Homework 2

CS 591Q/791V - Pattern Recognition
Instructor: Dr. Arun Ross
Due Date: March 4, 2010

Note: You are permitted to discuss the following questions with others in the class. However, you *must* write up your *own* solutions to these questions. Any indication to the contrary will be considered an act of academic dishonesty. Code developed as part of this assignment should be placed in a zip file and sent to arun.ross at mail.wvu.edu with the subject line "CS 591Q/791V : Homework 2".

---

1. [5 points] Consider the three-dimensional normal distribution $p(\boldsymbol{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (1, 2, 2)^t$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix}$. What is $p[(0.5, 0, 1)^t]$.

2. [15 points] Consider the following density function for feature vector $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$:

$$p(\boldsymbol{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}; \sigma_{12} = \sigma_{21}.$$

   (a) Write down the expression for the Euclidean distance between an arbitrary vector $\boldsymbol{x}$ and mean vector $\boldsymbol{\mu}$.

   (b) Write down the expression for the Mahalanobis distance between vector $\boldsymbol{x}$ and mean vector $\boldsymbol{\mu}$. Simplify this expression by expanding the quadratic term.

   (c) Compare the expressions in (2a) and (2b). How does the Mahalanobis distance differ from the Euclidean distance? When are the two distances equal? When is it more appropriate to use the Mahalanobis distance?

3. [15 points] Consider the following class-conditional densities for a three-class problem involving two-dimensional features:

$$p(\boldsymbol{x}|\omega_1) \sim N\left((0,0)^t, I\right);$$
$$p(\boldsymbol{x}|\omega_2) \sim N\left((1,1)^t, I\right);$$
$$p(\boldsymbol{x}|\omega_3) \sim \frac{1}{2}N\left((0.5,0.5)^t, I\right) + \frac{1}{2}N\left((-0.5,0.5)^t, I\right).$$

By explicit calculation of the *posterior* probabilities, classify the two-dimensional point $\boldsymbol{x} = (0.3, 0.3)^t$ based on the Bayes decision rule (assuming a 0-1 loss function and equal priors).

Note: For this problem, the *exact* values for the three posterior probabilities have to be computed prior to classifying the point.

4. [20 points] Consider a two-category classification problem involving one feature, x. Let both class-conditional densities conform to a Cauchy distribution as follows:

$$p(x \mid \omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a_i}{b}\right)^2}, \quad i = 1, 2.$$

Assume a 0-1 loss function and equal priors.

(a) Compute the Bayes decision boundary.

(b) Show that the probability of misclassification according to the Bayes decision rule is

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|.$$

(c) Plot $P(error)$ as a function of $|a_2 - a_1|/b$.

(d) What is the maximum value of $P(error)$. Under what conditions will this occur?

5. [25 points] Consider a two-category classification problem involving two-dimensional feature vectors of the form $\boldsymbol{x} = (x_1, x_2)^t$. The two categories are ω_1 and ω_2, and

$$p(\boldsymbol{x} \mid \omega_1) \sim N\left((0,0)^t, I\right),$$
$$p(\boldsymbol{x} \mid \omega_2) \sim N\left((1,1)^t, I\right),$$
$$P(\omega_1) = P(\omega_2) = \frac{1}{2}.$$

(a) Calculate the Bayes decision boundary and write down the Bayes decision rule assuming a 0-1 loss function.

(b) What are the Bhattacharyya and Chernoff bounds on the probability of misclassification, P(error)?

(c) Generate $n = 25$ patterns from *each* of the two class-conditional densities and plot them in a two-dimensional feature space using different markers for the two categories. Draw the Bayes decision boundary computed in (a) on this plot for visualization purposes. What is the empirical error rate when classifying the generated patterns using the Bayes decision rule.

(d) Repeat (c) above by generating $n = 10,000$ patterns for each of the two categories.

(e) Is it possible for the empirical error rate in (c) and (d) to exceed the theoretically derived error bounds in (b)? If so, under what condition is that likely to occur?

CS 791V students will have to solve the following problems in addition to the ones above

1. [15 points] In many pattern classification problems, the classifier is allowed to reject an input pattern by not assigning it to any one of the c classes. If the cost of rejection is not too high, it may be a desirable action in some cases. Let

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0, & i = j \quad (i, j = 1, \ldots c) \\ \lambda_r, & i = c + 1 \\ \lambda_s, & i \neq j \quad (i, j = 1, \ldots c), \end{cases}$$

where λ_r is the loss incurred for rejecting an input pattern and λ_s is the loss incurred for misclassifying the input pattern (known as substitution error).

   (a) Show that the minimum risk rule results in the the following decision policy.
   Assign pattern $\boldsymbol{x}$ to class ω_i if $P(\omega_i \mid \boldsymbol{x}) \geq P(\omega_j \mid \boldsymbol{x}) \forall j$ **and** $P(\omega_i \mid \boldsymbol{x}) \geq 1 - \lambda_r/\lambda_s$, else reject it.

   (b) Explain what happens if $\lambda_r = 0$? Also, explain what happens if $\lambda_r > \lambda_s$?

2. [10 points] Let $p(x \mid \omega_i) \sim N(\mu_i, \sigma^2)$ for a two-class one-dimensional problem with $P(\omega_1) = P(\omega_2) = 0.5$. Show that the minimum error classification rule results in the following probability of misclassification:

$$P(error) = \frac{1}{\sqrt{(2\pi)}} \int_a^\infty e^{-u^2/2} du,$$

where $a = \mid \mu_2 - \mu_1 \mid / (2\sigma)$.
