each want to classify patterns into two classes ω_1 and ω_2, whose prior probabilities are equal. They both measure quantity x, but with different measuring devices.

   (a) Hans uses the latest model of the device, which we can call feature x_1, to do the classification. The probability distribution of x_1 for each class is:

	$x_1 = 1$	$x_1 = 2$	$x_1 = 3$
$p(x_1\|\omega_1)$	0.80	0.055	0.145
$p(x_1\|\omega_2)$	0.15	0.05	0.80

   Hans wants to use the decision rule which minimizes his error rate. What is the rule, and what is its error rate?

(b) Frans uses an older model of the device, which we can call feature x_2, to do the classification. The probability distribution of x_2 for each class is:

	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$
$p(x_2\|\omega_1)$	0.26	0.73	0.01
$p(x_2\|\omega_2)$	0.026	0.803	0.171

Frans also wants minimize his error rate. What rule should he use, and what is its error rate?

(c) What is Hans's confidence in his classification (how certain is he that he made the right choice), **as a function of his measurement x_1?** That is, for each possible value of x_1 and the decision rule you stated above, how certain is Hans that he's making the right choice? What is Frans's confidence in his classification, as a function of his measurement x_2?

(d) What does this tell you about the relationship between a classifier's error rate and our confidence in its classification? Does Hans or does Frans have a better classifier? Why?

3. You are designing software for an airport radar system. The airport runway constantly emits a radar signal which is reflected when an aircraft is landing. The radar return is a time series of discrete measurements $x[t]$. If there is no aircraft, the return is zero. If there is an aircraft, the return is a deterministic signal $s[t]$ of length n. Furthermore, the return always has noise added to it which we can model as zero-mean Gaussian with variance σ^2. You can assume that the noise at different times is independent.

   To detect aircraft, you will inspect every window of length n to see if it contains the characteristic signal $s[t]$. Each window can be considered a fixed-length vector $\mathbf{x}$. Show that an optimal decision rule, regardless of the priors and costs, has the form $\mathbf{x}^{\mathrm{T}}\mathbf{y} < t$ for some vector $\mathbf{y}$ and threshold t. Give one possible $\mathbf{y}$.

4. You are given two datasets (find them by clicking on "datasets" on the course home page). Each dataset contains 100 examples of a 3D random vector. Your task is to discriminate between these two datasets. To do that, you will do the following in Matlab:

   - Calculate the sample covariance matrices of the datasets. If we assume that the classes have the same covariance, then the best estimate of the common covariance is the average of the two sample covariance matrices.
   - Compute the eigenvectors of your estimate of the common covariance matrix.
   - And then answer the following questions:
     - Which eigenvectors capture most of the energy in the data?
     - Which eigenvectors permit you to discriminate most easily?

   Discuss your results. Do you find that the eigenvectors that capture most of the energy in the data (also known as the Most Expressive Feature [MEF]) also are the Most Discriminating Feature? If so, why? If not, why not?

5. Billy's Beetle Repainting of LA is a shop specializing in repainting VW Beetle automobiles. When the car is white (about 30% of the cars), they must do a quick measurement on the paint to determine if it's original, and quote a total price depending on the measurement. They've

hired you as a signal detection expert to help them calculate their costs, and to see if they can achieve 95% accuracy.

If the car has original paint, it costs an average of \$30 to clean and prep it for repainting. If it has already been repainted, it costs \$75 to clean and prep. But, if you use the \$30 cleaning process on a repainted car, you waste the \$30 and still have to spend \$75 more cleaning the car. Worse still, if you use the \$75 process on original paint, it creates such a mess that it costs another \$65 to get the car cleaned and prepped.

This story wouldn't be complete without a description of the measurement. On original paint, it returns a value between 3 and 5 with the probability ramping up linearly from 3 to 4 and linearly back down from 4 to 5 (looks like the sum of two uniform r.v.'s). On a repainted car, the measurement yields a gaussian with a mean of 6.5 and standard deviation of 1.2. Finally, Billy's highly educated guess is that about 65% of the VW Beetles in Southern California have already been repainted once.

Billy figures his customers will be happiest if he just charges them a fixed price depending on how the measurement classifies the car. He wants to know what value of the measurement he should use as a threshold, and the expected cost of classifying a car in each of the two categories (above and below the threshold). As a consultant, you have to jazz up your report with a few graphs and stuff. So,

(a) Use matlab to generate a graph of the two distributions.

(b) Now make an R.O.C. curve where
ϵ_2 is the x-axis; $\epsilon_2 = Pr[\text{misclassifying original paint car}]$
$(1-\epsilon_1)$ is the y-axis; $\epsilon_1 = Pr[\text{misclassifying repainted car}]$

(c) Plot the expected cost over a range of thresholds (you need 3 curves on this plot: expected cost for cars classified as original paint, expected cost for cars classified as repainted, and the sum of those two which is total expected cost).

(d) Find the threshold which classifies original cars correctly 95% of the time and the threshold which minimizes the sum of expected costs. For each one, give Billy's confidence in each classification outcome and the expected cost for each outcome, so Billy can set his prices.

6. Wilbur wonders whether or not he should speak up in class when he thinks he sees a mistake on the board. If he speaks up and there turns out not to be a mistake, he feels unhappy. He decides to restrict the probability of this horrible event to 0.05. Let ω_1 be the case where there really is a mistake on the board, and let ω_2 be the case where the board is correct. Over the years, Wilbur has noticed a relationship between the level of murmuring in class and the probability that there is an error on the board. In particular, he has determined that the probability density function for y, the dB audio level of students murmuring, is

$$p_{y|\omega_1}(\hat{y}|\omega_1) = \begin{cases} e^{-y} & \text{if } y > 0; \\ 0 & \text{otherwise} \end{cases}$$

when there is a mistake on the board, and

$$p_{y|\omega_2}(\hat{y}|\omega_2) = \begin{cases} 3e^{-3y} & \text{if } y > 0; \\ 0 & \text{otherwise} \end{cases}$$

when there is no mistake. Wilbur makes his decision whether or not to speak up based on the murmuring level y. Let P_D be the probability that Wilbur speaks up correctly, i.e., when there is a mistake on the board, and let P_F be the probability that he speaks up when there is no mistake on the board. Design a decision rule so that P_D is maximized subject to the constraint that $P_F \leq 0.05$. What is the resulting value of P_D?

7. This problem addresses the issue of making decisions when the class distributions are themselves uncertain. Perhaps surprisingly, this situation can be handled without any extra machinery beyond what you have been given so far.

   The problem:

   You want to determine if a fish taken from a given pond is a carp or bass, based on its weight in pounds. You consult two experts on fish who agree on some things and disagree on others. They both agree that:

   - Carp and bass are equally likely to be drawn from the pond.
   - The weight of either type of fish is Gaussian-distributed with a standard deviation of one pound.
   - The mean weight of a carp is 3 pounds.

   However, one expert says that the mean weight of a bass is 2 pounds, while the other expert says the mean weight of a bass is 4 pounds.

   You trust both experts equally well, and you believe one of them is right. There is an optimal Bayesian decision rule based on this information alone. What is the rule?