

tante, es un ineludible deber, y para mí una gran satisfacción, expresar mi gratitud al profesor E. Chacón, a quien debo mi introducción ya hace años en las técnicas cuantitativas, sin cuyo magisterio, consejo y constante apoyo no hubiese podido escribir este libro.

W. S. Meisel, R. O. Duda y P. E. Hart, tantas veces referenciados en la obra, han contribuido en gran medida al análisis comparativo de las diversas técnicas descritas. M. Rebollo y A. Vázquez Muñiz han proporcionado los casos reales recogidos en la obra, y sin sus sugerencias difícilmente hubiese sido una obra de aplicación vívida en la realidad de los problemas.

Debo también mi gratitud a J. Thompson, cuyas apreciaciones sobre el muestreo irrestricto simple multivariante, así como sobre la cuantificación de ciertos criterios a optimizar en la creación de funciones discriminantes tanto han contribuido a la operatividad de los algoritmos recogidos en este trabajo.

Difícilmente podré agradecer las sugerencias, corrección de pruebas y validación de algoritmos y programas debidos a J. López Navalón, J. Jiménez y B. Orfila. Sin su ayuda, esta obra no hubiese traspasado los umbrales de la teoría.

LAUREANO F. ESCUDERO

Centro de Investigación UAM-IBM
Madrid

Introducción a las técnicas de Reconocimiento de Patrones

1.1. NOCIONES SOBRE RECONOCIMIENTO DE PATRONES Y CAMPO DE APLICACION

En sentido amplio, se puede considerar que patrones son los medios a través de los cuales se puede interpretar el mundo. Un niño aprende a distinguir los patrones visuales del padre y de la madre, los patrones auditivos de la música y de la palabra, los patrones sensitivos del calor y del frío, etc. A medida que el niño va creciendo es capaz de definir más detalladamente su reconocimiento de patrones, de tal forma que puede distinguir una sinfonía de Bach de una de Beethoven, o un cuadro de Velázquez de uno de Picasso. Algunos patrones tienen una manifestación física (tipos de caracteres en una cuartilla, tipos de cosas o animales). Otros, en cambio, sólo se perciben de una forma abstracta (patrones sociales o económicos).

¿Qué es un proceso de reconocimiento de patrones («Pattern Recognition»)? Cuando se analiza, por ejemplo, una página impresa y se reconocen sin vacilación sus caracteres, se están utilizando reglas fijas obtenidas a base de experiencia. Quizá no se puede definir en qué consisten exactamente estas reglas, pero al diferenciar las letras «o» y «a», realmente las reglas existen. Puede ser que con el transcurso del tiempo se hayan observado repetidamente muestras de los caracteres «o» y «a» aprendiendo a distinguir sus rasgos, de tal forma que el examen de esta muestra «controlada» («labeled sample») haya producido una regla de decisión. Esta regla de decisión sirve para decidir, sin ninguna otra tutoría, si un nuevo carácter pertenece o no a la «clase a» o a la «clase o».

Los dos aspectos fundamentales en todo reconocimiento de patrones son el desarrollo o creación de una regla de decisión (o clasificador) y la utilización de la misma. El *reconocimiento* se efectúa cuando se utiliza la regla; el *patrón* se define a base de un proceso de aprendizaje («learning process»). Para definir un patrón se utiliza una muestra controlada de dicho patrón. Los patrones de esta muestra son ejemplos de cada una de las clases de patrones (en este caso las letras «a» y «o»).

Un problema de diseño de un clasificador comienza, en líneas generales, con la definición de las clases objeto de estudio (en este caso las letras «a» y «o»), y la representación adecuada de la muestra controlada de cada una de ellas. El problema termina cuando se ha obtenido una regla de decisión en base a la cual se puede asignar un nuevo patrón «no controlado» (y, por tanto, cuya clase de pertenencia se desconoce) a aquella clase para la que se estime un menor riesgo en la asignación.

Aunque se puede abordar el problema del reconocimiento de patrones desde campos muy diversos, esta obra sólo trata del reconocimiento matemático de patrones y de sus aplicaciones. Estas técnicas consisten en un conjunto de algoritmos para resolver un tipo muy particular de problemas presentado en una forma muy peculiar. El éxito de su aplicación depende de la formulación del mismo y de la comprensión de las hipótesis que se consideran en su descripción. Debido a los extraordinarios resultados del reconocimiento sensorial de patrones efectuado por el hombre, existe cierta tendencia a esperar resultados automáticos del reconocimiento de patrones basado en algoritmos matemáticos diseñados y utilizados con ayuda de un ordenador. Como en cualquier otro problema de tipo práctico, se precisa un gran cuidado en la preparación y análisis de los datos, así como en la selección e implantación de los métodos apropiados; rápidos estudios de viabilidad en este campo producen, la mayor de las veces, rápidos fracasos en la obtención de resultados.

Los métodos numéricos en el reconocimiento matemático de patrones se basan en conceptos relativamente simples; su operatividad no depende de la sofisticación relativa al reconocimiento humano, sino de la posibilidad de tratar con ordenador patrones de grandes dimensiones. La percepción humana puede tratar, in mente, sofisticadas reglas y algoritmos, pero no con más de dos o tres dimensiones a visualizar en los patrones. En cambio, el reconocimiento matemático de patrones puede, a base de sencillos algoritmos, manejar masas de datos con mayores dimensiones cuyas relaciones internas no son fácilmente visualizables.

Aunque las reglas de decisión en reconocimiento de patrones deben ser sencillas y rápidas en su resolución, la obtención de las mismas puede

requerir considerable esfuerzo. Esta es la razón fundamental para acudir al concurso de ordenadores. El costo de la fase de creación de las reglas de decisión puede parecer elevado en términos absolutos, pero a menudo es significativamente inferior al coste correspondiente al sistema de recogida, depuración de datos y posterior implantación del clasificador.

Es preciso destacar, aunque sea brevemente, campos típicos de aplicación de las técnicas de reconocimiento de patrones que, sin carácter exhaustivo, expongan su aplicabilidad. Los campos más comunes son los siguientes:

1. Reconocimiento de caracteres, donde el objeto a clasificar son los miembros de un alfabeto dado.
2. Clasificación de huellas digitales.
3. Reconocimiento de fotografías, cuyo objetivo puede consistir en rechazar una gran proporción de fotografías que no sean de interés.
4. Identificación de recursos naturales a base de «fotografías» tomadas por satélites o aviones convencionales. El Centro de Investigación UAM-IBM está actualmente investigando, a base de estas técnicas, los recursos naturales de ciertas regiones españolas, en tipologías tales como oceanográficas, geológicas, mineras, agrícolas, forestales, etc. Técnicas análogas se emplean en análisis geográficos y catastrales.
5. Análisis clínicos, donde las clases pueden ser sano, normal, deficiente o indeterminado.
6. Control de calidad tanto de materiales y máquinas como procesos de fabricación, donde las clases pueden ser los diferentes tipos de calidad de la materia prima, por ejemplo. Son muy interesantes cuando en el objeto analizado se estudia más de una dimensión, ya que en otro caso se utilizan las tradicionales técnicas estadísticas de control de calidad que son un caso particular de reconocimiento de patrones.
7. Reconocimiento auditivo de patrones, incluyendo identificación de la palabra, identificación de oradores, etc.
8. Aplicaciones de ingeniería, tales como la identificación de tipos de aviones a base de análisis de señales emitidas por radar, la selección de diversos tipos de procesos de producción, etc.
9. Aplicaciones de investigación-operativa fundamentalmente como instrumento a utilizar en las técnicas de previsión, identificando comportamientos futuros a base de analizar patrones actuales.
10. Identificación de movimientos sísmicos.
11. El campo de la biomedicina tiene grandes aplicaciones en este tipo de técnicas, dado que los análisis clínicos tales como cardiogramas, encefalogramas, etc., y cualquier tipo de test clínico no tiene una respuesta

exacta en el diagnóstico. El gran volumen de variables a manejar exige, en muchas situaciones, identificar posibles causas o clases a base de técnicas matemáticas de reconocimiento de patrones. De este tipo es el reconocimiento de patrones-mujeres que en ciertas clínicas de maternidad se efectúa a base de clasificar en normal, fallecido, prematuro, diostrópico o indeterminado a un niño próximo a nacer. Las variables o características a utilizar es el historial clínico personal y familiar de la madre.

12. Aplicaciones macroeconómicas. Una aplicación muy interesante de estas técnicas consiste en «predecir» el grado de inflación de un país para el siguiente período estacional, a base de utilizar series económicas de tiempo, de tal forma que se pueda llegar a distinguir e.g. las tres siguientes situaciones: inflación por encima del 17 por 100, inflación entre el 17 por 100 y el 12 por 100 e inflación por debajo del 12 por 100.

13. Aplicaciones meteorológicas. Básicamente, el problema consiste en estratificar en «tipos de tiempo» o clases las situaciones meteorológicas diarias de una región para así poder predecir las más probables. A base de sensores meteorológicos de dirección y velocidad de viento, temperatura, humedad, turbulencia, presión, radiación solar... de superficie se puede obtener diariamente una «fotografía» de la situación meteorológica obteniendo los valores medios y desviaciones típicas de cada seis horas. Las técnicas de «Feature Selection» («Selección de Características») estiman cuáles son las variables meteorológicas más significativas, creando el banco de datos meteorológicos diarios de la región en estudio. Utilizando técnicas de análisis cluster (Parte II) se agrupan estos datos o patrones en grupos o tipos de tiempo, tal que aquellos patrones con mayor similitud deben formar parte del mismo grupo o «cluster». Estos tipos de tiempo serán las clases del problema. Para cada uno de los días recogidos en el banco de datos meteorológicos, se analiza en los mapas sinóptico-meteorológicos del área en que está enclavada la región en estudio, los parámetros que mejor representen la situación sinóptica. Con estos parámetros se diseña un clasificador cuyo objetivo ha de ser el asignar correctamente cada situación meteorológica de superficie diaria al correspondiente tipo de tiempo. Dado que es más factible la predicción sinóptica que la de superficie, la predicción diaria de los parámetros sinópticos sirve de ayuda para la predicción meteorológica de superficie, si el clasificador se ha diseñado adecuadamente.

14. Estudio de mercado. Hoy día es preciso, conocer con más detalle las características, tipologías y estratificaciones del mercado en que opera la empresa. Las técnicas de reconocimiento de patrones ayudan a conocer mejor dicho mercado, obteniendo las características discriminantes, las

motivaciones y el volumen de sus diversas estratificaciones. Aplicaciones análogas se encuentran en los campos de la Sociología y Psicología.

15. Aplicaciones en el Departamento de Personal. Uno de los problemas latentes en la empresa consiste en la tipificación de su personal, agrupando el mismo de acuerdo con su nivel técnico, predisposición para cierto tipo de trabajos, circunstancias personales, etc. Esta tipificación es útil para el mejor conocimiento del inventario humano y, por tanto, ayuda a una mejor evaluación de la rentabilidad a obtener en los diferentes puestos de trabajo.

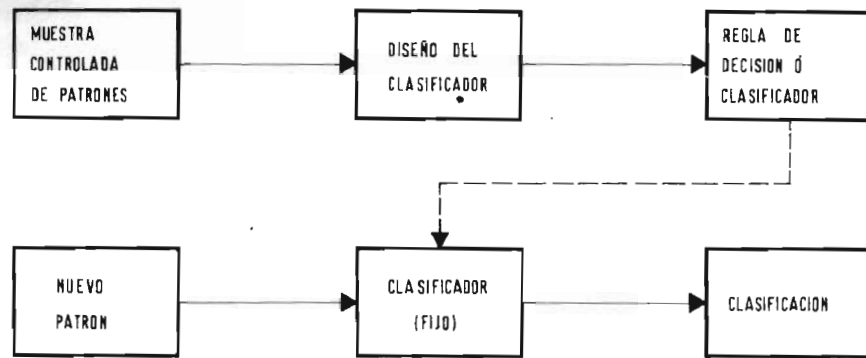
Se puede considerar, en general, como un buen candidato para la aplicación de las técnicas de reconocimiento de patrones a cualquier sistema en el que, conociendo una muestra controlada de cada clase, se desea asignar a una de ellas un nuevo dato cuya clase de pertenencia se desconoce. Puede, en cambio, producir resultados negativos una aparente aplicación de estas técnicas, si la muestra controlada es muy pequeña, no se han seleccionado correctamente las características discriminantes o las clases están muy poco discriminadas.

1.2. PROCESO DE RECONOCIMIENTO DE PATRONES

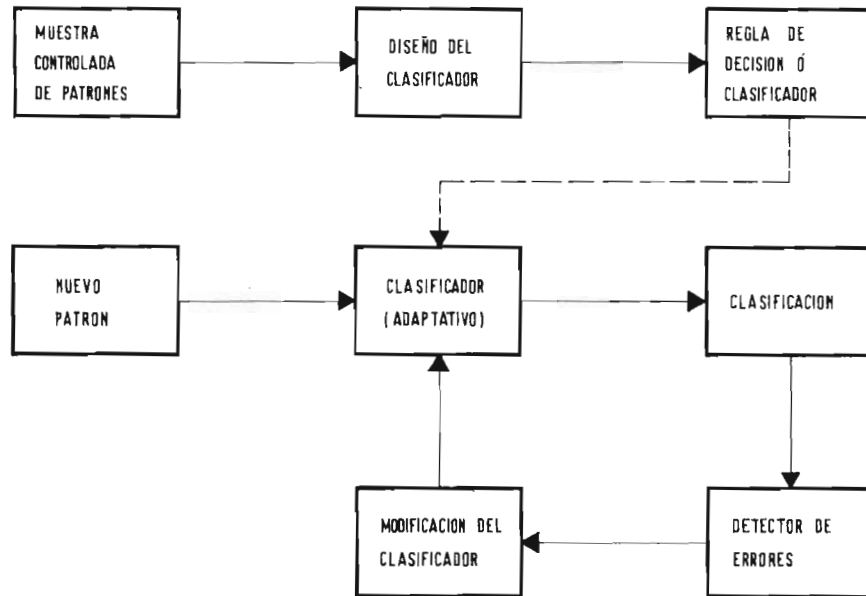
Las dos fases fundamentales de un proceso de reconocimiento de patrones son: obtención de la regla de decisión (o sea el diseño del clasificador) y utilización de la misma. Su utilización puede ser fija o adaptativa. En la figura 1.1a se recoge el procedimiento de utilización fija. Se diseña el clasificador a base de la muestra controlada de patrones y no sufre ninguna modificación una vez clasificados los patrones cuya clase de pertenencia se desconoce.

En la figura 1.1b se recoge el caso en que el clasificador se modifica a medida que se va utilizando. En este caso, una vez construido el clasificador, se asigna un nuevo patrón a una clase, según el resultado de su aplicación. Un detector de errores indica a continuación si la clasificación ha sido correcta, en cuyo caso no se modifica el clasificador, o si es preciso proceder a su modificación si el resultado ha sido erróneo. La utilización del clasificador no sería necesaria si el detector de errores se pudiera utilizar al mismo tiempo. Normalmente, éste no es el caso, pero en cambio es muy útil el detector de errores ya que «ajusta» el clasificador mejorando su utilización futura.

En la figura 1.2 se recoge el diagrama del proceso a seguir para la obtención de un clasificador. En primer lugar se ha de estudiar el sistema



(a)



(b)

Figura 1.1.—Estrategias de reconocimiento: (a) fija y (b) adaptativa.

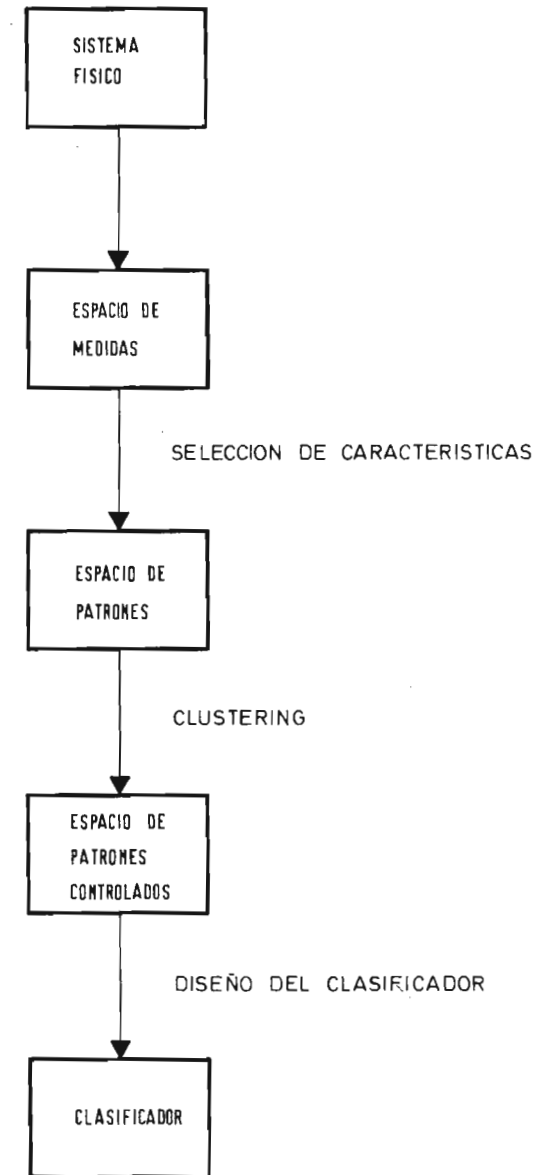


Figura 1.2.—Fases en el diseño de un clasificador

físico del que deben de partir los patrones a clasificar. Este sistema se caracteriza por una serie de mediciones. De este espacio de medidas se obtienen las características más representativas mediante la técnica de Selección de Características. Si la muestra de patrones, obtenida en el espacio físico, no es controlada, es preciso utilizar técnicas de análisis «cluster» (Parte II) para encontrar las «clases naturales» en que se agrupan estos patrones. En cualquier caso, la fase siguiente consiste en diseñar el clasificador a base de utilizar en cada clase la muestra controlada de patrones, considerando únicamente las características estimadas como más discriminantes del patrón.

Sea el ejemplo de aplicaciones meteorológicas recogido en el apartado anterior. El espacio físico consiste en la superficie de la región en estudio así como los sensores que miden las condiciones meteorológicas. El espacio de medidas lo componen los valores diarios de las diferentes condiciones meteorológicas, obteniendo de los mismos el promedio y, en algunos casos, la desviación típica, de cada seis horas. El número de características de cada patrón diario normalmente es muy elevado, ya que e. g. si sólo se contempla el período diurno, considerando para las siete variables meteorológicas el promedio de cada seis horas y la desviación típica de la dirección del viento, el número total de características de cada patrón meteorológico diario es de 16. Es preciso utilizar la técnica de Selección de Características eliminando aquellas que no sean discriminantes y/o obteniendo unas nuevas mediante combinación de las anteriores, de tal forma que, sin perder información *relevante*, se reduzca el espacio dimensional y así sea más manejable. Si el espacio de patrones resultante no es controlado, es decir si se desconoce la clase de pertenencia de cada uno de ellos, es preciso mediante técnicas de análisis «cluster» encontrar las clases naturales en que, por su mayor grado de similitud, se agrupan los patrones, considerando entonces que éstos ya son controlados. La fase final del proceso de creación del algoritmo del reconocimiento de patrones consiste en el diseño del clasificador, de tal forma que los nuevos patrones meteorológicos se asignen correctamente a su correspondiente clase de pertenencia.

1.3. CASO TÍPICO

A) Aproximación Inicial

En este apartado se recoge un ejemplo (Duda, 1970) a través del cual se pueden analizar, numéricamente, las fases fundamentales del reconocimiento de patrones, así como alguna de las técnicas más utilizadas. Sólo se

pretende dar una panorámica del proceso y, por tanto, de las líneas generales del diseño de un clasificador de patrones.

Sea el caso en el que ante una muestra controlada de los caracteres alfabéticos *B* y *8*, escritos manualmente, se desea diseñar una regla de decisión en virtud de la cual un nuevo carácter, escrito manualmente, pueda ser identificado como una *B* o un *8* sin otro instrumento que un rápido clasificador matemático.

El proceso de reconocimiento de patrones (Fig. 1.3) comienza con el análisis del espacio físico, en este caso el hombre y los instrumentos que

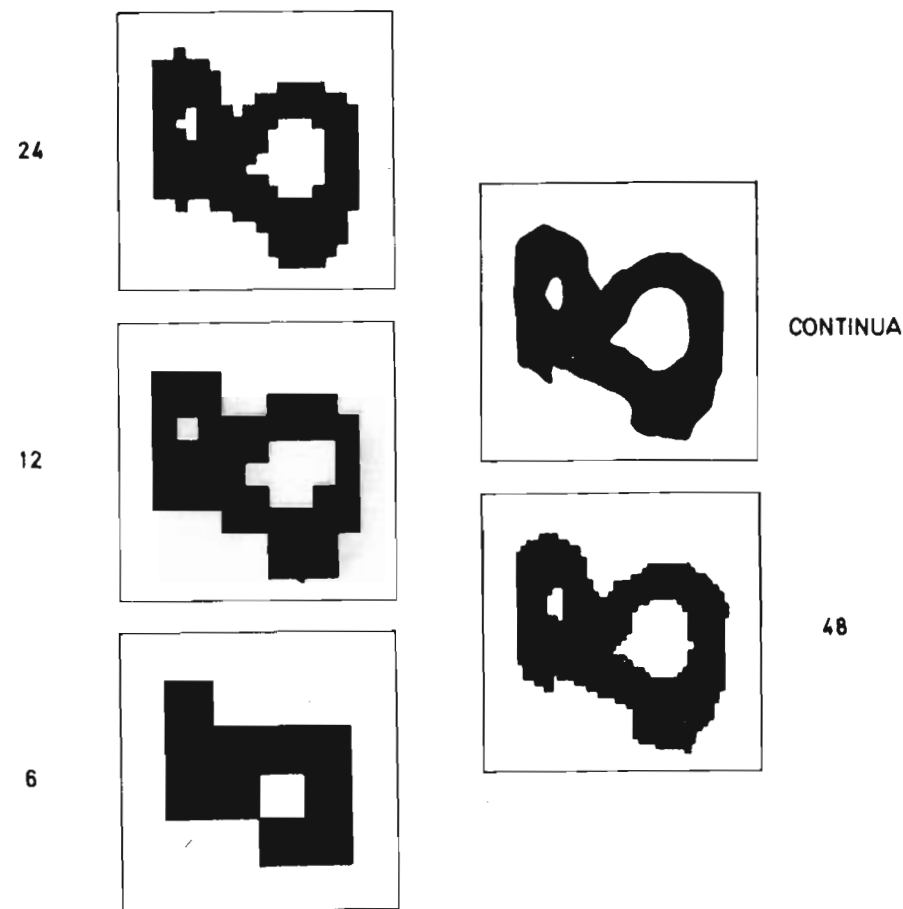


Figura 1.3.—Representación del carácter *B* escrito manualmente. La cuantificación se ha efectuado con 6, 12, 24, 48 líneas y de una forma continua.

utiliza para escribir. El espacio de medidas consiste en la cuantificación de los caracteres *B* y 8 que son la muestra de patrones controlados producidos por el hombre. Es importante determinar adecuadamente el grado de precisión en esta cuantificación, porque de lo contrario los resultados pueden ser muy costosos o muy erróneos.

En la figura 1.3 se recoge el resultado de la representación del carácter *B* que, aunque producido manualmente, se ha cuantificado con diferentes grados de precisión. Claramente se observa que con un grado de precisión inferior a 24 líneas, difícilmente se podría estimar si es una *B* o un 8. Es aconsejable analizar cuidadosamente el espacio de medidas ya que los instrumentos que las efectúan o propagan en el sistema físico introducen, numerosas veces, ruidos o distorsiones que enmascaran la realidad dando ocasión a diseños falseados. En este caso se ha elegido una representación del carácter en 24×24 puntos ya que visualmente parece adecuada.

La siguiente fase consiste en la selección de las características que más puedan discriminar un patrón *B* de un patrón 8. Es preciso notar que el carácter *B* está «cortado» en su parte izquierda, de tal forma que una de las posibles características diferenciadoras con respecto al carácter 8, podría ser la ratio (denominada *X*) entre la distancia de la parte superior a la parte inferior del carácter y la longitud de su arco izquierdo. Así, una *B* perfecta tendrá una ratio de 1 y un 8 perfecto tendrá una ratio muy inferior a 1 y siempre superior a 0. Por tanto, en principio se elige esta característica, cuyo rango va de 0 a 1, para discriminar ambos caracteres. Si se hubiese sido preciso utilizar más características, intuitivamente discriminantes, hubiese sido preciso utilizar la técnica de Selección de Características al objeto de reducir el nivel de dimensionalidad de las mismas si éste fuese elevado.

Considerando el espacio de medidas del carácter *B* recogido en la figura 1.4 y del carácter 8 recogido en la figura 1.5, se obtiene la característica de cada patrón tal como se ha indicado. Dado que la muestra es controlada, se procede a continuación al diseño del clasificador. El promedio de la característica *X* en la clase *B* es 0,92, siendo 0,85 para la clase 8. Sería razonable considerar un valor intermedio X_0 , p. ej. $X_0 = 0,89$, tal que se considera que un futuro carácter es una *B* si su característica es $X \geq 0,89$, estimando que dicho carácter es 8 si $X < 0,89$.

El resultado que obtiene Duda, 1970 indica que se ha clasificado erróneamente el 23 por 100 de los patrones de la muestra controlada. Las causas que lo hayan podido motivar son las siguientes: a) insuficiente grado de precisión

con 24 líneas, b) la característica considerada no tiene el adecuado grado de discriminación, c) ha sido medida erróneamente, o d) el umbral X_0 no ha sido elegido correctamente.

B) Mejora del clasificador

De las cuatro causas que han podido provocar un mal resultado del clasificador, la elección del umbral se puede verificar de la forma siguiente. En la figura 1.6 se recoge la función de distribución (probabilidad acumulada) obtenida empíricamente a base de las muestras controladas de *B* y 8 cuya representación se recoge en las figuras 1.4 y 1.5. Se puede observar que el rango va desde 0,57 a 1. Así p. ej. $F(X | 8)$ para $X = 0,8$ es $F(0,8 | 8) = 0,30$, lo que significa que el 30 por 100 de los patrones controlados de 8 tienen un valor en la característica *X* menor o igual que 0.8.

El umbral ideal X_0 sería aquel para el cual $F(X_0 | 8) = 1$ y $F(X_0 | B) = 0$ ya que, al menos en la muestra controlada, discriminaría totalmente ambos

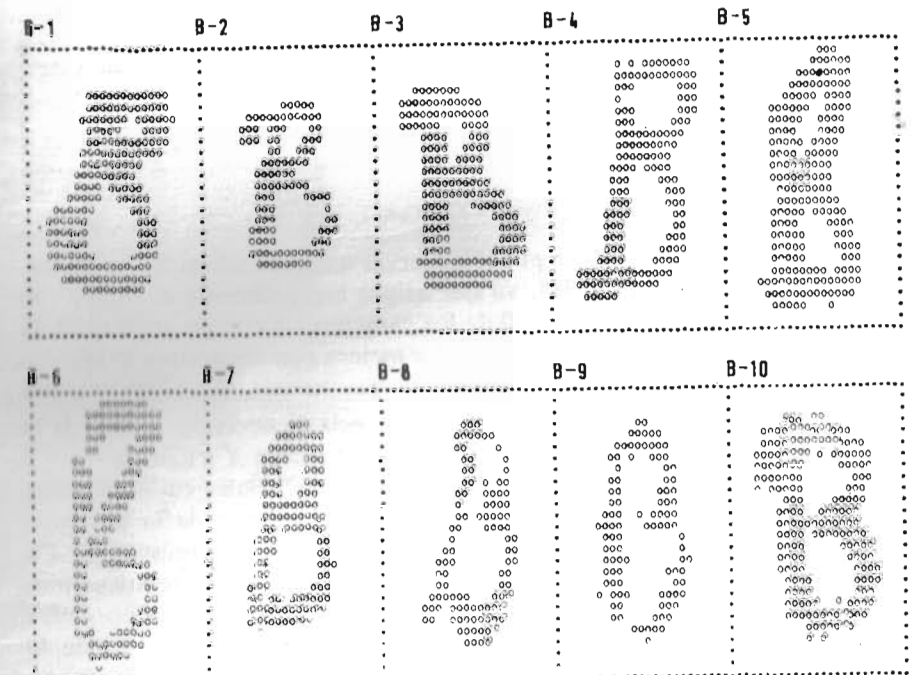


Figura 1.4.—Representación de la muestra controlada de caracteres *B* escritos manualmente.

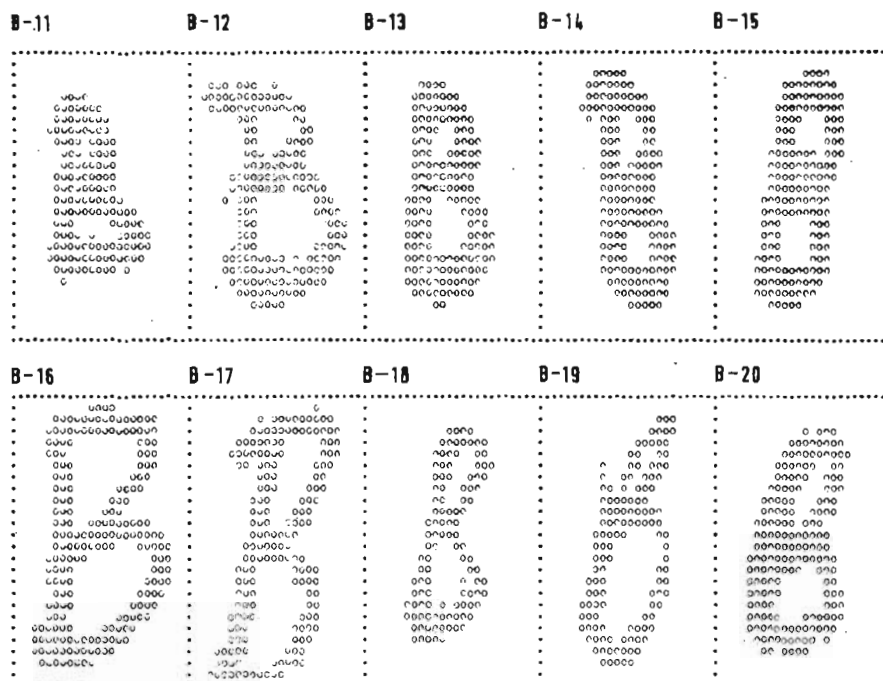


Figura 1.4.—(Cont.)

caracteres. En la figura 1.6 se puede observar que tal umbral X_0 no existe. El mejor umbral es $X_0 = 0,885$, ya que asigna erróneamente el 10 por 100 de los caracteres B y el 30 por 100 de los caracteres 8 con un error total de $0,10 \cdot 20 + 0,30 \cdot 20 = 8 = 20 \% \text{ s}/40$. La mejora con respecto a la aproximación inicial no es significativa.

Otra de las típicas formas de obtener una regla de decisión, a base de la muestra controlada, consiste en asignar el nuevo patrón X a aquella clase a la cual corresponda el patrón X_j de la muestra controlada cuya distancia $(X - X_j)^2$ sea mínima. Esta técnica se denomina estrategia de la distancia mínima, también muy utilizada en las técnicas clustering (capítulo 12). En el apartado 6.6 se recogen las hipótesis de partida. Si no se repitiese ningún valor X en clases distintas, con lo cual no habría lugar a ambigüedades al aplicar la estrategia de la distancia mínima a la muestra controlada de 40 patrones, se obtendría un resultado del 0 por 100 de error. En cambio, en este caso es del 12,5 por 100, reduciendo considerablemente el error del 20 por 100 de la aproximación anterior. No obstante, hay que considerar que

el resultado obtenido al probar el clasificador con la misma muestra controlada que se ha utilizado en su diseño, no es muy significativo si la muestra no lo es del colectivo total. En el apartado 1.4.4 se estudia la posibilidad de dividir la muestra controlada en muestra («training field») para diseñar el clasificador y muestra («test field») para estimar el error de la futura clasificación.

C) *Aproximación estadística*

Se considera *probabilidad a priori* de la clase, p. ej. B , a la probabilidad de que, sin conocer el valor de la característica X , el patrón pertenezca a la clase B . Si el carácter, cuya clase se desconoce, sólo puede ser B ó 8 , entonces la suma de las probabilidades a priori $P(B)$ y $P(8)$ es uno:

$$P(B) + P(8) = 1 \quad (1.1)$$

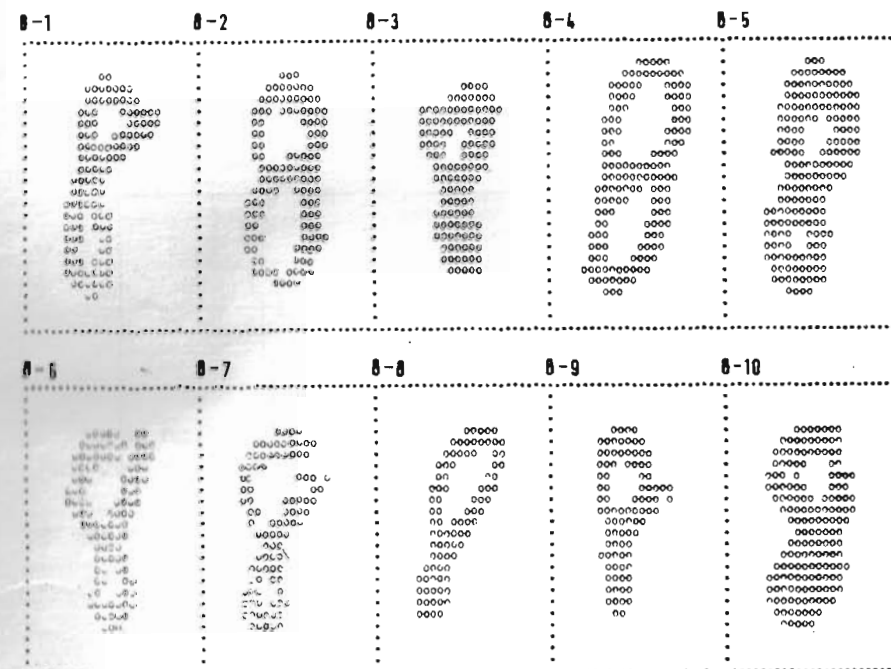


Figura 1.5.—Representación de la muestra controlada de caracteres 8 escritos manualmente.

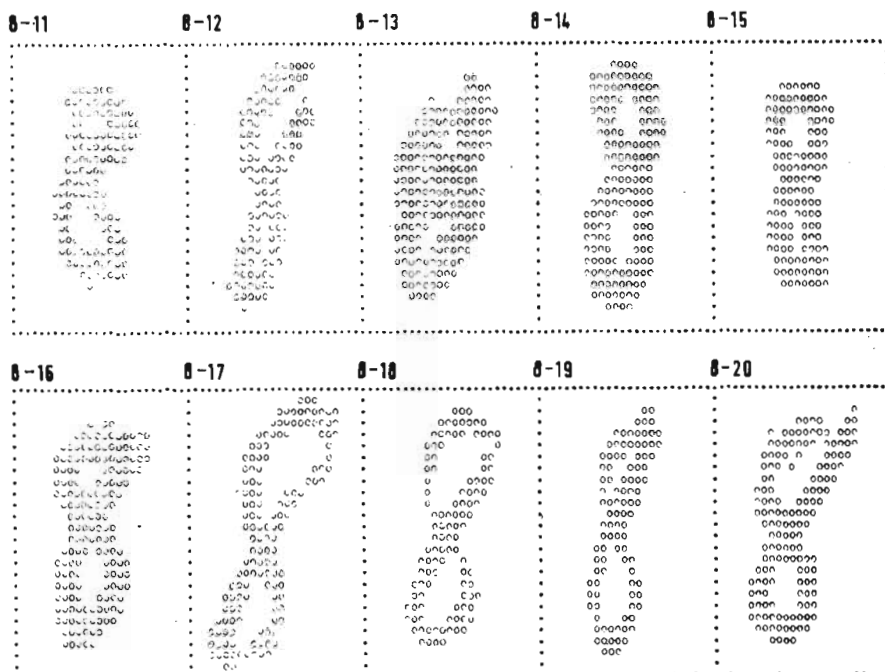


Figura 1.5.—(Cont.)

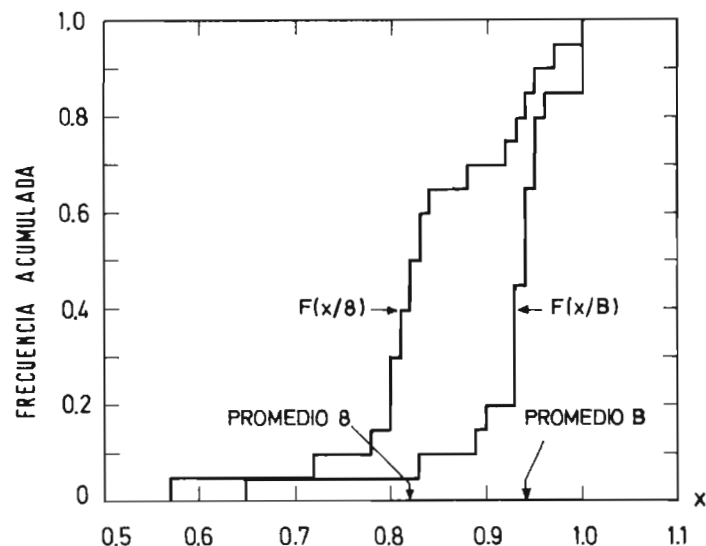


Figura 1.6.—Frecuencia acumulada de la característica X.

Se define como *probabilidad condicional* $P(X|B)$ a la probabilidad de que la característica de un patrón que pertenece a la clase B tome el valor X . En la eq. (1.2) se recoge una aproximación de este valor, diferenciando la distribución empírica acumulada $F(X|B)$.

$$P(X|B) = \frac{dF(X|B)}{dX} \quad (1.2)$$

$$P(X|8) = \frac{dF(X|8)}{dX} \quad (1.3)$$

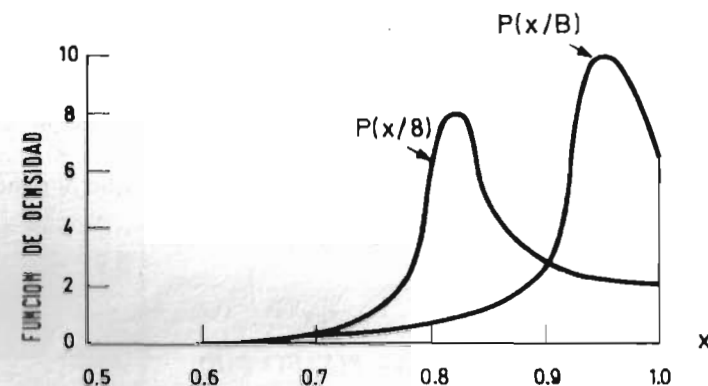


Figura 1.7.—Función de densidad de la característica X.

En la figura 1.7 se recoge la representación de las correspondientes funciones de densidad de los caracteres B y 8 .

La probabilidad $P(X)$ de que el valor de la característica del patrón sea X , sea cual fuese la clase de su pertenencia, será:

$$P(X) = P(X|B) * P(B) + P(X|8) * P(8) \quad (1.4)$$

Se define como *probabilidad a posteriori* a la probabilidad de que el patrón pertenezca a la clase p. ej. B , conociendo el valor X de su característica.

Según el teorema de Bayes, su valor será:

$$P(B|X) = \frac{P(X|B) * P(B)}{P(X)} \quad (1.5)$$

$$P(8|X) = \frac{P(X|8) * P(8)}{P(X)} \quad (1.6)$$

De acuerdo con esta aproximación estadística, se puede considerar que la regla óptima de decisión, en el sentido de que minimiza la probabilidad de error, es aquella que asigna el patrón X a la clase para la cual la probabilidad a posteriori es mayor, de tal forma que si $P(8|X) \geq P(B|X)$ se asigna al carácter 8, y en caso contrario al carácter B .

A falta de mayor información, se asume que para el caso recogido en las figuras 1.4 y 1.5, ambas probabilidades a priori son iguales ($P(B) = P(8) = 20/40 = 0,5$). Las probabilidades condicionales $P(X|B)$ y $P(X|8)$ se han estimado, según las eqs. (1.2 y 1.3), a base de considerar que las frecuencias acumuladas $F(X|B)$ y $F(X|8)$ representan las correspondientes probabilidades acumuladas.

Considerando las eqs. (1.5) y (1.6) se puede observar que la condición

$$P(8|X) \geq P(B|X) \quad (1.7)$$

es equivalente a la condición

$$P(X|8) * P(8) \geq P(X|B) * P(B) \quad (1.8)$$

donde, para el caso de probabilidades a priori iguales, resulta:

$$P(X|8) \geq P(X|B) \quad (1.9)$$

En la figura 1.7 se puede observar que el valor del umbral X_0 para el cual $P(X|8) = P(X|B)$ es muy cercano a 0,9, tal como se había estimado anteriormente.

C) Nueva selección de características

Aunque el clasificador resultante de la aplicación de la aproximación estadística anterior no ha mejorado la efectividad de la clasificación, abre en cambio un nuevo campo de posibilidades en la obtención de la regla de decisión.

Considerando que el grado de precisión de la cuantificación en 24 líneas es suficiente para la discriminación de cada patrón, una de las posibilidades de mejora del clasificador consiste en la prospección de nuevas características. Parece que es adecuada la selección de la característica X_1 ratio entre la «altura» del carácter y la longitud de su arco izquierdo. Pero otra característica que puede ser muy significativa es la ratio (X_2) entre las alturas de las partes superior e inferior del carácter, ya que existe cierta tendencia a escribir manualmente el carácter B con mayor altura en la parte inferior y, en cambio, el carácter 8 con mayor altura en la parte superior.

Por tanto, de acuerdo con esta particular selección de características, cada uno de los 40 patrones que componen la muestra controlada (Figs. 1.4 y 1.5) está definido por el vector $X = (X_1, X_2)'$ que recoge los valores de ambas características. En la figura 1.8 se recoge la representación del correspondiente plano.

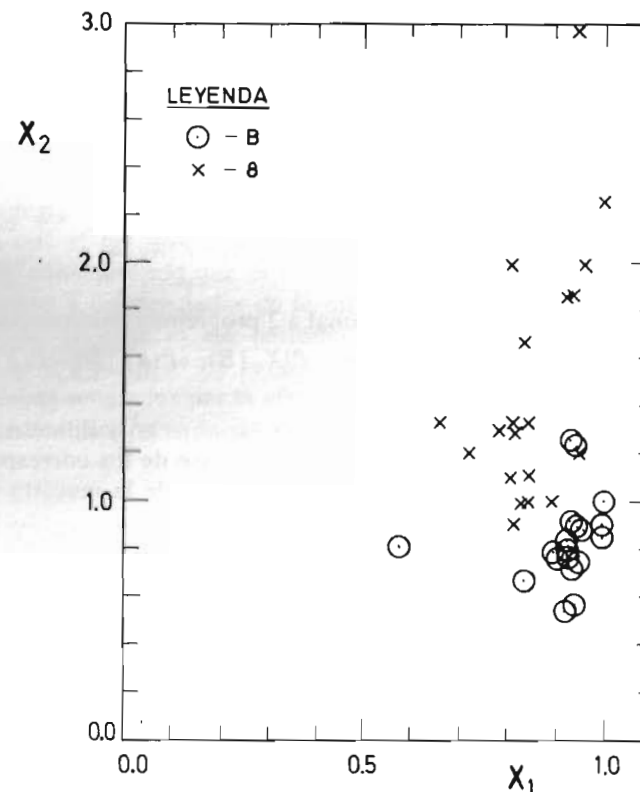


Figura 1.8.—Representación de patrones bidimensionales.

E) Diseño del nuevo clasificador

De acuerdo con las ideas recogidas anteriormente, se asignará un nuevo patrón a aquella clase B ó 8 cuya probabilidad a posteriori $P(B|X)$ o $P(8|X)$ sea mayor, con la particularidad de que ahora X es multidimensional (bidimensional en este caso) y que, por tanto, las funciones de densidad $P(X)$, $P(X|B)$ y $P(X|8)$ son distribuciones conjuntas y, en cambio en la exposición anterior eran distribuciones marginales o también llamadas simples, ya que sólo eran de una variable.

Al igual que en el caso anterior, de nuevo el problema consiste en estimar las probabilidades a priori $P(B)$ y $P(8)$ y condicionales conjuntas $P(X|B)$ y $P(X|8)$. En este caso se asume que, dada la composición de la muestra controlada, las probabilidades a priori son iguales. Dado que el vector X está compuesto por más de una variable, la estimación de las funciones de densidad $P(X_1, X_2|B)$ y $P(X_1, X_2|8)$ es más difícil. Una aproximación general se recoge en el capítulo 4. En este caso, y siguiendo una práctica muy generalizada, se hacen las tres hipótesis siguientes:

a) Las características son estadísticamente independientes, lo que significa que la matriz de covarianzas es diagonal, y por tanto:

$$P(X|B) = P(X_1, X_2|B) = P(X_1|B) * P(X_2|B) \quad (1.10)$$

$$P(X|8) = P(X_1, X_2|8) = P(X_1|8) * P(X_2|8) \quad (1.11)$$

lo cual reduce un problema 2-dimensional a 2 problemas unidimensionales.

b) Las distribuciones marginales $P(X_1|B)$, $P(X_2|B)$, $P(X_1|8)$ y $P(X_2|8)$ siguen la ley Normal.

c) Estas distribuciones tienen la misma varianza σ^2 y difieren sólo en su promedio $\mu_1^{(B)}$, $\mu_2^{(B)}$, $\mu_1^{(8)}$ y $\mu_2^{(8)}$ estimado a base de los correspondientes promedios (figuras 1.8 y 1.9) de los 40 patrones de la muestra controlada.

Como resultado de estas hipótesis, se tiene que:

$$P(X|B) = \frac{1}{\sqrt{2\pi\sigma^2}} * \exp \left\{ -\frac{(X_1 - \mu_1^{(B)})^2}{2\sigma^2} \right\} * \frac{1}{\sqrt{2\pi\sigma^2}} * \exp \left\{ -\frac{(X_2 - \mu_2^{(B)})^2}{2\sigma^2} \right\} =$$

$$= \frac{1}{2\pi\sigma^2} * \exp \left\{ -\frac{1}{2\sigma^2} * [(X_1 - \mu_1^{(B)})^2 + (X_2 - \mu_2^{(B)})^2] \right\} =$$

$$= \frac{1}{2\pi\sigma^2} * \exp \left\{ -\frac{1}{2\sigma^2} * \|X - \mu_B\|^2 \right\} \quad (1.12)$$

$$P(X|8) = \frac{1}{2\pi\sigma^2} * \exp \left\{ -\frac{1}{2\sigma^2} * \|X - \mu_8\|^2 \right\} \quad (1.13)$$

donde

$$\mu_B = (\mu_1^{(B)}, \mu_2^{(B)})^t \quad \text{y} \quad \mu_8 = (\mu_1^{(8)}, \mu_2^{(8)})^t$$

Dadas las hipótesis anteriores, para que se cumpla que $P(B|X) \geq P(8|X)$ (en cuyo caso el patrón X se asigna a la clase B), es preciso que:

$$P(X|B) \geq P(X|8) \quad (1.14)$$

es decir,

$$\|X - \mu_B\|^2 \leq \|X - \mu_8\|^2 \quad (1.15)$$

Esta regla de decisión basada en la distancia euclídea es simple e intuitiva, ya que el patrón X será asignado a aquella clase a la que esté más próximo, considerando que el vector promedio es el *único* estadístico representativo y diferenciador de la misma. La utilización de este clasificador divide el espacio de patrones en dos regiones, una en la cual $\|X - \mu_B\|^2 \leq \|X - \mu_8\|^2$ y la otra en la cual $\|X - \mu_B\|^2 > \|X - \mu_8\|^2$, estando ambas separadas por la «frontera de decisión» («decision boundary») cuyos puntos satisfacen la condición:

$$\|X - \mu_B\|^2 = \|X - \mu_8\|^2 \quad (1.16)$$

donde

$$X_1^2 + \mu_1^{(B)2} - 2 * X_1 * \mu_1^{(B)} + X_2^2 + \mu_2^{(B)2} - 2 * X_2 * \mu_2^{(B)} =$$

$$= X_1^2 + \mu_1^{(8)2} - 2 * X_1 * \mu_1^{(8)} + X_2^2 + \mu_2^{(8)2} - 2 * X_2 * \mu_2^{(8)} \quad (1.17)$$

resultando la siguiente discriminación lineal:

$$\frac{1}{2} * (\mu_1^{(B)2} - \mu_1^{(8)2} + \mu_2^{(B)2} - \mu_2^{(8)2}) = (\mu_1^{(B)} - \mu_1^{(8)}) * X_1 + (\mu_2^{(B)} - \mu_2^{(8)}) * X_2 \quad (1.18)$$

Dado que no se conocen los valores promedios, su estimación se efectúa a base de los promedios de la muestra controlada, siendo éstos $m_1^{(B)} = 0,92$, $m_2^{(B)} = 0,84$, $m_1^{(8)} = 0,85$ y $m_2^{(8)} = 1,51$. En la figura 1.9 se recoge la solución obtenida.

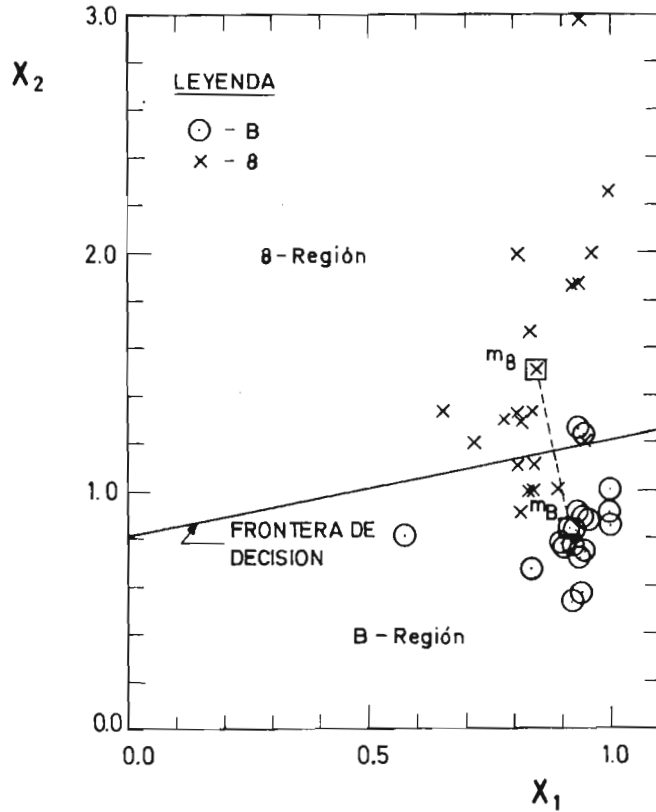


Figura 1.9.—Frontera de decisión basada en la distancia euclídea.

Se puede observar que el 20 por 100 de los patrones se han clasificado erróneamente. Como este clasificador es óptimo, el resultado sólo puede ser debido a que las hipótesis de partida no son correctas. En la figura 1.8 se puede observar que las características no son independientes, que las distribuciones no son gaussianas y que las características tienen un grado de variabilidad diferente. Considerando como ciertas las dos primeras hi-

pótesis y que σ_1^2 y σ_2^2 son, respectivamente, las varianzas de las características X_1 y X_2 sea cual fuere la clase, las nuevas probabilidades condicionales serán:

$$P(X|B) = \frac{1}{2\pi\sigma_1\sigma_2} * \exp \left\{ -\frac{1}{2} * \left[\left(\frac{X_1 - \mu_1^{(B)}}{\sigma_1} \right)^2 + \left(\frac{X_2 - \mu_2^{(B)}}{\sigma_2} \right)^2 \right] \right\} = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} * \exp \left\{ -\frac{1}{2} * (X - \mu_B)^t * \Sigma^{-1} * (X - \mu_B) \right\} \quad (1.19)$$

$$P(X|8) = \frac{1}{2\pi\sigma_1\sigma_2} * \exp \left\{ -\frac{1}{2} * \left[\left(\frac{X_1 - \mu_1^{(8)}}{\sigma_1} \right)^2 + \left(\frac{X_2 - \mu_2^{(8)}}{\sigma_2} \right)^2 \right] \right\} = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} * \exp \left\{ -\frac{1}{2} * (X - \mu_8)^t * \Sigma^{-1} * (X - \mu_8) \right\} \quad (1.20)$$

donde Σ es la matriz de covarianzas de las características que en este caso particular será

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad (1.21)$$

Las eqs. (1.19) y (1.20) son, respectivamente, similares a las eqs. (1.12) y (1.13). La diferencia estriba en que en las eqs. (1.12) y (1.13), las características no están normalizadas considerando sólo la distancia euclídea, y en las eqs. (1.19) y (1.20) las características están normalizadas con sus varianzas y, en sentido amplio, con la matriz Σ con independencia de la correlación existente entre ellas considerando, por tanto, en este caso la distancia de Mahalanobis. En los capítulos 11 y 12 se recogen los diferentes campos de aplicación y subsiguientes resultados en la utilización de ambos tipos de estrategias.

Considerando nuevamente que las probabilidades a priori son iguales, para que se cumpla que $P(B|X) \geq P(8|X)$ (en cuyo caso el patrón X se asigna a la clase B), es preciso que

$$P(X|B) \geq P(X|8) \quad (1.22)$$

es decir,

$$(X - \mu_B)^t * \Sigma^{-1} * (X - \mu_B) \leq (X - \mu_8)^t * \Sigma^{-1} * (X - \mu_8) \quad (1.23)$$

La regla de decisión recogida en la eq. (1.23) asignará el patrón X a aquella clase cuya distancia de Mahalanobis sea menor. Al no ser el vector promedio el único estadístico diferenciador de las clases, puede ocurrir que no se asigne un patrón a aquella clase cuya distancia «geométrica» sea menor. Ello ocurrirá cuando el patrón no esté en el «área de influencia» de la misma. Para mayor explicación ver capítulo 11.

La frontera de decisión que separa las dos regiones en que el espacio de patrones queda dividido por el clasificador, es aquella para la cual los puntos correspondientes satisfacen la condición

$$(X - \mu_B)^t * \Sigma^{-1} * (X - \mu_B) = (X - \mu_8)^t * \Sigma^{-1} * (X - \mu_8) \quad (1.24)$$

donde

$$\begin{aligned} (X - \mu_B)^t * \Sigma^{-1} * X - (X - \mu_B)^t * \Sigma^{-1} * \mu_B = \\ = (X - \mu_8)^t * \Sigma^{-1} * X - (X - \mu_8)^t * \Sigma^{-1} * \mu_8 \end{aligned} \quad (1.25)$$

$$\begin{aligned} X^t * \Sigma^{-1} * X - \mu_B^t * \Sigma^{-1} * X - X^t * \Sigma^{-1} * \mu_B + \mu_B^t * \Sigma^{-1} * \mu_B = \\ = X^t * \Sigma^{-1} * X - \mu_8^t * \Sigma^{-1} * X - X^t * \Sigma^{-1} * \mu_8 + \mu_8^t * \Sigma^{-1} * \mu_8 \end{aligned} \quad (1.26)$$

$$\begin{aligned} -\mu_B^t * \Sigma^{-1} * X + \mu_8^t * \Sigma^{-1} * X - X^t * \Sigma^{-1} * \mu_B + X^t * \Sigma^{-1} * \mu_8 = \\ = \mu_8^t * \Sigma^{-1} * \mu_8 - \mu_B^t * \Sigma^{-1} * \mu_B \end{aligned} \quad (1.27)$$

$$\begin{aligned} (\mu_8 - \mu_B)^t * \Sigma^{-1} * X + X^t * \Sigma^{-1} * (\mu_8 - \mu_B) = \\ = (\mu_8^t * \Sigma^{-1} * \mu_8 + \mu_8^t * \Sigma^{-1} * \mu_8 - \mu_B^t * \Sigma^{-1} * \mu_8 - \mu_B^t * \Sigma^{-1} * \mu_B) \end{aligned} \quad (1.28)$$

$$2 * X^t * \Sigma^{-1} * (\mu_8 - \mu_B) = (\mu_8 + \mu_B)^t * \Sigma^{-1} * (\mu_8 - \mu_B) \quad (1.29)$$

$$X^t * \Sigma^{-1} * (\mu_8 - \mu_B) = \frac{1}{2} * (\mu_8 + \mu_B)^t * \Sigma^{-1} * (\mu_8 - \mu_B) \quad (1.30)$$

Dado que tampoco se conocen las varianzas σ_1^2 y σ_2^2 que integran la matriz Σ , la varianza σ_1^2 se estima a base de la varianza S_1^2 de la característica X_1 en la muestra controlada de 40 patrones (siendo $S_1^2 = 0,0075$) y la varianza σ_2^2 se estima con el correspondiente valor $S_2^2 = 0,350$. En la figura 1.10 se recoge el clasificador lineal (1.30).

El error en la clasificación de los patrones de la muestra controlada es del 7,5 por 100 (3 patrones). Teóricamente, se podría incluso mejorar este resultado analizando más detenidamente las hipótesis de las probabilidades condicionales gaussianas, así como de la independencia de las características, pero dado que la muestra controlada puede no ser significativa, que el

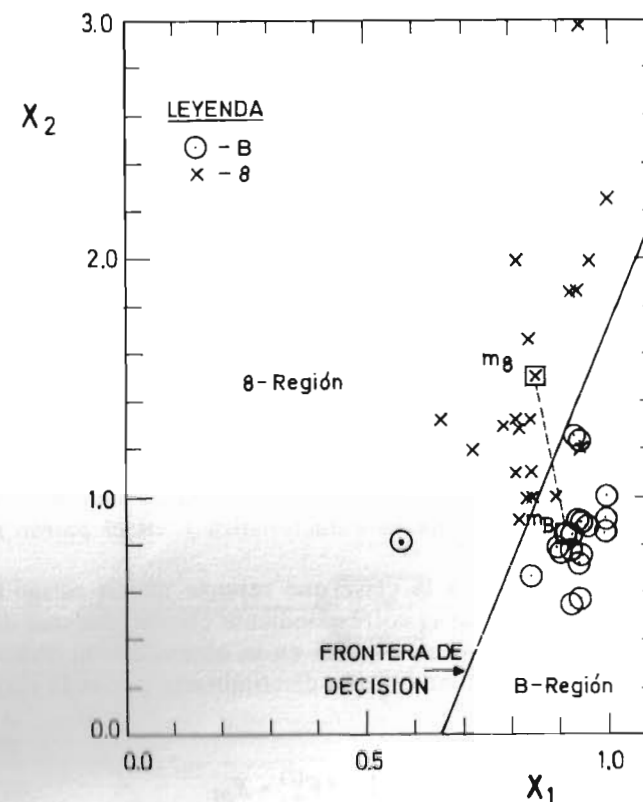


Figura 1.10.—Frontera de decisión basada en la distancia de Mahalanobis.

tanto por 100 de error es pequeño y que éste no se ha obtenido sobre una muestra independiente a la que ha servido para diseñar el clasificador, se puede considerar a efectos prácticos que el proceso de diseño del clasificador de futuros patrones ha terminado. En el apartado 1.4.4 se estudian las posibilidades de las muestras de diseño y de validación.

El objetivo de este apartado ha consistido en presentar, en líneas generales, las diversas etapas en la creación de un algoritmo de reconocimiento de patrones, así como las técnicas más sencillas. En los siguientes apartados de este capítulo se describen formalmente los principales conceptos, introduciendo los diversos tipos de técnicas a utilizar en el diseño de un clasificador.

1.4. CONCEPTOS BASICOS EN RECONOCIMIENTO DE PATRONES

1.4.1. Definición del problema

El objetivo final de las técnicas de Reconocimiento de Patrones consiste en asignar un patrón a la clase que reporte un menor riesgo en la clasificación. El patrón j está definido por el conjunto $X_j = (X_{1j}, X_{2j}, \dots, X_{fj}, \dots, X_{Fj})^T$ de F variables unidimensionales, de tal forma que X_{fj} recoge el valor de la característica f en el patrón j .

En todo problema de Reconocimiento de Patrones existe un número finito de clases, de tal forma que $i = 1, 2, \dots, C$ será la identificación genérica de una clase y ω_i el colectivo de patrones que la integran. Cuando se conoce una muestra $x_i = (X_1^{(i)}, X_2^{(i)}, \dots, X_n^{(i)})$ de n_i patrones de una clase se la denomina muestra controlada («labeled sample»), de tal forma que $X_{fj}^{(i)}$ sería el valor de la característica f en el patrón j que pertenece a la clase i .

Para asignar el patrón X_j a la clase que reporte menor riesgo en la clasificación, es preciso diseñar el correspondiente clasificador que de una forma más o menos directa se materializa en la obtención, normalmente con muestras controladas, de una función discriminante por cada clase. Si la función es lineal, sería:

$$g_i(X_j) = AV_0^{(i)} + \sum_{p=1}^F AV_p^{(i)} * X_{pj} \quad (1.31)$$

y para el caso cuadrático, sería:

$$g_i(X_j) = AV_0^{(i)} + \sum_{p=1}^F AV_p^{(i)} * X_{pj} + \sum_{p=1}^F \sum_{f=p}^F AM_{pf}^{(i)} * X_{pj} * X_{fj} \quad (1.32)$$

donde $AV_0^{(i)}$ recoge el peso ó umbral de la función $g_i(X)$, $AV_p^{(i)}$ es el coeficiente de influencia de la característica p en la función anterior y

$AM_{pf}^{(i)}$ recoge la influencia conjunta de las características p y f . Si ambas características no está correlacionadas, estos coeficientes teóricamente deberían ser cero. Normalmente, una discriminación superior a la cuadrática, aparte de necesitar un tamaño de la muestra controlada muy elevado y mayor tiempo de CPU en un ordenador, no ofrece mejores resultados.

La obtención de las funciones discriminantes debe efectuarse de tal forma que para el patrón $X_j^{(i)}$, perteneciendo a la clase i , el valor de la función discriminante $g_i(X_j^{(i)})$ debe ser superior a los valores $g_k(X_j^{(i)})$ para $k = 1, 2, \dots, C, i \neq k$. Por tanto, la estimación de los coeficientes $AV_0^{(i)}, AV_p^{(i)}$ y en su caso $AM_{pf}^{(i)} \forall p, f \in F, p \leq f$ debe ser tal que:

$$g_i(X_j^{(i)}) = \max_{\forall k \in C} \{g_k(X_j^{(i)})\} \text{ para } j = 1, 2, \dots, n_i, i, k = 1, 2, \dots, C \quad (1.33)$$

De acuerdo con este tipo de diseño, se asignará el nuevo patrón X_j a aquella clase para la cual (figura 1.11) $g_i(X_j)$ sea máximo. En la figura 1.12 se recogen dos típicos casos de clases lineal y no-linealmente separables. Será preciso recurrir a la discriminación cuadrática cuando no se puedan

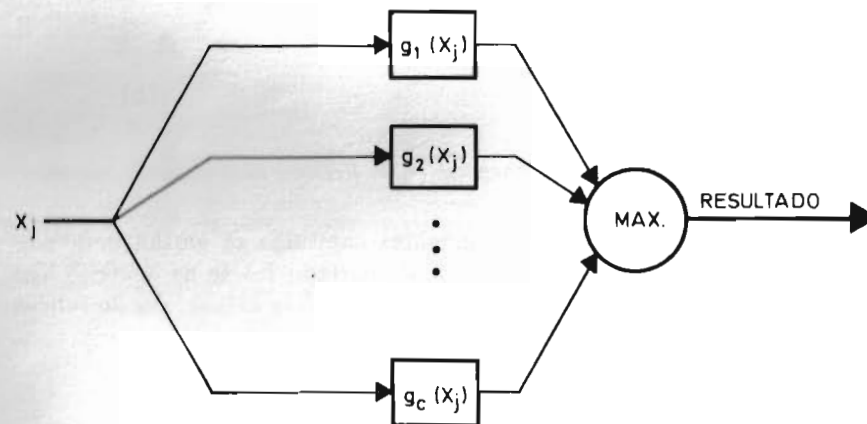


Figura 1.11.—Clasificador definido en términos de funciones discriminantes.

discriminar linealmente las muestras controladas, pero tal como se ha indicado no es preciso, normalmente, recurrir a otro tipo más elevado de discriminación. Cuando la discriminación cuadrática no separe totalmente las regiones o clases del problema, el objetivo en vez de maximizar la «distan-

cia mínima de discriminación», deberá consistir en minimizar el número de patrones mal clasificados, y en su caso la «distancia de mala clasificación».

Existen diversos métodos para la obtención de las funciones discriminantes. Su aplicación depende de las hipótesis de partida así como de la

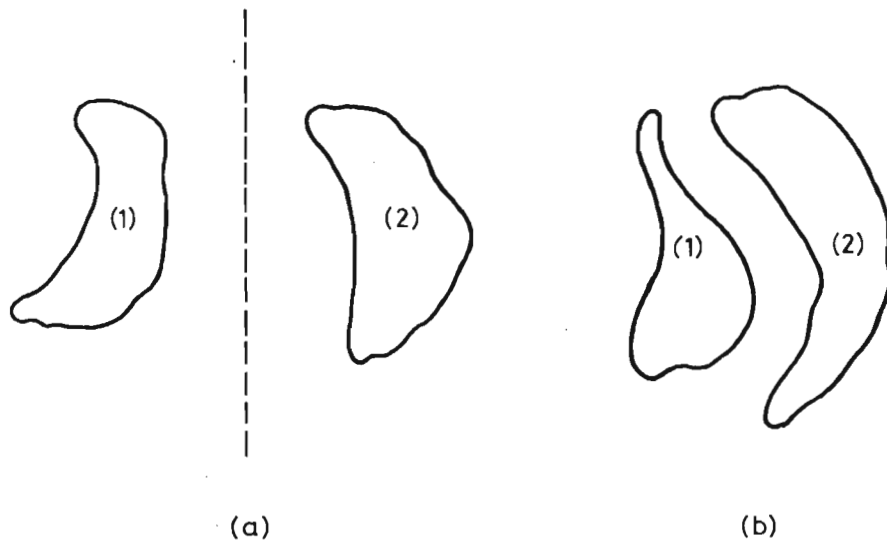


Figura 1.12.—(a) Regiones linealmente separables, (b) Regiones no linealmente separables.

disponibilidad de datos. En los siguientes capítulos se glosan detalladamente cada uno de estos métodos. En el apartado 1.3 se ha descrito ligeramente la utilización del teorema de Bayes para la estimación de funciones discriminantes.

1.4.2. Fronteras de decisión y funciones discriminantes

Cuando sólo hay $C = 2$ clases en el problema (apartado 1.3), o habiendo más clases se combinan éstas de una forma biclásica, se pueden obtener, a base de las funciones discriminantes, las denominadas «fronteras de decisión». Tal como anteriormente se ha indicado, al aplicar el patrón X_j a las funciones $g_1(X_j)$ y $g_2(X_j)$ se asignará a la clase 1 si $g_1(X_j) > g_2(X_j)$, se asignará a la clase 2 si $g_1(X_j) < g_2(X_j)$, y en caso de igualdad se asignará arbitrariamente a la clase 1.

Ahora bien, si se crea la función lineal

$$g(X_j) = g_1(X_j) - g_2(X_j) = AV_0 + \sum_{p=1}^F AV_p * X_{pj} \quad (1.34)$$

o cuadrática

$$g(X_j) = g_1(X_j) - g_2(X_j) = AV_0 + \sum_{p=1}^F AV_p * X_{pj} + \sum_{p=1}^F \sum_{f=p}^F AM_{pf} * X_{pj} * X_{fj} \quad (1.35)$$

se asignará el patrón X_j a la clase 1 si $g(X_j) \geq 0$ y en caso contrario se asignará a la clase 2. Los valores de X que hacen que $g(X) = 0$, integran lo que se denomina una frontera (y, en su caso, superficie) discriminante.

Si hubiese en el problema más de dos clases, el número de fronteras distintas sería $C * (C - 1)/2$, siendo C el número de clases. En la figura 1.13 se recoge un caso con $C = 3$ clases y, por tanto, $C * (C - 1)/2 = 3$ fronteras lineales y en la figura 1.14 otro caso con 3 fronteras cuadráticas. En ambos casos se han clasificado correctamente los patrones de la muestra controlada. Las correspondientes fronteras de decisión serían aquellas para las cuales se cumpliesen las condiciones:

$$g_{12}(X) = 0 \quad g_{13}(X) = 0 \quad g_{23}(X) = 0$$

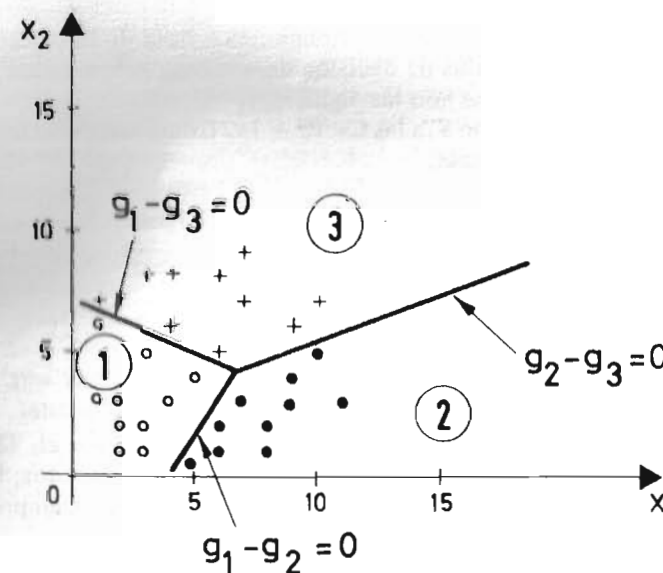


Figura 1.13.—Fronteras discriminantes lineales.

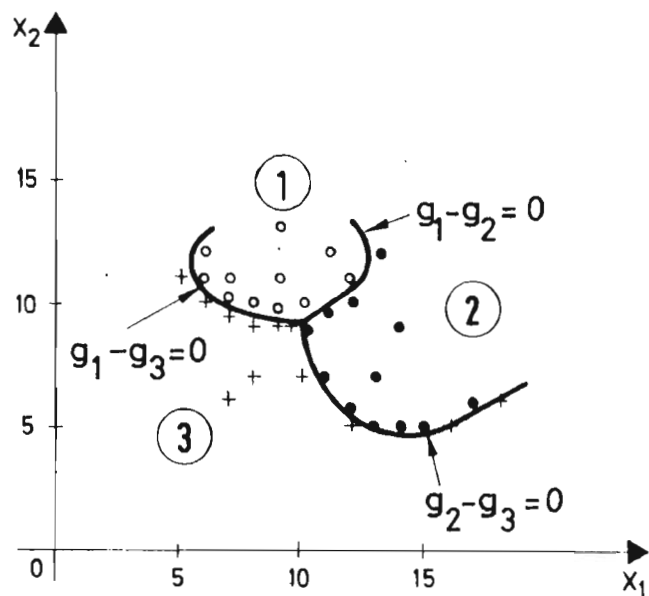


Figura 1.14.—Fronteras discriminantes cuadráticas.

Una vez creadas las fronteras discriminantes a base de la muestra controlada de patrones, las reglas de decisión de uso más general para asignar un nuevo patrón a una clase son las siguientes:

a) Aplicación del patrón X_j a las $C \cdot (C - 1)/2$ fronteras, de tal forma que si $C = 3$, se asignará a la clase:

$$\begin{array}{lll} 1 & \text{si } g_{12}(X_j) \geq 0 & \text{y } g_{13}(X_j) \geq 0 \\ 2 & \text{si } g_{12}(X_j) < 0 & \text{y } g_{23}(X_j) \geq 0 \\ 3 & \text{si } g_{13}(X_j) < 0 & \text{y } g_{23}(X_j) < 0 \end{array} \quad (1.36)$$

En general, se asignará el patrón X_j a aquella clase que en las $C \cdot (C - 1)/2$ fronteras haya alcanzado el mayor número de votos favorables. Así, si $C = 5$ (10 fronteras) y la clase 4 ha tenido tres votos (ya que $g_{14}(X_j) < 0$, $g_{24}(X_j) < 0$, $g_{45}(X_j) \geq 0$), la clase 1 dos votos, la clase 2 dos votos, la 3 dos votos y la 5 un voto, el patrón será asignado a la clase 4, ya que, en principio, es la que tiene más posibilidad de pertenencia.

Si se utilizan fronteras, y no funciones discriminantes, quizá sea ésta una de las formas más correctas de aplicar el clasificador, pero el tiempo

de ejecución de la asignación es bastante elevado. Así, si $C=10$ hay 45 fronteras y en cambio sólo 10 funciones discriminantes. Cuando sólo hay $C=2$ clases es preferible la utilización de fronteras y a partir de $C=4$ clases es preferible la utilización de funciones.

b) Otra regla de utilización de las fronteras discriminantes es la siguiente:

- 1) Aplicación del patrón X_j a la frontera discriminante (1,2). Si $g_{12}(X_j) \geq 0$, $i=1$. Si $g_{12}(X_j) < 0$, $i=2$.
- 2) Aplicación del patrón X_j a la frontera discriminante ($i,3$). Si $g_{i3}(X_j) \geq 0$, el valor de i no se cambia, y en caso contrario $i=3$.
- 3) Se continúa la operativa hasta la frontera discriminante (i,C). De tal forma que si $g_{iC}(X_j) \geq 0$, se asume que el patrón analizado pertenece a la clase i . De lo contrario, se asume que pertenece a la clase C .

En esta operativa el total de iteraciones sólo es $(C - 1)$, siendo C el número de clases analizadas. Ahora bien, existe cierto riesgo en su aplicación ya que la asignación del patrón depende, en cierto modo, del resultado del primer test que se efectúe, que en este caso es sobre la frontera (1,2).

c) Una tercera alternativa (Meisel, 1972) consiste en convertir un problema de C clases en series de problemas biclasicos, a base de efectuar dicotomías sucesivas de cada uno de los correspondientes grupos.

Si p. ej. el problema tiene $C = 8$ clases, es preciso obtener, a base del total de las muestras controladas, la frontera discriminante (a) entre las clases 1 a 4 y 5 a 8. Así mismo se obtiene la frontera (b) entre las clases 1 a 2 y 3 a 4 pero sólo utilizando las muestras controladas de las clases 1 a 4. De forma análoga se obtiene la frontera (c) entre las clases 5 a 6 y 7 a 8, utilizando las muestras controladas de las clases 5 a 8. Por fin, utilizando las correspondientes muestras controladas se obtienen las fronteras (1,2), (3,4), (5,6) y (7,8).

La utilización de las $(C - 1)$ fronteras se efectúa de la forma siguiente:

1. Aplicar el nuevo patrón X_j a la frontera (a). Si $g_a(X_j) \geq 0$ se bifurca al paso 2. Si $g_a(X_j) < 0$ se bifurca al paso 3.
2. Aplicar el nuevo patrón a la frontera (b). Si $g_b(X_j) \geq 0$, se aplica el patrón a la frontera $g_{12}(X_j)$, y en caso contrario se aplica a la frontera $g_{34}(X_j)$. Se asignará el patrón a la clase que le corresponda según sea el resultado de las fronteras simples, tal como se ha indicado en el resto del apartado.
3. Aplicar el nuevo patrón a la frontera (c). Si $g_c(X_j) \geq 0$, se aplica el patrón a la frontera $g_{56}(X_j)$, y en caso contrario se aplica a la frontera $g_{78}(X_j)$. En cualquier caso se continúa el procedimiento como en la etapa 2.

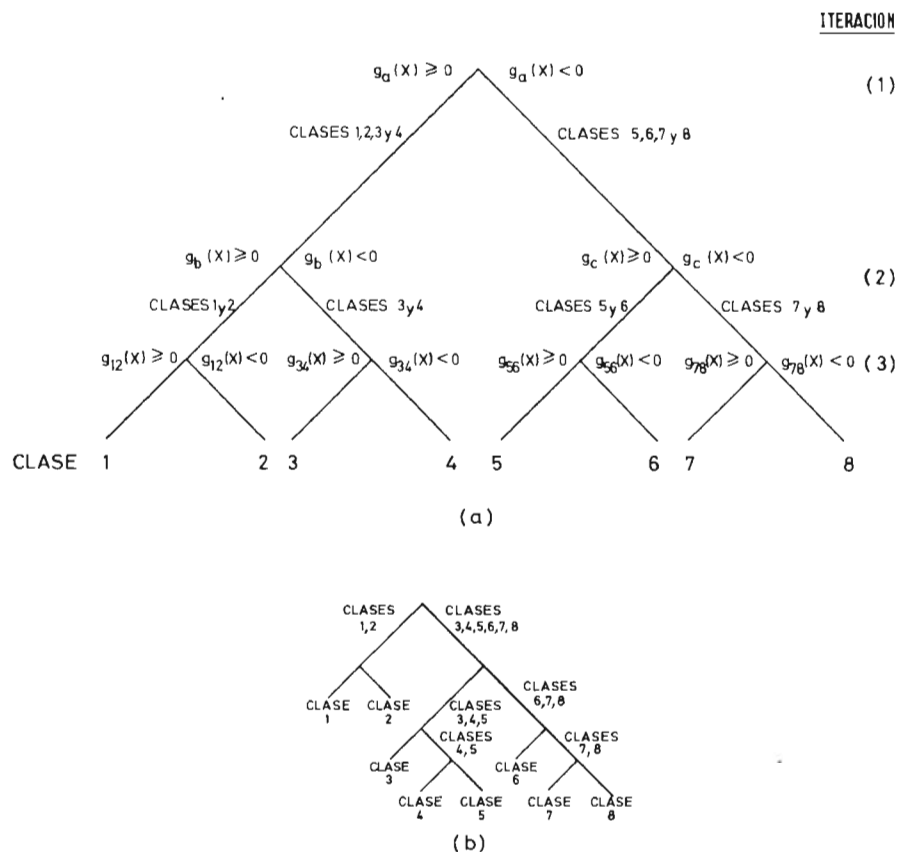
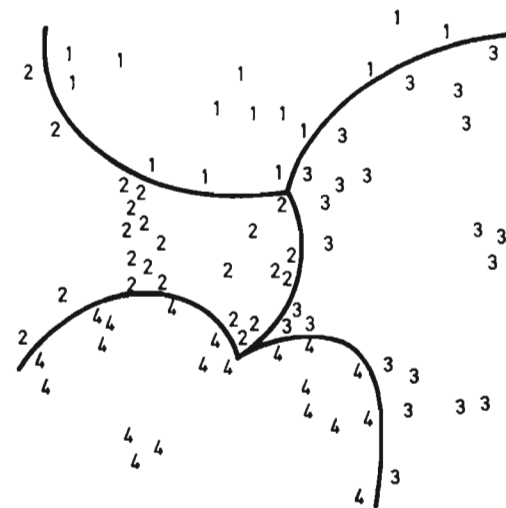


Figura 1.15.—Tipos diversos de dicotomías sucesivas en reconocimiento de patrones.

Aunque la operativa pueda parecer compleja, el ahorro de tiempo de ejecución es considerable, ya que sólo se crean $(C - 1)$ fronteras, el número de test para asignar un patrón a una clase es bajo (para $C = 8$ sólo se precisan 3 iteraciones, siendo necesarias $C = 8$ iteraciones si se utilizan funciones discriminantes) y las fronteras «superiores», al utilizar mayor número de datos, ofrecen menor riesgo de mala clasificación. En la figura 1.15a se recoge el caso que ha servido de introducción a esta operativa.

En la figura 1.15b se recoge otra modalidad de dicotomías sucesivas en la que para $C = 8$ también sólo es necesaria la obtención de $C - 1 = 7$ fronteras discriminantes, precisando de dos (clases 1 y 2), tres (clases 3 y 6) ó cuatro iteraciones para la asignación de un nuevo patrón. Si en el problema considerado, las clases 1 y 2 son más frecuentes, la dicotomía de la figura 1.15b es más interesante que la de la figura 1.15a.

La regla de decisión de dicotomías sucesivas, en comparación con la utilización de funciones discriminantes, crea un menor número de ecuaciones y precisa en su explotación un menor número de iteraciones. Pero, en ciertas circunstancias, sus resultados son menos satisfactorios. Así, en el caso recogido en la figura 1.16 se han generado las fronteras discriminantes a base de restar las correspondientes funciones cuadráticas, no habiendo discriminado incorrectamente ningún patrón de la muestra contro-

Figura 1.16.—Fronteras discriminantes generadas a base de $C = 4$ funciones discriminantes.

lada. En cambio, si se hubiese intentado la operativa de crear las fronteras discriminantes a base de dicotomías sucesivas, dado que en los primeros niveles hay que separar las muestras por grupos de clases, sólo se obtiene el mismo resultado si la frontera de cada nivel es cuadrática. De lo contrario, se discriminarían erróneamente algunos patrones.

Así en el primer nivel de la operativa de dicotomías sucesivas, sólo se puede formar una de las siguientes alternativas de discriminación (1, (2, 3, 4)), (2, (1, 3, 4)), (3, (1, 2, 4)), (4, (1, 2, 3)), ((1,2), (3,4)), ((1,3), (2,4)), ((1,4), (2,3)). Ninguna de estas posibilidades logra separar correctamente en forma cuadrática sus respectivos grupos. Para que con las dicotomías sucesivas se obtengan los mismos resultados que con las funciones discriminantes (y, por tanto, sea ventajosa su utilización), es preciso que las fronteras discriminantes a diseñar en cada uno de los niveles de la formación arborescente discriminen los correspondientes patrones al menos con el mismo resultado de las funciones discriminantes. Es una restricción más severa que la impuesta a esta otra operativa y que, desde luego, no cumple el caso recogido en la figura 1.16.

1.4.3. Problemática en el reconocimiento de patrones

En la aplicación concreta de las técnicas de reconocimiento de patrones normalmente aparecen los cuatro siguientes problemas que, de no tener una solución adecuada, pueden hacer fracasar una aplicación que en principio podía ser viable. Estos problemas consisten en la elección de la distancia o medida de similitud de los patrones, la normalización y dimensionalidad de las características y la estimación y tratamiento de distribuciones complejas de variables.

Distancia o medida de similitud

La forma más obvia para medir la similitud o la divergencia entre dos patrones o grupos de patrones consiste en obtener la distancia entre ambos. Si la distancia seleccionada es una buena medida de divergencia, se puede considerar entonces que dos patrones son tanto más similares cuanto menor sea el valor de la medida de distancia seleccionada. ¿Cómo se puede obtener la distancia entre dos patrones no unidimensionales?

En la eq (1.37) se recoge la distancia euclídea entre los patrones j y g

$$D_{jg}^2 = \|X_j - X_g\|^2 = \sum_{f=1}^F (X_{fj} - X_{fg})^2 \quad (1.37)$$

donde X_{fj} y X_{fg} son los valores de la característica f , respectivamente, en los patrones j y g , de tal forma que en principio se podría considerar que existe más probabilidad de que, p. ej., el patrón j pertenezca a la clase del patrón g en vez de a la clase del patrón h si $D_{jg}^2 < D_{jh}^2$. Dado que la medida de similitud o de divergencia es uno de los tópicos más importantes en las técnicas de reconocimiento de patrones, es fundamental una elección adecuada de la distancia a utilizar. Según se tome un tipo de distancia u otro, los resultados del problema pueden ser completamente distintos. En el capítulo 11 se recogen detalladamente los tipos de distancias más usuales, sus ventajas e inconvenientes y el apropiado campo de aplicación.

No obstante, las propiedades esenciales de una correcta medida de divergencia son las siguientes:

1. La divergencia será positiva para dos patrones distintos

$$D^2(j, g | 1, 2, \dots, F) > 0 \quad (1.38a)$$

2. La divergencia de una patrón consigo mismo es nula

$$D^2(j, j | 1, 2, \dots, F) = 0 \quad (1.38b)$$

3. La divergencia entre g y j no debe estar afectada por la denominación de los patrones

$$D^2(j, g | 1, 2, \dots, F) = D^2(g, j | 1, 2, \dots, F) \quad (1.38c)$$

4. La divergencia debe ser aditiva para características independientes

$$D^2(j, g | 1, 2, \dots, F) = \sum_{f=1}^F D^2(j, g | f) \quad (1.38d)$$

5. La divergencia no debe decrecer al añadir más características

$$D^2(j, g | 1, 2, \dots, F) \leq D^2(j, g | 1, 2, \dots, F, F+1) \quad (1.38e)$$

6. La divergencia debe ser invariante a rotaciones y transformaciones. Para ello, es preciso por un lado normalizar las características bien a base de promedio cero y varianza la unidad, bien a base de escalar sus valores con el rango en la muestra de cada característica, o con el mayor o menor de sus valores, etc. Por otro lado, para lograr la invarianza respecto a rotaciones, se puede rotar el eje de coordenadas según los autovalores de la matriz de covarianzas de la muestra.

7. La medida de divergencia seleccionada debe tener en cuenta la interdependencia de las características. De tal forma que si en una clase dos características están fuertemente correlacionadas, la divergencia de un patrón con respecto a la clase (representada, p. ej., por el vector promedio) ha de ser tanto mayor cuanto menor relación en el patrón tenga la desviación de una característica con respecto a la otra. Así, la desviación de un patrón con respecto a un grupo, motivada por la desviación correlacionada (en el mismo sentido del grupo) de *ambas* características, deberá tener menor repercusión en la medida de divergencia que si la desviación sólo es motivada por *una* característica, sin mayor correlación con la otra. En el caso recogido en la figura 1.17, el patrón 1 deberá tener mayor divergencia con respecto al grupo que el patrón 2, ya que en éste la desviación de ambas características responde a la correlación que tienen en el grupo.

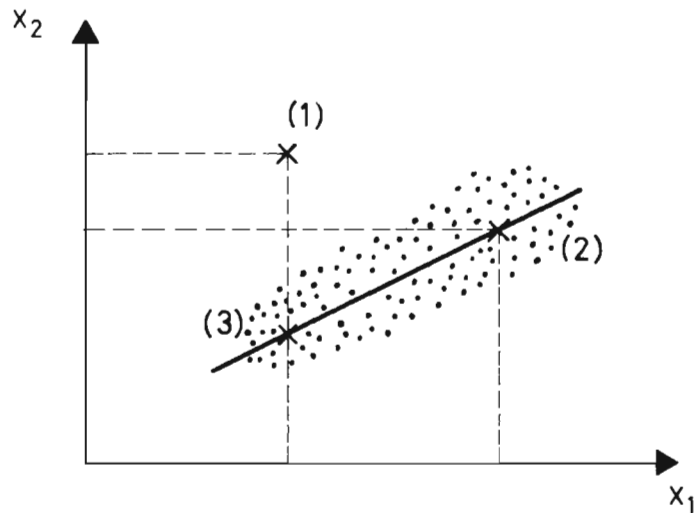


Figura 1.17.—Desviación de patrones con mayor o menor interdependencia en las características.

Normalización

Uno de los aspectos fundamentales al diseñar una medida de similitud es la elección de la unidad de medida y de ponderación de las características. Así el peso puede medirse en gramos y en kilos, la altura en metros y en centímetros, el producto nacional bruto en miles de millones, el índice del coste de vida en porcentajes, etc. El resultado de aplicar p. ej., la distancia euclídea

con un tipo u otro de escala en las medidas es completamente distinto. Por otra parte, si no se homogeneizan y ponderan las características, puede resultar que una gran desviación en una de ellas quede muy enmascarada por la desviación en otra característica, que, si en términos absolutos es fuerte, quizá relativamente no lo sea. Así, X_1 (p.ej., PNB) se mide en miles de millones y X_2 (p. ej., incremento del coste de vida) se mide en fracciones, la segunda característica no tiene prácticamente influencia en la distancia, ya que $(X_1 - X_2)^2 \approx X_1^2$.

La medida de similitud o de divergencia entre dos patrones no debería estar influenciada por el tipo concreto de escala seleccionado en las características. Agrupaciones obtenidas a base de la distancia euclídea (eq. 1.37) son invariantes a traslaciones o rotaciones de los valores de las características, pero en cambio son sensibles a transformaciones lineales o, en general, a cambios de escala de las características. En la figura 1.18 se puede observar cómo un cambio de escala en el eje de coordenadas puede provocar un diferente agrupamiento de patrones.

Las formas más usuales de lograr que la medida de similitud sea invariante a cambios de escala son las siguientes:

a) Normalizar el valor X_{fj} de la característica f del patrón j , resultante $X_{fj}/a_f \forall f \in F$ para todos los patrones de la muestra controlada, donde a_f es el rango de la característica, es decir

$$a_f = [\max \{X_{f1}^{(i)}, X_{f2}^{(i)}, \dots, X_{fn_i}^{(i)}\} - \min \{X_{f1}^{(i)}, X_{f2}^{(i)}, \dots, X_{fn_i}^{(i)}\}] \quad i = 1, 2, \dots, C \quad (1.39)$$

Este método sólo considera los valores extremos.

b) Una alternativa consiste en utilizar las varianzas de las características, de tal forma que el valor normalizado sería X_{fj}/σ_f . La estimación de σ_f a base de la muestra controlada es

$$S_f = \frac{1}{n_1 + n_2 + \dots + n_C} * \sum_{i=1}^C \sum_{j=1}^{n_i} (X_{fj}^{(i)} - \bar{X}_f^{(i)})^2 \quad (1.40)$$

donde $\bar{X}_f^{(i)}$ es el promedio de la característica f en la clase i .

Sea el caso recogido en la figura 1.18. La distancia euclídea del patrón 1 con respecto a los patrones 2 y 4 según la escala de medidas de la figura 1.18a será:

$$D_{12}^2 = \sum_{f=1}^2 (X_{f1} - X_{f2})^2 = (0 - 1)^2 + (1 - 4)^2 = 10 \quad (1.41a)$$

$$D_{14}^2 = \sum_{f=1}^2 (X_{f1} - X_{f4})^2 = (0 - 5)^2 + (1 - 0)^2 = 26 \quad (1.41b)$$

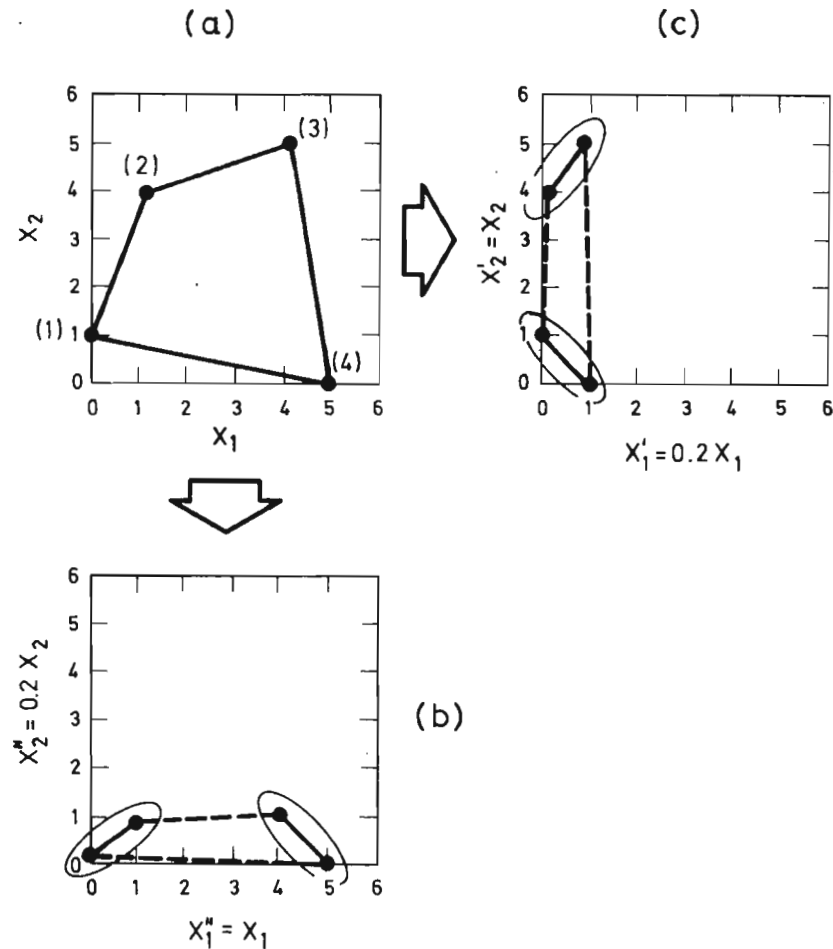


Figura 1.18.—Efectos del cambio de escala en el agrupamiento de patrones.

En cambio, si se escala en un 20 por 100 la característica 1 (Fig. 1.18c), la nueva distancia euclídea será:

$$D_{12}^2 = (0,2 * 0 - 0,2 * 1)^2 + (1 - 4)^2 = 0,2^2 * (0 - 1)^2 + (1 - 4)^2 = 9,04 \quad (1.42a)$$

$$D_{14}^2 = (0,2 * 0 - 0,2 * 5)^2 + (1 - 0)^2 = 0,2^2 * (0 - 5)^2 + (1 - 0)^2 = 2 \quad (1.42b)$$

Se puede observar que con el primer tipo de escala el patrón 2 parece más similar al patrón 1 que el patrón 4 ($D_{12}^2 < D_{14}^2$) (Fig. 1.18a) y, en cambio, con el segundo tipo de escala la relación se invierte y los patrones 1 y 4 parecen muy similares (Fig. 1.18c). ¿Cuál es el tipo de escala correcto? Ambos tipos de escala son correctos. Pero sí es errónea su utilización en la medida de similitud sin efectuar ningún tipo de normalización.

Normalizando el caso anterior a base de las varianzas $S_f^2 = \frac{1}{n} * \sum_{j=1}^n (X_{fj} - \bar{X}_f)^2 \quad \forall f \in F$, la distancia euclídea es invariante a los cambios de escala, ya que si en la fórmula

$$D_{fg}^2 = \left(\frac{X_{1j}}{S_1} - \frac{X_{1g}}{S_1} \right)^2 + \left(\frac{X_{2j}}{S_2} - \frac{X_{2g}}{S_2} \right)^2 = \frac{(X_{1j} - X_{1g})^2}{S_1^2} + \frac{(X_{2j} - X_{2g})^2}{S_2^2} \quad (1.43a)$$

se escala e. g. la característica 1 en un 20 por 100, la nueva varianza

$$S_1'^2 = \frac{1}{n} * \sum_{j=1}^n (0,2 * X_{1j} - 0,2 * \bar{X}_1)^2 = \frac{0,2^2}{n} * \sum_{j=1}^n (X_{1j} - \bar{X}_1)^2 = 0,2^2 * S_1^2 \quad (1.43b)$$

queda en la misma proporción, y por tanto la «nueva» distancia euclídea

$$D_{fg}'^2 = \frac{(X_{1j}' - X_{1g}')^2}{S_1'^2} + \frac{(X_{2j} - X_{2g})^2}{S_2^2} = \frac{0,2^2 * (X_{1j} - X_{1g})^2}{0,2^2 * S_1^2} + \frac{(X_{2j} - X_{2g})^2}{S_2^2} = D_{fg}^2 \quad (1.43c)$$

no varía, de tal forma que el resultado desde el punto de vista de la similitud entre patrones siempre es el mismo.

No obstante, los anteriores tipos de normalización no son a veces los más apropiados. Considerando la normalización con la varianza, se evita efectivamente que ciertas características puedan dominar sobre otras debido al mayor rango de sus valores numéricos. Con esta normalización se logra que la dispersión de los valores sólo sea debida a la propia divergencia de las características. Pero habría que estudiar con sumo cuidado esta normalización, ya que la divergencia podría ser debida a la presencia de

subclases en el grupo de patrones considerado. En la figura 1.19 se puede observar cómo patrones, pertenecientes a clases distintas, aparentan pertenecer a la misma clase cuando se normalizan sus características.

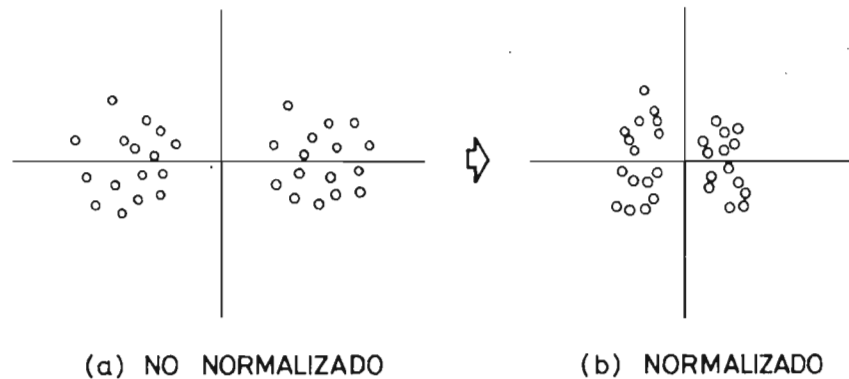


Figura 1.19.—Efecto de la normalización de características.

Dimensionalidad

Se puede visualizar fácilmente un problema de reconocimiento de patrones en dos o tres dimensiones. Si el espacio unidimensional se cubre con, p. ej., 2 puntos por unidad de intervalo (Fig. 1.20a) se precisarían en un espacio bidimensional $2^2 = 4$ puntos (Fig. 1.20b) por unidad de intervalo, y si el espacio fuese tridimensional (Fig. 1.20c) serían necesarios $2^3 = 8$ puntos. En 1, 2 ó 3 dimensiones claramente se observa que respectivamente una muestra de 2, 4 u 8 puntos no es muy representativa. Si el problema fuese de, p. ej., 10 dimensiones el número de puntos, aunque elevado ($2^F = 2^{10} = 1.024$), tampoco tendría una gran representatividad. Si $F = 20$, el número de puntos para cubrir deficientemente el amplio espectro de la muestra sería (2^{20}) superior al millón.

¿Esto significa que en problemas con un número elevado de características es preciso tener un tamaño de la muestra muy alto? La contestación depende del tipo de objetivo y método que para lograrlo se emplee. Dada la dificultad que, en numerosas ocasiones, existe para obtener muestras fiables, el problema debe orientarse en el sentido de no trabajar con toda la información disponible del patrón, sino sólo con aquella que sea *relevante*. Es decir, es preciso utilizar métodos de Selección de Características tratando de reducir al mínimo el número que de ellas se utilicen. Por otra

parte, si el problema consistiese, p. ej., en discriminar dos clases, sólo sería relevante la parte de las muestras controladas que formen la zona conflictiva. Así, en la figura 1.21 no serían necesarios los espacios (b) y (c) de las muestras controladas, ya que no contienen información relevante para el objetivo perseguido.

Del párrafo anterior se deduce que no hay ninguna norma general, rigurosamente científica, para estimar en cada caso concreto el tamaño óptimo de una muestra de patrones.

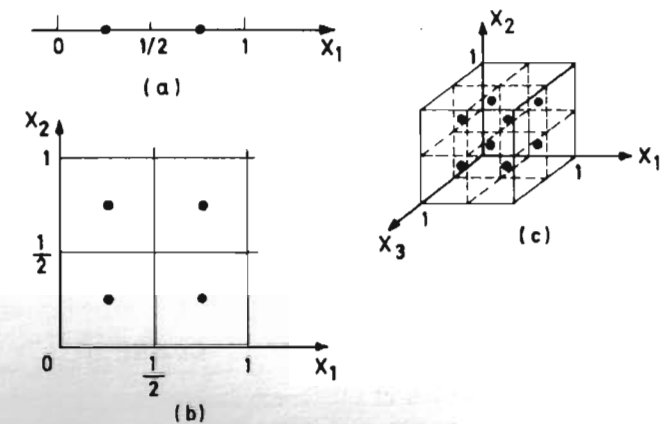


Figura 1.20.—Muestras de patrones con 1, 2 y 3 dimensiones.

En el apartado 8.5 se recoge la operativa para estimar el tamaño de la muestra, pero contemplando sólo las distribuciones marginales (es decir, bajo la hipótesis de que las características son independientes). No obstante, cada caso requiere un estudio individual.

Existen en la literatura numerosos estudios (Foley, 1971; Foley et al., 1970; Abend et al., 1969; Allais, 1964; Hughes, 1969; Ullmann, 1969) tratando de estimar la proporción mínima del número (n_i) de patrones controlados en cada clase en relación con el número (F) de características. En la mayoría de los casos aplicados, el clasificador diseñado ofrece resultados muy poco satisfactorios si la relación n_i/F (o $n_i/(F*(F+3)/2)$ en el caso de discriminación cuadrática) es inferior a 5.

Ahora bien, no sólo es interesante que la relación n_i/F sea lo más alta posible para así lograr una mayor representatividad de la muestra y, por tanto, que las características relevantes intervengan en todo su rango en la

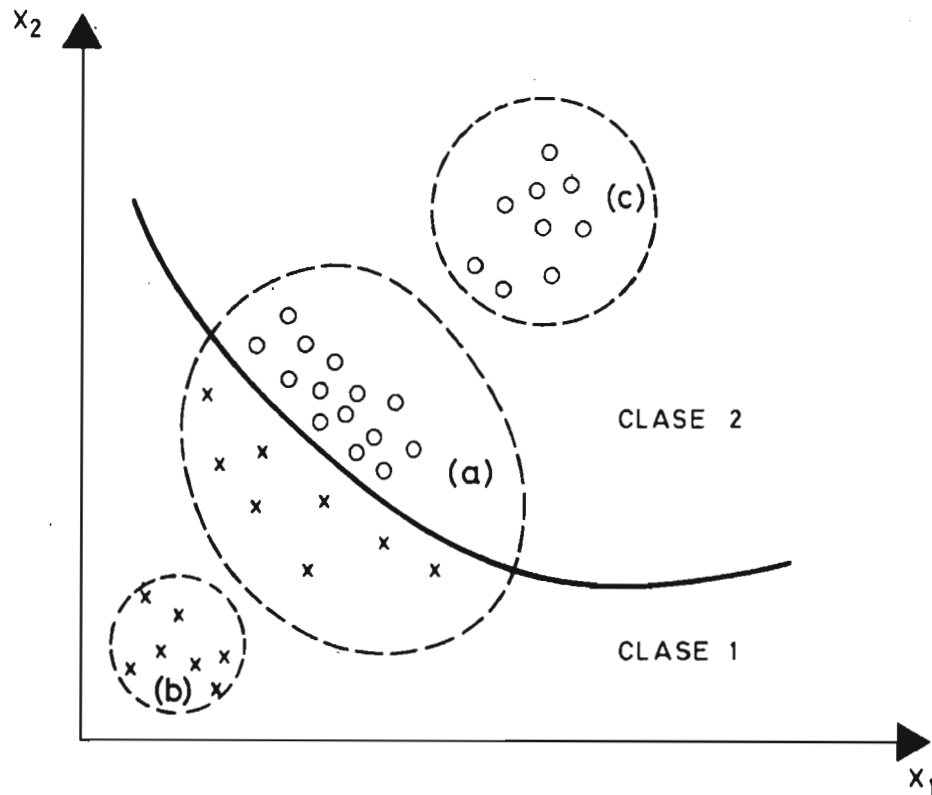


Figura 1.21.—Muestra relevante (a) y no relevante (b) y (c) de patrones.

obtención del clasificador, sino también para evitar que su variabilidad resulte enmascarada por las características que no lo son. Sea el siguiente caso hipotético. En un problema de discriminación cuadrática se han seleccionado $F = 2$ características como realmente discriminadoras. El tamaño de la muestra por clase es $n_i = 100$. La relación $n_i/F = 50$ es muy buena y los resultados serán probablemente satisfactorios. Si se añadiesen 98 características seleccionadas al azar sin ninguna relación con el objetivo de discriminar las dos clases, la nueva relación $n_i/F = 1$ sería muy baja. No obstante, se podría pensar que si las nuevas características no son relevantes y no tienen mayor relación con las otras dos características, sus coeficientes lineales y cuadráticos en la frontera discriminante serían nulos. En cambio, dado el escaso número de patrones disponibles, es muy probable que estos coeficientes no sean nulos, enmascarando por tanto la

importancia de los otros, y produciendo resultados muy poco satisfactorios en la explotación del clasificador obtenido.

En resumen, es preciso, como paso previo a diseñar un clasificador, seleccionar las características relevantes para el caso objeto en estudio y elegir una muestra representativa de la zona conflictiva, no esperando grandes resultados si la relación anterior es inferior a 3 ó 5.

Distribuciones complejas

La muestra controlada de cada clase no siempre se distribuye de una forma compacta (Fig. 1.22a), sino que la muestra puede ser multi-modal (Fig. 1.22b). Un ejemplo tipo lo constituye la clase de la primera letra del abecedario, de tal forma que en la muestra puede haber patrones de la subclase «A» y patrones de la subclase «a». Otras veces las muestras forman «hojas» (Fig. 1.22c) en lugar de conglomerados, otras las muestras forman circunferencias concéntricas con radio diferente (Fig. 1.22d), etc. A medida que crece el número F de características, se incrementa la complejidad de las distribuciones.

Una gran mayoría de las técnicas de reconocimiento de patrones se efectúa bajo la base de que las características siguen la distribución normal unidimensional (Fig. 1.22e) o multidimensional (para $F = 2$ ver figura 1.22f), cuando en numerosas ocasiones la distribución es exponencial, binomial, de Poisson, etc. En los casos en que se consideran estas hipótesis, se asume que las características son independientes.

La escasa generalidad que en este aspecto tienen las técnicas de reconocimiento de patrones es debida a la complejidad que supone el tratamiento de patrones con dimensiones elevadas. No obstante, en los capítulos 7 y 8 se recoge una operativa que, en numerosas ocasiones, ofrece resultados muy satisfactorios en aquellos casos en los que la distribución seguida por la muestra controlada es sensiblemente diferente de la distribución multivariante normal.

1.4.4. Muestra de diseño y muestra de validación

Una vez diseñado un clasificador a base de una muestra controlada, existe cierta tendencia a estimar el % de patrones que clasificará erróneamente, utilizando los mismos patrones de la muestra controlada. Implícitamente se está considerando que en la muestra se encuentra toda la información significativa del problema. Dada la dificultad de conseguir muestras muy representativas, la validación del clasificador efectuada de

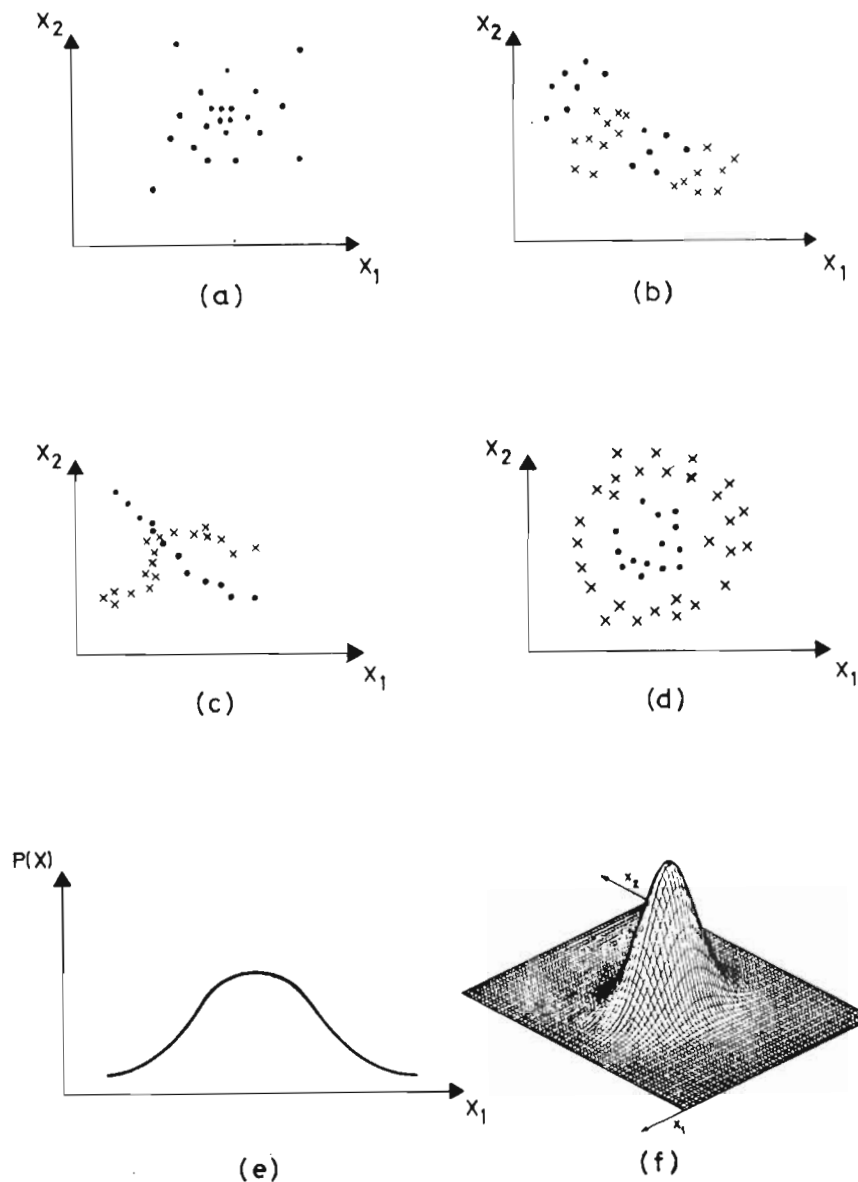


Figura 1.22.—Tipos diversos de agrupación de patrones.

esta forma es sesgada ya que, en realidad, el clasificador ha sido diseñado sólo para clasificar la muestra controlada.

Por ello es preferible, aun a costa de disminuir la relación n_i/F , dividir la muestra controlada en muestra de diseño («training field») y muestra de validación («test field»). La primera muestra sirve para construir el clasificador y la segunda se utiliza para estimar el resultado de su aplicación a nuevos patrones cuya clase se desconoce.

Aunque existen otras formas (Meisel, 1972) de abordar el problema, la de mayor aplicación consiste en dividir la muestra controlada en los dos conjuntos indicados. El problema consiste en estimar el tamaño de la muestra de validación. Si es pequeño la validación no será significativa, y si es grande provocará que la muestra de diseño sea inoperante. Tradicionalmente, y sin ninguna razón específica, se utiliza el 25 por 100 de la muestra controlada como muestra de validación. Un tratamiento más completo consiste en analizar, a base de esta muestra, el % de error del clasificador y a continuación rediseñarlo con el 100 por 100 de la muestra controlada. No se conoce el resultado final, pero muy probablemente el % de error será inferior al ya estimado.

Otro problema típico consiste en distribuir la muestra de validación entre las clases que componen el caso en estudio. Una forma razonable de determinar el número de patrones de validación de cada clase, una vez estimado el tamaño de esta muestra, consiste en distribuir ésta en proporción a la probabilidad a priori de cada clase, que muy frecuentemente consiste en dividir la muestra en proporción al número de patrones que hay de cada clase en la muestra total.

1.5. TECNICAS DE RECONOCIMIENTO DE PATRONES

Existe cierta tendencia a considerar que el diseño de un clasificador es el único objetivo de las denominadas técnicas de reconocimiento de patrones. Es muy importante la obtención de una regla de decisión para asignar un nuevo patrón a una clase determinada, pero no es el objetivo único. Se podría indicar que cualquier objetivo que se pretenda con patrones multidimensionales forma parte de la aplicación e investigación de las técnicas de reconocimiento de patrones.

Así uno de los grandes problemas recogidos en el caso del apartado 1.3 consiste en una adecuada Selección de Características de tal forma que recojan la información relevante del problema. Hay fundamentalmente dos tipos de técnicas de Selección de Características. Ante la presentación in-

tuitiva de las características de un sistema, un tipo de técnicas consiste en seleccionar un número determinado de ellas, de acuerdo con su grado de discriminación. El otro tipo no exige necesariamente aceptar ciertas características, sino que, p. ej., a través de transformaciones lineales de las antiguas, trata de estimar nuevas características, menores en número, pero que siendo factores comunes a las anteriores, recojan la discriminación intrínseca de las clases (Hartman, 1967 y Meisel, 1972).

Cuando la muestra disponible no es controlada e incluso se desconocen las clases del problema, puede ser útil tratar de obtener los grupos o clases naturales en que, de acuerdo con su similitud, se agrupan los patrones de la muestra. Se denominan técnicas de «análisis cluster» (Parte II) al conjunto de métodos y técnicas que describen y localizan estas agrupaciones.

Su utilidad fundamental es la siguiente:

1. Mejor interpretación de la estructura de los patrones analizando las causas intrínsecas de la agrupación de los mismos.
2. Ayudan a descubrir las posibles subclases subyacentes en la muestra controlada de una clase. Así la clase puede ser la primera letra del abecedario y las letras «A» y «a» las subclases.
3. Pueden desconocerse las clases a que pertenecen los patrones que integran la muestra. Pero, en cambio, puede que, de acuerdo con las características seleccionadas, si se encuentran grupos o clusters «naturales» éstos correspondan a las clases del problema. Estas clases sirven para diseñar un clasificador de futuros patrones cuya clase de pertenencia es desconocida. Es decir, esta aplicación del análisis cluster consiste en estimar las clases de pertenencia de tal forma que, así controladas las muestras, se pueda diseñar un clasificador de reconocimiento de patrones.

La Parte I trata de los métodos de diseño de un clasificador. Estos métodos se agrupan de la forma siguiente:

1. Estructura probabilística conocida. Se conocen las probabilidades a priori $P(\omega_i)$ y condicional $P(X|\omega_i)$, así como los estadísticos θ de estas distribuciones. A base de la misma metodología se tratan las técnicas basadas en las leyes normal, exponencial y binomial.
2. Estructura probabilística no conocida totalmente.

A) Métodos supervisados.—Estos métodos parten de que se dispone de una muestra controlada de cada clase, así como de que se conoce la probabilidad a priori $P(\omega_i)$, aunque si ésta se desconoce se podría estimar a base de la muestra controlada.

a) *Métodos paramétricos.*—En estos métodos se conoce también la probabilidad condicional $P(X|\omega_i)$ pero, en cambio, se desconocen los co-

rrespondientes estadísticos θ . Se efectúa el diseño del clasificador a base del método de máxima verosimilitud (en el cual se considera que los estadísticos θ son fijos pero desconocidos) y del método de la estimación de Bayes (considerando que los estadísticos son una variable cuya distribución es conocida).

b) *Métodos no paramétricos.*—En estos métodos no se conoce la función de probabilidad condicional $P(X|\omega_i)$. Se efectúa el diseño del clasificador a base de estimar esta probabilidad o sustituirla en su caso por transformaciones ortogonales y ortonormales.

c) *Métodos directos.*—Estos métodos estiman directamente la probabilidad a posteriori $P(\omega_i|X)$ de que un patrón pertenezca a una clase. Su objetivo consiste en obtener las correspondientes funciones discriminantes, prescindiendo de la distribución condicional $P(X|\omega_i)$. Los métodos más comúnmente empleados son: Método de la distancia mínima («Nearest-Neighbour rule»), método del incremento fijo («fixed increment procedure») y método del gradiente. El inconveniente de estos métodos estriba en que convergen muy lentamente en problemas que no tienen gran separación lineal en las clases, ofreciendo una solución «buena» pero sin garantizar su optimalidad.

d) *Análisis discriminante.*—Este método parte de las mismas hipótesis que los métodos directos, pero tiene la ventaja de que, en la creación de las funciones discriminantes lineales o cuadráticas, optimizan un determinado criterio impuesto por el usuario. En el marco del análisis discriminante se recoge el sistema DISMIP que actualmente es el que mejores resultados está proporcionando.

e) *Análisis secuencial.*—Como variante de los métodos paramétricos y no-paramétricos, el reconocimiento secuencial de patrones analiza una a una las características del patrón, de tal forma que el número medio de características analizadas que se precisan para efectuar la asignación del patrón es inferior al número que precisan los métodos anteriores.

B) Métodos no supervisados.—Estos métodos tratan los casos en los que la muestra no es controlada. La diferencia con las técnicas de análisis cluster consiste en que se conocen las clases a discriminar así como la función de probabilidad condicional $P(X|\omega_i)$. Se desconoce, normalmente, la probabilidad a priori $P(\omega_i)$. El clasificador se diseña de una forma iterativa, ya que tampoco se conocen los estadísticos θ , y para estimarlos es preciso, a su vez, estimar a qué clase corresponden los patrones que integran la muestra.

Aunque en los párrafos que anteceden se hayan presentado las diversas técnicas de reconocimiento de patrones con una sistemática que puede

parecer que son independientes, es preciso puntualizar que la división entre Selección de Características, Análisis Cluster y Diseño de Clasificadores es puramente artificial, ya, p. ej., el diseño de un clasificador no es más que la transformación unidimensional de las características en principio contempladas. Lo mismo ocurre con los diversos métodos presentados. Todos tienen la misma finalidad y su aplicación no es única en el sentido de que en un mismo problema se utilizan muy frecuentemente varios tipos de métodos.

REFERENCIAS

- ABEND, T.; CHANDRASEKARAN, B., y HARLEY, T. J.: *Comments on the Mean Accuracy of Statistical Pattern Recognizers*, «IEEE Trans. on Info. Theory», 15, pp. 420-423 (1969).
- ALLAIS, D. C.: *Selection of Measurements for Prediction*, «Stanford Electron Labs.», SELL-64-115, Stanford (1964).
- ANDERSON, T. W.: *Introduction to Multivariate Statistical Analysis*, J. Wiley (1958).
- ANDREWS, H. C.: *Introduction to Mathematical techniques in Pattern Recognition*, J. Wiley (1972).
- BALL, G. B.: *Data Analysis in the social sciences*, «Proceedings FJCC», Las Vegas (1965).
- BELLMAN, R. E.: *Adaptive Control Processes*, «Princeton University Press» (1961).
- BONGARD, N.: *Pattern Recognition*, Spartan, N. Y. (1970).
- DUDA, R. O.: *Elements of Pattern Recognition* en «Adaptive, Learning and Pattern Recognition Systems» (ed. Mendel y Fu), Academic Press, pp. 3-33 (1970).
- DUDA, R. O. y HART, P. E.: *Pattern Classification and Scene Analysis*, J. Wiley (1973).
- FOLEY, D.: *The Probability of error on the design set as a function of the sample size and feature size*, «Ph. D. dissertation», Rome Air Development Center. Report RADCR-71-171 (1971).
- FOLEY, D.; SAMMON, J. D., y PROCTOR, A.: *Considerations of dimensionality versus sample size*, «IEEE Symposium on Adaptive Processes», Austin Texas (1970).
- FU, K. S.: *Sequential Methods in Pattern Recognition and Machine Learning*, Academic Press (1969).
- FUKANAGA, K.: *Introduction to Statistical Pattern Recognition*, Academic Press (1972).
- GIULIANO, V. E.: *How we find patterns*, «Intern. Science and Technology» (1967).
- GOSE, E. A.: *Introduction to biological and mechanical pattern recognition* en «Methodologies of Pattern Recognition» (ed. Watanabe), Academic Press (1969).
- HARTMAN, H. H.: *Modern Factor Analysis*, Chicago University Press (1967).
- HIGHLEYMAN, W. H.: *The design and analysis of pattern recognition experiments*, «Bell Systems Technique Journal», 41, pp. 723-744 (1962).
- HO, Y. C., y AGRAWALA, A. K.: *On Pattern Classification Algorithms. Introduction and Survey*, «Proceedings IEEE», 56, pp. 836-862 (1968).
- KANAL, L. N.: *Pattern Recognition*, Thompson, Washington (1968).
- KRISHNAIAH, P. R.: *Multivariate Analysis*, Academic Press (1966).
- KRISHNAIAH, P. R.: *Multivariate Analysis II*, Academic Press (1969).
- MASON, C. J. W.: *Pattern Recognition Bibliography*, «IEEE Sys. Sci. and Cyb», (1970).
- MEISEL, W. S.: *Computer-oriented approaches to Pattern Recognition*, Academic Press (1972).

- MENDEL, J. M., y FU, K. S. (eds): *Adaptive, Learning and Pattern Recognition Systems*, Academic Press (1970).
- NAGY, G.: *State of the Art in Pattern Recognition*, «Proceedings IEEE», 56, pp. 836-861 (1968).
- NILSSON, N. J.: *Learning Machines*, McGraw-Hill (1965).
- RÍOS, S.: *Análisis de Decisiones*, 1.^a Parte, Madrid (1973).
- SABESTYEN, G.: *Decision-Making Processes in Pattern Recognition*, MacMillan, N. Y. (1962).
- SAYRE, K. M.: *Recognition. An Study in the Philosophy of Artificial Intelligence*, University of Notre Dame Press, Ohio (1965).
- UHR, L.: *Pattern Recognition, Learning and Thought*, Prentice Hall (1973).
- ULLMANN, J. R.: *Experiments with the N-Tuple Method of Pattern Recognition*, «IEEE Trans. comp.», 18, pp. 1135-1137 (1969).
- ULLMANN, J. R.: *Pattern Recognition Techniques*, Butterworths, Londres (1973).
- VERHAGEN, C. J. D. M.: *Some general remarks about Pattern Recognition*, «Pattern Recognition», 7, pp. 109-116 (1975).
- WATANABE, S.: *Knowing and Guessing*, J. Wiley (1969a).
- WATANABE, S. (ed.): *Methodologies of Pattern Recognition*, Academic Press (1969b).