ECSE-6610: PR Homework Set 1

Chengjiang Long

February 6, 2018

Assigned Date: Feb 2, 2018.
Due Date: Feb 9, 2018.
Collaboration Policy. Homeworks will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited.
Late Policy. No late submissions will be allowed without consent from the instructor. If urgent or unusual circumstances prohibit you from submitting a homework assignment in time, please e-mail me explaining the situation.
Submission Format. Electronic submission of a zip file is mandatory. Include code in your pdf file as needed to make your answers clear. Submit all code separately.

Problem 1 (10 points) Let $\omega_{max}(\mathbf{x})$ be the state of nature for which $P(\omega_{max}|\mathbf{x}) \geq P(\omega_i|\mathbf{x})$ for all i, $i = 1, \ldots, c$.

(a) [**2 points**] Show that $P(\omega_{max}|\mathbf{x}) \geq \frac{1}{c}$.

(b) [**2 points**] Show that for the minimum-error-rate decision rule the average probability of error is given by

$$P(error) = 1 - \int P(\omega_{max}|\mathbf{x})P(\mathbf{x})d\mathbf{x} \tag{1}$$

(c) [**3 points**] Use these two results to show that $P(error) \leq \frac{c-1}{c}$.

(d) [**3 points**] Describe a situation for which $P(error) = \frac{c-1}{c}$.

Problem 2 (20 points) In many pattern classification problems one has the option either to assign the pattern to one of c classes, or to reject it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$\lambda(a_i|\omega_j) = \begin{cases} 0 & i = j, j = j = 1, ..., c \\ \lambda_r & i = c+1 \\ \lambda_s & otherwise \end{cases} \tag{2}$$

where λ_r is the loss incurred for choosing the (c + 1)-th action, rejection, and λ_s is the loss incurred for making any substituting error. Show that the minimizing risk is obtained if we decide ω_i if $P(\omega_i|\mathbf{x}) \geq P(\omega_j|\mathbf{x})$ for all j and if $P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$, and reject otherwise.

(a) [**10 points**] What happens if $\lambda_r = 0$?

(b) [**10 points**] What happens if $\lambda_r > \lambda_s$?

Hints: The input pattern x still belongs to one of the c classes and the posteriors $P(\omega_i|\mathbf{x})$ obey the law of total probability. if one classifies the input as belonging to class $i \neq c+1$, what is the risk of rejection?

Problem 3 (20 points) Consider a two-category classification problem in two dimensions with $p(\mathbf{x}|\omega_1) \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{I})$, $p(\mathbf{x}|\omega_2) \sim N(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{I})$, and $P(\omega_1) = P(\omega_2) = 0.5$.

(a) [**5 points**] Calculate the Bayes decision boundary.

(b) [**5 points**] Calculate the Bhattacharyya error bound.

(c) [**10 points**] Repeat the above for the same prior probabilities, but $p(\mathbf{x}|\omega_1) \sim N(\mathbf{0}, \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix})$ and $p(\mathbf{x}|\omega_1) \sim N(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix})$.

Problem 4 (10 points) Let x be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

$$P(\mathbf{x}|\theta) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i}, \tag{3}$$

where $\theta = (\theta_1, \ldots, \theta_d)^T$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum likelihood estimate for θ is

$$\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k. \tag{4}$$

Problem 5 (40 points + Extra 20 points) Generate 10,000 samples from each 2D Gaussian distribution specified by the following parameters:

$$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 6 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \tag{5}$$

Please, note that this is not the same as sampling the functions shown above; see "Generating Gaussian Random Numbers" (Link: https://www.taygeta.com/random/gaussian.html) for more information on how to generate the samples using the Box-Muller transformation. A link to C code (Link: ftp://ftp.taygeta.com/pub/c/boxmuller.c) has been provided. Since the code generates samples for 1D distributions, you would need to call the function twice to get a 2D sample (x, y); use the x-mean, x-variance for the x sample and the y-mean, y-variance for the y sample.

(a) [**10 points**] Assuming $P(\omega_1) = P(\omega_2)$, design and implement a Max Likelihood classifier and a naive Bayes classifier. Use the first 80% for training, and the rest of 20% as testing. Report the classification accuracy for each class separately and the overall accuracy at both the training and testing stages.

(b) [**10 points**] Repeat part (a) for $P(\omega_1) = 0.2$ and $P(\omega_2) = 0.8$.

(c) [**20 points**] Repeat the above (a) using $P(\omega_1) = 0.2$, $P(\omega_2) = 0.8$ and the following parameters:

$$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 6 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix} \tag{6}$$

(d)[**Extra 20 points**] Plot ROC curves to evaluate the performance of the classifiers used in above (c).

Hints: You can use $P(\omega_2|\mathbf{x})$ as the scores to draw the ROC curve, assuming that you consider ω_2 as positive class. You also can use $P(\mathbf{x}|\omega_2)/P(\mathbf{x}|\omega_1)$ as the scores to draw the ROC curve.