MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Media Arts and Sciences Section

Course 4.902: Pattern Recognition and Analysis — Assigned: 2/9/93

Problem Set No. 2 — Due: 2/16/93

READING ASSIGNMENT: Therrien, Chapters 3 and 4.

This problem set consists of six warm-up exercises and four problems. Remember that you are *not* allowed to look at old problem sets. Collaboration is allowed and encouraged.

Exercise 1: (5 points)

Let $\mathbf{A}$ be an $m \times n$ matrix with real elements a_{ij}, and let $\mathbf{B} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$.

(a) Show that $\mathbf{B}$ is symmetric.

(b) Show that $\mathbf{B}$ is positive semidefinite.

Exercise 2: (5 points)

In Problem 1.2 (p. 11) in the text, the following two sets of data are presented, corresponding to classes ω_1 and ω_2:

Class ω_1

1.21	2.07	3.00	3.31	3.78	4.30	4.97	6.58
1.32	2.21	3.11	3.37	3.90	4.70	5.10	7.00
1.40	2.22	3.12	3.45	3.97	4.86	5.70	
1.56	2.73	3.21	3.50	4.12	4.92	6.21	

Class ω_2

6.89	9.33	10.45	11.23	11.45	11.99	12.37	14.60
8.01	9.76	10.56	11.25	11.60	12.22	13.10	
8.76	10.03	10.72	11.34	11.71	12.32	13.50	
9.25	10.31	10.80	11.37	11.82	12.33	13.57	

The text asks you to draw histograms for these two sets (it is not necessary to do so here), and to choose a decision boundary, which we can call τ.

a) By inspection, how many classification errors result if $\tau = 6.7$? If $\tau = 6.9$? If $\tau = 7.1$?

b) Based on your answer to part (a), would you say that minimizing the number of errors in a set of sample data is a good way to fine-tune the decision boundary? Why or why not? Limit your answer to one or two sentences.

In Exercises 3-5, complete the given statement with the choice that results in the strongest true statement. Justify each answer in a line or two. (Statement A is "stronger" than statement B if A implies B. For instance, both statements "a covariance matrix is symmetric" and "a covariance matrix is square" are true, but the first is stronger than the second.)

In answering the questions, remember that a positive definite matrix is also positive semidefinite.

Exercise 3: (5 points)

Suppose that all of the eigenvalues of a particular symmetric matrix are nonzero. Then the matrix is ______.

a) positive semidefinite
b) nonsingular
c) square
d) a covariance matrix
e) none of the above necessarily follows

Exercise 4: (5 points)

Suppose that all of the eigenvalues of a particular symmetric matrix are positive. Then the matrix is ______.

a) positive semidefinite
b) nonsingular
c) square
d) a covariance matrix
e) none of the above necessarily follows

Exercise 5: (5 points)

Suppose that all of the diagonal elements in a symmetric matrix are positive. Then the matrix is ______.

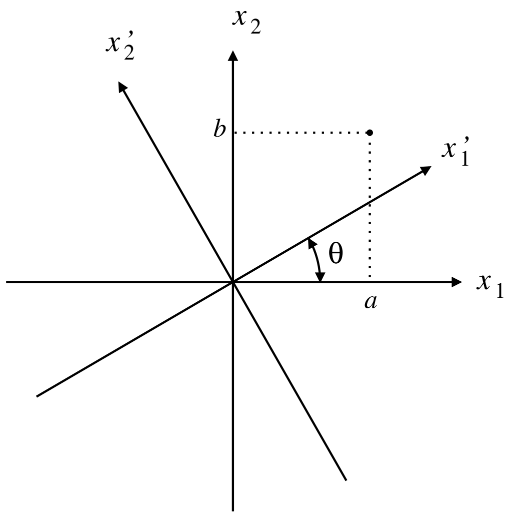
a) positive semidefinite
b) nonsingular
c) square
d) a covariance matrix
e) none of the above necessarily follows

Exercise 6: (5 points)

Show that the correlation matrix $\mathbf{R}$ and the covariance matrix $\mathbf{K}$ are always positive semidefinite.

Problem 1: (15 points)

In the diagram below, two coordinate systems are shown. They share the same origin, but the primed coordinate system is rotated by an angle of θ radians relative to the unprimed system.



A point is shown having coordinate vector

$$\mathbf{p} = \begin{bmatrix} a \\ b \end{bmatrix}$$

relative to the unprimed coordinate system.

a) Find the coordinate vector $\mathbf{p}'$ of the same point with respect to the primed system. Express your answer in the form $\mathbf{p}' = \mathbf{A}\mathbf{p}$.

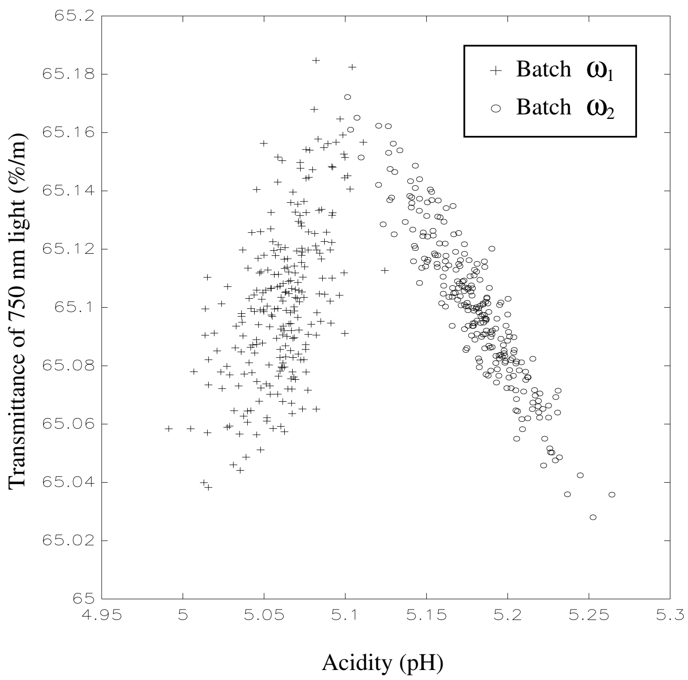
b) Let $\mathbf{x}$ be a 2-dimensional random vector with covariance matrix $\mathbf{K_x}$, and let $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is the matrix you found in part (a). Find $\mathbf{K_y}$, the covariance matrix of $\mathbf{y}$, and also find $\det(\mathbf{K_y})$ in terms of $\det(\mathbf{K_x})$.

c) Show that $\mathrm{tr}(\mathbf{K_y}) = \mathrm{tr}(\mathbf{K_x})$ by explicitly computing the diagonal elements of $\mathbf{K_y}$.

Problem 2: (15 points)

A pharmaceutical plant has discovered that a particular crate of analgesic pills in its inventory has not been labeled with its batch number. If the batch number can be identified with probability of error no greater than 10^{-4}, the plant will label the crate and ship it. Otherwise, it will destroy the crate.

It is believed that the crate came from either of two batches, ω_1 and ω_2, with equal probability for each. To provide a basis for classification, data are collected in the following way. A correctly labeled bottle of 250 pills is obtained from each of the two batches. Each pill from each of the two bottles is dissolved in 100 ml distilled water, and the resulting solution tested for acidity and transmittance of infrared light. Because of slight variations in the manufacturing process, there are statistical differences between the two batches in these features, as can be seen in the scatter plot below. Note that there are 250 points for each batch.



a) Suppose the following classification rule is used. First, the feature space is divided into two regions, R_1 and R_2, by a straight line. A single pill from the unclassified crate is then dissolved and tested. If it falls into R_1, the entire crate is labeled ω_1; otherwise, it is labeled ω_2. To minimize the probability of classification error, what should the equation of the line be (approximately)? Express your answer in the form $y = ax + b$, where y is transmittance and x is pH.

b) Using the test described in part (a), do you think the goal of 10^{-4} error probability will be met? Justify your answer. How might the test be modified to reduce the probability of error? Hint: a whole bottle had to be opened to get the pill used in part (a). Limit your answer to one-half page.

c) Now suppose that there are sufficent resources to allow either IR transmittance measurements or acidity measurements, but not both. Which would you choose, and why? Describe a strategy for meeting the 10^{-4} error probability goal, based on this one type of measurement.

Problem 3: (20 points)

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

$$p_{y|\omega_1}(\hat{y}|\omega_1) = k_1 \exp(-\hat{y}^2/6).$$

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 $C_{21} = \sqrt{3}$, 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.

Problem 4: (20 points)

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?

Puzzle: (0 points, but fun if you haven't seen it before)

This problem is completely optional, and will not count in any way to your grade.

The Petersburg Paradox

Consider the following game, in which there are two sides: bank and player. A fair coin is flipped until it comes up heads. Let n denote the number of times it was flipped. For instance, if the sequence is TH then $n = 2$ and if the sequence is $TTTTH$ then $n = 5$. After n is determined, the bank pays the player 2^n dollars.

A game is *fair* when the expected value of either side's gain (or loss) is zero. How much should the player pay the bank as an entry fee to make the game fair? Does your answer strike you as paradoxical?

A legitimate objection to the foregoing scenario is that the bank does not have unlimited resources with which to honor its debt. Suppose that the bank is unable to pay out more than 2^M dollars, where M is a fixed integer. That is, the bank will pay 2^n dollars if $n \leq M$, but only 2^M dollars if $n > M$. Under this new condition, what is the expected value of the player's winnings? If the player pays an entry fee equal to his expected winnings (to make the game fair), what is the probability that the player will come away with a profit, as a function of M?