Homework 1

Shao-Ting Chiu (UIN:433002162)

9/17/22

Homework Description

Problems (from Chapter 2 in the book): 2.1 , 2.3 (a,b), 2.4, 2.7, 2.9, 2.17 (a,b)

Note: the book is available electronically on the Evans library website.

- Deadline: `Sept. 26th, 11:59 pm`

Problem 2.1

Suppose that X is a discrete feature vector, with distribution concentrated over a countable set $D = \{x^1, x^2, \dots\}$ in R^d. Derive the discrete versions of (2.3), (2.4), (2.8), (2.9), (2.11), (2.30), (2.34), and (2.36)

Hint: Note that if X has a discrete distribution, then integration becomes summation, $P(X = x_k)$, for $x_k \in D$, play the role of $p(x)$, and $P(X = x_k|Y = y)$, for $x_k \in D$, play the role of $p(x|Y = y)$, for $y = 0, 1$.

Problem 2.3

This problem seeks to characterize the case $\epsilon^* = 0$.

(a)

Prove the "Zero-One Law" for perfect discrimination:

$$\epsilon^* = 0 \Leftrightarrow \eta(X) = 0 \text{ or } 1 \quad \text{with probability 1.} \tag{1}$$

(b)

Show that

$$\epsilon^* = 0 \Leftrightarrow \text{ there is a function } f \text{ s.t. } Y = f(X) \text{with probability 1}$$

Problem 2.4

This problem concerns the extension to the multiple-class case of some of the concepts derived in this chapter. Let $Y \in \{0, 1, \dots, c-1\}$, where c is the number of classes, and let

$$\eta_i(x) = P(Y = i | X = x), \quad i = 0, 1, \dots, c-1,$$

for each $x \in R^d$. We need to remember that these probabilities are not indpendent, but satisfy $\eta_0(x) + \eta_1(x) + \cdots + \eta_{c-1}(x) = 1$, for each $x \in R^d$, so that one of the functions is redundant. In the two-class case, this is made explicit by using a single $\eta(x)$, but using the redundant set above proves advantageous in the multiple-class case, as seen below.

Hint: you should answer the following items in sequence, using the previous answers in the solution of the following ones

(a)

Given a classifier $\psi : R^d \to \{0, 1, \dots, c-1\}$, show that its conditional error $P(\psi(X) \neq Y | X = x)$ is given by

$$P(\psi(X) \neq Y | X = x) = 1 - \sum_{i=1}^{c-1} I_{\psi(x)=i} \eta_i(x) = 1 - \eta_{\psi(x)}(x)$$

(b)

Assuming that X has a density, show that the classification error of ψ is given by

$$\epsilon = 1 - \sum_{i=0}^{c-1} \int_{\{x | \psi(x) = i\}} \eta_i(x) p(x) dx$$

.

(c)

Prove that the Bayes classifier is given by

$$\psi^*(x) = \arg \max_{i=0,1,\ldots,c-1} \eta_i(x), \quad x \in R^d$$

Hint: Start by considering the difference between conditional expected errors $P(\psi(X) \neq Y | X = x) - P(\psi^*(X) \neq Y | X = x)$.

(d)

Show that the Bayes error is given by

$$\epsilon^* = 1 - E[\max_{i=0,1,\ldots,c-1} \eta_i(X)]$$

(e)

Show that the maximum Bayes error possible is $1 - \frac{1}{c}$.

Problem 2.7

Consider the following univariate Gaussian class-conditional densities:

$$p(x|Y=0) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{(x-3)^2}{2})$$

$$p(x|Y=1) = \frac{1}{3\sqrt{2\pi}} \exp(-\frac{(x-4)^2}{18})$$

Assume that the classes are equally likely, i.e., $P(Y=0) = P(Y=1) = \frac{1}{2}$

(a)

Draw the densities and determine the Bayes classifier graphically.

(b)

Determine the Bayes classifier.

(c)

Determine the specificity and sensitivity of the Bayes classifier.

Hint: use the standard Gaussian CDF $\psi(x)$

(d)

Determine the overall Bayes error.

Problem 2.9

Obtain the optimal decision boundary in the Gaussian model with $P(Y=0) = P(Y=1)$ and

In each case draw the optimal decision boundary, along with the class means and class conditional density contours, indicating the 0- and 1-decision regions.

(a)

$$\mu_0 = (0,0)^T, \mu_1 = (2,0)^T, \Sigma_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

(b)

$$\mu_0 = (0,0)^T, \mu_1 = (2,0)^T, \Sigma_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$$

(c)

$$\mu_0 = (0,0)^T, \mu_1 = (0,0)^T, \Sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

(d)

$$\mu_0 = (0,0)^T, \mu_1 = (0,0)^T, \Sigma_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

Python Assignment: Problem 2.17

This problem concerns the Gaussian model for synthetic data generation in Braga-Neto (2020, sec. A8.1).

(a)

Derive a general expression for the Bayes error for the homoskedastic case with $\mu_0 = (0, \dots, 0), \mu_1 = (1, \dots, 1)$, and $P(Y = 0) = P(Y = 1)$. Your answer should be in terms of $k, \sigma_1^2, \dots, \sigma_k^2, l_1, \dots, l_k$, and $\sigma_1, \dots, \sigma_k$.

Hint: Use the fact that

$$\begin{bmatrix} 1 & \sigma & \cdots & \sigma \\ \sigma & 1 & \cdots & \sigma \\ \vdots & \vdots & \ddots & \vdots \\ \sigma & \sigma & \cdots & 1 \end{bmatrix}_{l \times l}^{-1} = \frac{1}{(1-\sigma)(1+(l-1)\sigma)} \tag{2}$$

(b)

References

Braga-Neto, Ulisses. 2020. *Fundamentals of Pattern Recognition and Machine Learning.* Springer.