Ejercicios de “Pattern Classification”, Duda & Hart

Sebastián Sampayo

Primer Cuatrimestre de 2014

Índice

Parte I
[Capítulo 2]

1. Ejercicio 1

In the two-category case, under the Bayes decision rule the conditional error is given by Eq. 7. Even if the posterior densities are continuous, this form of the conditional error virtually always leads to a discontinuous integrand when calculating the full error by Eq. 5.

1.1. (a)

Show that for arbitrary densities, we can replace Eq. 7 by $P(error|x) = 2P(\omega_1|x)P(\omega_2|x)$ *in the integral and get an upper bound on the full error.*

Solución

$$Eq.\,7: \;\; P(\varepsilon|x) = \text{mín}\left\{P(\omega_1|x), P(\omega_2|x)\right\} \tag{1}$$

$$Eq.\,5: \;\; P(\varepsilon) = \int_{-\infty}^{+\infty} P(\epsilon, x)dx = \int_{-\infty}^{+\infty} P(\varepsilon|x)p(x)dx \tag{2}$$

Si $P(\omega_1|x) > P(\omega_2|x)$ entonces,

$$P(\varepsilon) = \int_{-\infty}^{+\infty} P(\omega_2|x)p(x)dx$$

Si $P(\omega_1|x) < P(\omega_2|x)$ entonces,

$$P(\varepsilon) = \int_{-\infty}^{+\infty} P(\omega_1|x)p(x)dx$$

Se propone $P(\varepsilon|x) = 2P(\omega_1|x)P(\omega_2|x)$. Se quiere mostrar que:

$$P(\varepsilon) = \int_{-\infty}^{+\infty} \text{mín}\left\{P(\omega_1|x), P(\omega_2|x)\right\} p(x)dx \; \leq \int_{-\infty}^{+\infty} 2P(\omega_1|x)P(\omega_2|x)p(x)dx$$

Si se demuestra que para todo x,

$$2P(\omega_1|x)P(\omega_2|x) \geq \text{mín}\left\{P(\omega_1|x), P(\omega_2|x)\right\}$$

entonces queda demostrado lo enunciado.

Defino: $\text{mín}\left\{P(\omega_1|x), P(\omega_2|x)\right\} \triangleq \text{mín}$, para facilitar la lectura.
Teniendo en cuenta que $P(\omega_1|x) + P(\omega_2|x) = 1$, entonces

$$\text{mín} < \frac{1}{2} \qquad y \qquad (1 - \text{mín}) > \frac{1}{2} \tag{3}$$

pues de otra manera no sería *el mínimo*.
Por otro lado:

$$2P(\omega_1|x)P(\omega_2|x) = 2\,(1 - \text{mín})\,\text{mín} \geq \text{mín}$$

Si $mín \neq 0$,

$$2\,(1 - mín) \geq 1$$

$$1 \geq 2mín$$

$$\boxed{\frac{1}{2} \geq mín} \tag{4}$$

que se cumple siempre por lo mencionado en la ecuación (3).
Si $mín = 0$, la inecuación se cumple igual, sin necesidad de operar sobre ella.

1.2. (b)

Show that if we use $P(error|x) = \alpha P(\omega_1|x)P(\omega_2|x)$ for $\alpha < 2$, then we are not guaranteed that the integral gives an upper bound on the error.

Solución
En este caso se tiene:

$$\alpha\,(1 - mín)\,mín \geq mín$$

Si $mín \neq 0$,

$$\alpha - 1 \geq \alpha mín$$

$$\boxed{1 \geq mín + \frac{1}{\alpha}} \tag{5}$$

⇒Para $\alpha < 2$ no se asegura la validez de la inecuación. Esto se ve en los casos límites más claramente:
Si $mín = \frac{1}{2}$ y $\alpha = 2$, reemplazando,

$$1 \geq \frac{1}{2} + \frac{1}{2} = 1$$

entonces, se cumple la inecuación.
Si $mín = \frac{1}{2}$ y $\alpha < 2$, reemplazando,

$$1 \geq \frac{1}{2} + \frac{1}{1{,}999...} = 1{,}00..{,}001$$

entonces: no se cumple la inecuación.
NOTA: Igualmente, puede darse el caso de $\alpha < 2$ con un $mín = m$ tal que se cumpla igual la inecuación. Ejemplo:
Si $\alpha = \frac{3}{2}$ y $mín \leq \frac{1}{3}$, reemplazando

$$1 \geq \frac{1}{3{,}00..{,}01} + \frac{2}{3} = 0{,}999...$$

⇒Se cumple.

1.3. (c)

Analogously, show that we can use instead $P(error|x) = P(\omega_1|x)P(\omega_2|x)$ and get a lower bound on the full error.

Solución
Ahora se pide mostrar que:

$$P(\varepsilon) = \int_{-\infty}^{+\infty} mín\left\{P(\omega_1|x), P(\omega_2|x)\right\} p(x)dx \geq \int_{-\infty}^{+\infty} P(\omega_1|x)P(\omega_2|x)p(x)dx$$

Por analogía con el ejercicio (a), se buscará:

$$(1 - mín)\, mín \leq mín$$

Si $mín \neq 0$,

$$1 \leq 1 + mín$$
$$\boxed{0 \leq mín} \tag{6}$$

lo cual se cumple siempre por ser una probabilidad.

1.4. (d)

Show that if we use $P(error|x) = \beta P(\omega_1|x)P(\omega_2|x)$ for $\beta > 1$, then we are not guaranteed that the integral gives an lower bound on the error.

Solución
De manera similar se llega a:

$$\beta\,(1 - mín)\, mín \leq mín$$

Operando se obtiene:

$$\boxed{1 - \frac{1}{\beta} \leq mín} \tag{7}$$

lo cual puede no tiene razón para cumplirse siempre.

2. Ejercicio 2

Suppose two equally probable one-dimensional densities are of the form $p(x|\omega_i) \propto e^{-|x-a_i|/b_i}$ for $i = 1, 2$ and $0 < b_i$.

2.1. (a)

Write an analytic expression for each density, that is, normaliz each function for arbitrary a_i and b_i.

Solución
Se trata de encontrar una constante K tal que:

$$\int_{-\infty}^{+\infty} K e^{-\frac{|x-a_i|}{b_i}}\, dx = 1$$

Dada la simetría de la función, esta integral es equivalente a pedir:

$$2K \int_{0}^{+\infty} e^{-\frac{x}{b_i}}\, dx = 1$$

$$\int_{0}^{+\infty} e^{-\frac{x}{b_i}}\, dx = \frac{1}{2K} = b_i$$

Por lo tanto:

$$K = \frac{1}{2b_i}$$

$$\boxed{p(x|\omega_i) = \frac{1}{2b_i} e^{-\frac{|x-a_i|}{b_i}}}$$

2.2. (b)

Calculate the likelihood ratio as a function of your four variables.

Solución

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} \underset{\alpha_2}{\overset{\alpha_1}{\gtrless}} 1$$

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{\frac{1}{2b_1} e^{-\frac{|x-a_1|}{b_1}}}{\frac{1}{2b_2} e^{-\frac{|x-a_2|}{b_2}}}$$

$$\boxed{\frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{b_2}{b_1} e^{\frac{|x-a_2|}{b_2} - \frac{|x-a_1|}{b_1}}}$$

2.3. (c)

Sketch a graph of the likelihood ratio $p(x|\omega_1)/p(x|\omega_2)$ for the case $a_1 = 0$, $b_1 = 1$, $a_2 = 1$ and $b_2 = 2$.

Solución

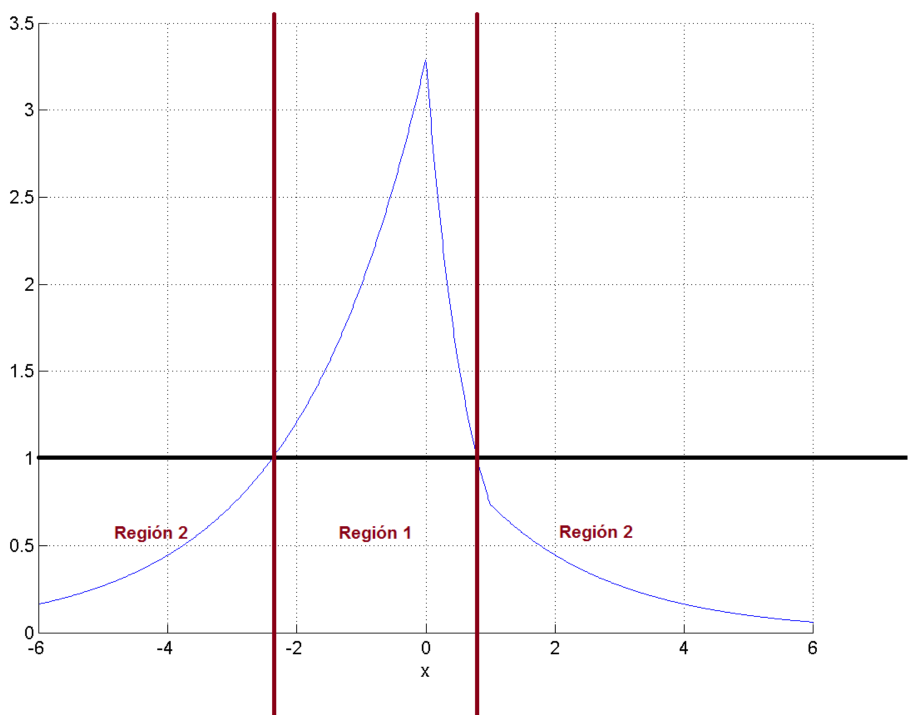

Figura 1: Ejercicio 2 (c)

3. Ejercicio 11

Suppose that we replace the deterministic decision function $\alpha(x)$ with a randomized rule, namely, one giving the probability $P(\alpha_i|x)$ of taking action α_i upon observing x.

3.1. (a)

Show that the resulting risk is given by

$$R = \int \left[\sum_{i=1}^{a} R(\alpha_i|x) P(\alpha_i|x) \right] p(x)dx$$

Solución
Normalmente, en el caso en que $\alpha(x)$ es determinística, el riesgo total está dado por:

$$R = \int R(\alpha|x)p(x)dx$$

Sin embargo, en el caso de la decisión aleatoria, $R(\alpha|x)$ no está determinada, por lo tanto se toma el promedio:

$$R(\alpha|x) = \sum_{i=1}^{a} R(\alpha_i|x) P(\alpha_i|x)$$

3.2. (b)

In addition, show that R is minimized by choosing $P(\alpha_i|x) = 1$ for the action α_i associated with the minimum conditional risk $R(\alpha_i|x)$, thereby showing that no benefit can be gained from randomized the best decision rule.

Solución
Si se demuestra que se minimiza $R(\alpha|x)$ para todo x, entonces la integral R queda minimizada.
En el caso aleatorio, como se vio anteriormente,

$$R(\alpha|x) = \sum_{i=1}^{a} R(\alpha_i|x) P(\alpha_i|x)$$

esto es, se tiene un promedio ponderado de todos los $R(\alpha_i|x)$.
Por el contrario en el caso determinístico:

$$R(\alpha|x) = mín\{R(\alpha_i|x)\}$$

Como el promedio de cualquier conjunto numérico nunca puede ser menor al mínimo de dicho conjunto, entonces se concluye que el caso determinístico minimiza el riesgo.

4. Ejercicio 12

Let $\omega_{max}(x)$ be the state of nature for wich $P(\omega_{max}|x) \geq P(\omega_i|x)$ for all i, i=1,...,c.

4.1. (a)

Show that $P(\omega_{max}|x) \geq 1/c$.

Solución
El enunciado del problema establece que:

$$P(\omega_{max}|x) = máx\{P(\omega_i|x)\}$$

Por otro lado, el promedio del conjunto $\{P(\omega_i|x)\}$ es:

$$\frac{1}{c} \cdot \sum_{i=1}^{c} P(\omega_i|x) = \frac{1}{c} \cdot 1$$

Por lo tanto, por ser el máximo de un conjunto, siempre será mayor o igual (en caso de que seán todos los valores del conjunto iguales), que el promedio de dicho conjunto.

$$\boxed{P(\omega_{max}|x) \geq \frac{1}{c}}$$

4.2. (b)

Show that for the minimum-error-rate decision rule the average probability of error is given by

$$P(error) = 1 - \int P(\omega_{max}|x)p(x)dx$$

Solución
La regla de decisión que minimiza el error es elegir la clase que tiene mayor probabilidad *a posteriori*, i.e., $\omega_{max}(x)$. Por lo tanto, la probabilidad de acertar es:

$$P(acierto) = \int P(\omega_{max}|x)p(x)dx$$

La probabilidad de error puede escribirse como:

$$P(error) = 1 - P(acierto)$$

que es igual a:

$$\boxed{P(error) = 1 - \int P(\omega_{max}|x)p(x)dx}$$

4.3. (c)

Use these two results to show that $P(error) \leq (c-1)/c$.

Solución
Dado que

$$P(\omega_{max}|x) \geq \frac{1}{c}$$

entonces

$$P(acierto) = \int P(\omega_{max}|x)p(x)dx \geq \frac{1}{c} \cdot \int p(x)dx = \frac{1}{c}$$

Luego, con el mínimo de la probabilidad de acierto, se obtiene una cota del máximo de la probabilidad de error:

$$P(error) \leq 1 - \frac{1}{c}$$

$$\boxed{P(error) \leq \frac{c-1}{c}}$$

4.4. (d)

Describe a situation for wich $P(error) = (c-1)/c$.

Solución

Para que se de la igualdad, se debe cumplir que $P(acierto) = 1/c$ y por lo tanto una posibilidad es que se cumpla que:

$$P(\omega_{max}|x) = \frac{1}{c}$$

en cuyo caso, todas las probabilidades *a posteriori* son iguales, en consecuencia no se tiene preferencia probabilística por ninguna clase.

De esta forma:

$$P(error) = 1 - P(acierto) = 1 - \frac{1}{c}$$

$$\boxed{P(error) = \frac{c-1}{c}}$$

5. Ejercicio 13

In many pattern classification problems one has the option either to assingn the pattern to one of c classes, or to reject ir as being unrecognizable. If the cost for reject is not too high, rejection may be a desirable action. Let

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

where λ_r is the loxx incurred for choosing the (c+1)th action, rejection, and λ_s is the loss incurred for making any substitution error. Show that the minimum risk obtained if we decide ω_i if $P(\omega_i|x) \geq P(\omega_j x)$ for all j and if $P(\omega_i|x) \geq 1 - \lambda_r/\lambda_s$, and reject otherwise. What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?

Solución

Sea el riesgo de tomar la decisión α_i dada una observación x :

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|x)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|x) + \underbrace{\lambda(\alpha_i|\omega_i)}_{=0} P(\omega_i|x)$$

$$= \lambda_s [1 - P(\omega_i|x)]$$

Por otro lado, el riesgo de rechazar la observación está dado por

$$R(\alpha_{c+1}|x) = \lambda_r$$

De esta manera, se elegirá aceptar o rechazar la observación según cual presente menos riesgo:

$$R(\alpha_{c+1}|x) \underset{\alpha_{c+1}}{\overset{\alpha_i}{\gtrless}} R(\alpha_i|x)$$

$$\lambda_r \underset{\alpha_{c+1}}{\overset{\alpha_i}{\gtrless}} \lambda_s[1 - P(\omega_i|x)]$$

$$\boxed{P(\omega_i|x) \underset{\alpha_{c+1}}{\overset{\alpha_i}{\gtrless}} 1 - \frac{\lambda_r}{\lambda_s}}$$

6. Ejercicio 14

Consider the classification problem with rejection option.

6.1. (a)

Use the results of Problem 13 to show that the following discriminant functions are optimal for such problems:

$$g_i(x) = \begin{cases} p(x|\omega_i)P(\omega_i) & i = 1, ..., c \\ \frac{\lambda_s - \lambda_r}{\lambda_s} \sum_{j=1}^{c} p(x|\omega_j)P(\omega_j) & i = c + 1 \end{cases}$$

Solución

Con el uso de una función discriminante ($g(x)$) se decide si aceptar o rechazar según que $g(x)$ sea mayor. Esto es, rechazo si:

$$g_{c+1}(x) > g_i(x)$$

Reemplazando con el enunciado se obtiene:

$$\frac{\lambda_s - \lambda_r}{\lambda_s} \underbrace{\sum_{j=1}^{c} p(x|\omega_j)P(\omega_j)}_{=p(x)} > p(x|\omega_i)P(\omega_i)$$

$$\frac{\lambda_s - \lambda_r}{\lambda_s} > \frac{p(x|\omega_i)P(\omega_i)}{p(x)} = P(\omega_i|x)$$

$$\boxed{1 - \frac{\lambda_r}{\lambda_s} > P(\omega_i|x)}$$

que como se vio en el ejercicio anterior es la regla de decisión que minimiza el riesgo tomado, por lo tanto la función discriminante $g(x)$ es óptima.

6.2. (b)

Plot these discriminant functions and the decision regions for the two-category one-dimensional case having

- $p(x|\omega_1) \sim N(1, 1)$,
- $p(x|\omega_2) \sim N(-1, 1)$,
- $P(\omega_1) = P(\omega_2) = 1/2$, and
- $\lambda_r/\lambda_s = 1/4$.

Solución

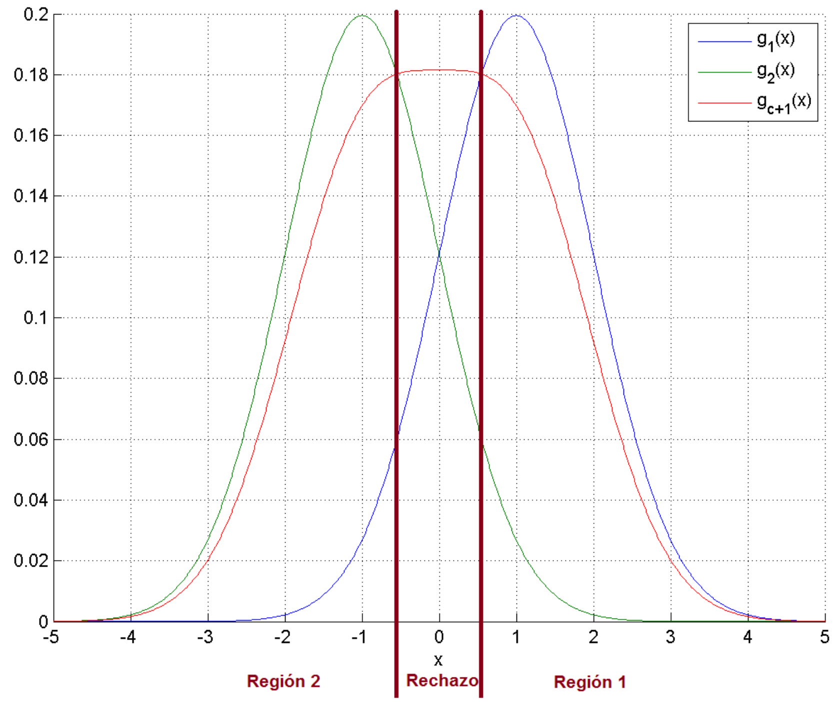

Figura 2: Regiones de decisión

6.3. (c)

Describe qualitatively what happens as λ_r/λ_s is increased from 0 to 1.

Solución

Si λ_r/λ_s tiende a 0, implica que el riesgo de rechazar es demasiado bajo, tendiendo a 0. Por lo tanto, si es 0, siempre tendrá menor riesgo en consecuencia se rechazará siempre.

Analizando el resultado del ejercicio 13, de aceptar siempre que:

$$P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

entonces si λ_r/λ_s tiende a 1,

$$P(\omega_i|x) \geq 0$$

para aceptar, lo cual se cumplirá siempre. En conclusión: siempre se acepta.

Por otro lado, puede verse fácilmente que para el caso de 2 clases, si $\lambda_r/\lambda_s = 1/2$, queda:

$$P(\omega_i|x) \geq \frac{1}{2}$$

Esto se dará siempre, por lo tanto se aceptará siempre. Además si λ_r/λ_s crece, $1 - \frac{\lambda_r}{\lambda_s}$ se achica. Finalmente, se puede decir que para el caso binario si $\frac{\lambda_r}{\lambda_s} \geq \frac{1}{2}$ siempre se aceptará.

Por extensión, se puede ver que en el caso de c clases, siempre se cumple que: $P(\omega_i|x) \geq \frac{1}{c}$, ya que en este caso $\omega_i = \omega_{máx}$ (ver ejercicio 12), por lo tanto si $\lambda_r/\lambda_s \leq 1 - 1/c$, siempre se aceptará.

6.4. (d)

Repeat for the case having

- $p(x|\omega_1) \sim N(1, 1)$,
- $p(x|\omega_2) \sim N(0, 1/4)$,
- $P(\omega_1) = 1/3,\ P(\omega_2) = 2/3$, and
- $\lambda_r/\lambda_s = 1/2$.

Solución

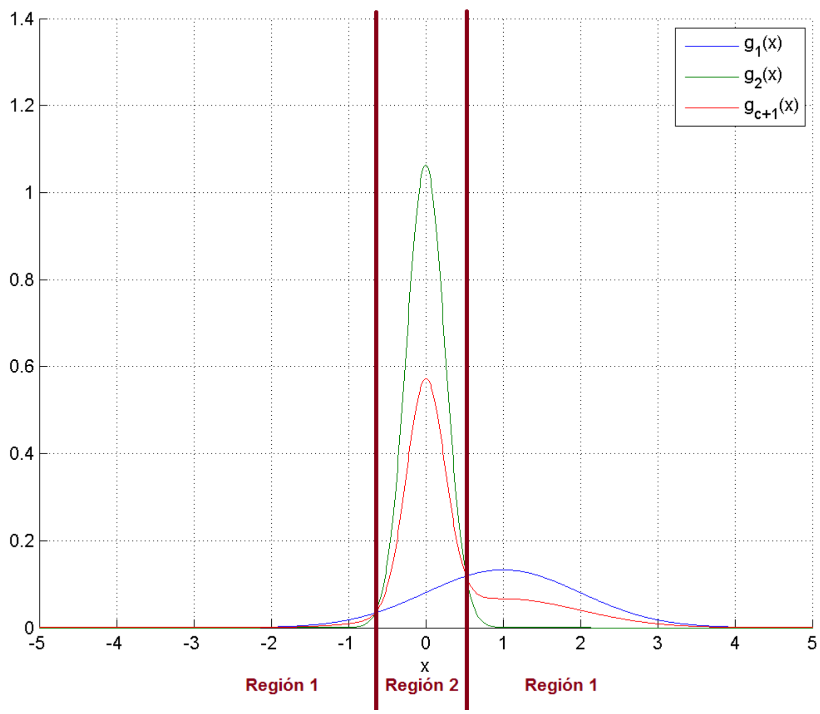

Figura 3: Regiones de decisión

Como se puede ver, no existe región de rechazo dado que la curva roja nunca supera a ninguna otra.

7. Ejercicio 25

Fill in the steps in the derivation from Eq. 59 to Eqs. 60-65.

Solución

$$Eq.\,59:\ g_i(x) = -\frac{1}{2}(x-\mu_i)^t\Sigma^{-1}(x-\mu_i) + ln\,P(\omega_i)$$

$$g_i(x) = -\frac{1}{2}(x^t\Sigma^{-1} - \mu_i^t\Sigma^{-1})(x-\mu_i) + ln\,P(\omega_i)$$

$$g_i(x) = -\frac{1}{2}(x^t\Sigma^{-1}x - x^t\Sigma^{-1}\mu_i - \mu_i^t\Sigma^{-1}x + \mu_i^t\Sigma^{-1}\mu_i) + ln\, P(\omega_i)$$

Como Σ es simétrica

$$g_i(x) = -\frac{1}{2}(x^t\Sigma^{-1}x - 2\mu_i^t\Sigma^{-1}x + \mu_i^t\Sigma^{-1}\mu_i) + ln\, P(\omega_i)$$

Además, como el término $x^t\Sigma^{-1}x$ es igual para todos los discriminantes, puede despreciarse.

$$g_i(x) = \mu_i^t\Sigma^{-1}x - \frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + ln\, P(\omega_i)$$

$$\boxed{Eq.\,60:\ g_i(x) = w_i^t x + \omega_{i0}}$$

$$\boxed{Eq.\,61:\ w_i = \Sigma^{-1}\mu_i}$$

$$\boxed{Eq.\,62:\ \omega_{i0} = -\frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + ln\, P(\omega_i)}$$

Luego, para obtener el límite entre las regiones de decisión se busca que las funciones discriminantes sean iguales:

$$g_i(x) = g_j(x)$$

$$w_i^t x + \omega_{i0} = w_j^t x + \omega_{j0}$$

$$(w_i^t - w_j^t)x + \omega_{i0} - \omega_{j0} = 0$$

$$(w_i^t - w_j^t)\left(x + \frac{\omega_{i0} - \omega_{j0}}{(w_i^t - w_j^t)}\right) = 0$$

$$\boxed{Eq.\,63:\ w^t(x - x_0) = 0}$$

$$\boxed{Eq.\,64:\ w = w_i - w_j = \Sigma^{-1}(\mu_i - \mu_j)}$$

$$\begin{aligned}
-x_0 = \frac{\omega_{i0} - \omega_{j0}}{(w_i^t - w_j^t)} &= \frac{-\frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + ln\, P(\omega_i) + \frac{1}{2}\mu_j^t\Sigma^{-1}\mu_j - ln\, P(\omega_j)}{(\mu_i - \mu_j)^t\Sigma^{-1}} \\
&= \frac{-\frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + \frac{1}{2}\mu_j^t\Sigma^{-1}\mu_j + ln\, [P(\omega_i)/P(\omega_j)]}{(\mu_i - \mu_j)^t\Sigma^{-1}} \\
&= \frac{-\frac{1}{2}(\mu_i - \mu_j)^t\Sigma^{-1}(\mu_i + \mu_j)}{(\mu_i - \mu_j)^t\Sigma^{-1}} + \frac{ln\, [P(\omega_i)/P(\omega_j)]}{(\mu_i - \mu_j)^t\Sigma^{-1}} \cdot \frac{(\mu_i - \mu_j)}{(\mu_i - \mu_j)} \\
&= -\frac{1}{2}(\mu_i + \mu_j) + \frac{ln\, [P(\omega_i)/P(\omega_j)]}{(\mu_i - \mu_j)^t\Sigma^{-1}(\mu_i - \mu_j)} \cdot (\mu_i - \mu_j)
\end{aligned}$$

Por lo tanto:

$$\boxed{Eq.\,65:\ x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{ln\, [P(\omega_i)/P(\omega_j)]}{(\mu_i - \mu_j)^t\Sigma^{-1}(\mu_i - \mu_j)} \cdot (\mu_i - \mu_j)}$$

8. Ejercicio 43

Let the components of the vector $x = (x_1, ..., x_d)^t$ be binary-valued (0 or 1), and let $P(\omega_j)$ be the prior probability for the state of nature ω_j and $j = 1, ..., c$. Now define:

$$p_{ij} = Pr[x_i = 1|\omega_j] \qquad \begin{matrix} i = 1, ..., d \\ j = 1, ..., c, \end{matrix}$$

with the components of x_i being statistically independent for all x in ω_j.

8.1. (a)

Interpret in words the meaning of p_{ij}.

Solución
En cada clase "j" existe un $\underline{x} = [x_1, x_2, ..., x_i, ..., x_d]^t$ cuyas componentes son variables aleatorias independientes *Bernoulli* de parámetro p_{ij}.

8.2. (b)

Show that the minimum probability of error is achieved by the following decision rule: Decide ω_k if $g_k(x) \geq g_j(x)$ for all j and k, where

$$g_j(x) = \sum_{i=1}^{d} x_i \ln \frac{p_{ij}}{1 - p_{ij}} + \sum_{i=1}^{d} \ln(1 - p_{ij}) + \ln P(\omega_j).$$

Solución
Sea la regla de decisión que minimiza la probabilidad de error:

$$P(\omega_i|\underline{x}) \underset{\alpha_j}{\overset{\alpha_i}{\gtrless}} P(\omega_j|\underline{x})$$

$$P(\underline{x}|\omega_i)P(\omega_i) \underset{\alpha_j}{\overset{\alpha_i}{\gtrless}} P(\underline{x}|\omega_j)P(\omega_j) \tag{8}$$

Como las x_i son independientes:

$$P(\underline{x}|\omega_i) = \prod_{i=1}^{d} p_{ij}^{x_i}(1 - p_{ij})^{(1-x_i)} \tag{9}$$

Reemplazando en la ecuación anterior:

$$\prod_{k=1}^{d} p_{ki}^{x_k}(1 - p_{ki})^{(1-x_k)} P(\omega_i) \underset{\alpha_j}{\overset{\alpha_i}{\gtrless}} \prod_{h=1}^{d} p_{hj}^{x_h}(1 - p_{hj})^{(1-x_h)} P(\omega_j)$$

Por lo tanto la función discriminante será:

$$g_i(x) = P(\omega_i) \prod_{k=1}^{d} p_{ki}^{x_k}(1 - p_{ki})^{(1-x_k)}$$

Tomando logaritmo:

$$= \ln\left(P(\omega_i)\right) + \sum_{k=1}^{d} \left[x_k \ln(p_{ki}) + (1 - x_k)\ln(1 - p_{ki})\right]$$

Operando con las propiedades del logaritmo se llega al enunciado propuesto:

$$\boxed{g_i(x) = \ln\left(P(\omega_i)\right) + \sum_{k=1}^{d}\left[x_k \ln\left(\frac{p_{ki}}{1-p_{ki}}\right) + \ln(1-p_{ki})\right]} \tag{10}$$