

8. Reconocimiento De Patrones

Podemos definir este tipo de acciones como la “búsqueda de la estructura de los datos”. En este ámbito los seres vivos hemos demostrado tener muchas ventajas: somos capaces de reconocer rostros, letras manuscritas (en el caso humano) y otros patrones, aún cuando la información de la que disponemos para reconocerlos no siempre sea idéntica. Somos capaces de reconocer el rostro de un ser querido ya sea en una foto en blanco y negro o a color, incluso a veces un tanto borrosa.

Ejemplo 1: Capacidad de reconocer caracteres manuscritos.

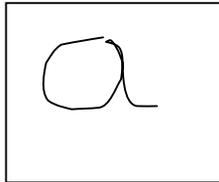


Figura 8.1. “a” manuscrita

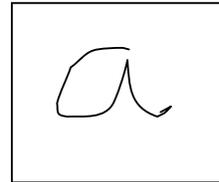


Figura 8.2. “a” manuscrita

El ser humano es capaz de reconocer que las figuras 1 y 2 corresponden a dos formas de escribir la letra “a” minúscula. ¿Sería capaz un sistema computacional de hacerlo?.

Este tipo de acciones representa un problema de difícil solución. Para la implementación de algoritmos de clasificación, que sean capaces de reconocer determinada información como la correspondiente a un patrón específico, se han propuesto diferentes alternativas, y se han obtenido interesantes resultados con aquellas basadas en redes neuronales y lógica difusa.

Sea $\{x_i\}_{i=1}^n$ n objetos caracterizados por p indicadores, es decir $x_i = \{x_i^1, x_i^2, \dots, x_i^p\}$. Deseamos ser capaces, a través de un algoritmo, de encontrar una partición de este conjunto en c “regiones de decisión” o “clusters”. Los objetos que pertenecen a un mismo cluster deben presentar similitudes; por otro lado, si tomamos dos objetos de diferentes clusters, estos deben presentar diferencias.

El número de clusters “ c ” es una cantidad que, dependiendo del tipo de problema, es un dato conocido o desconocido. Si, por ejemplo, nos enfrentamos a un proceso de reconocimiento de dígitos, sabemos que el número de clusters es 10. Si en cambio nos encontramos efectuando un estudio del riesgo que representan ciertos componentes químicos dentro de los procesos de contaminación atmosférica, probablemente desconoceremos cuántos niveles es importante distinguir. Dicha cantidad desconocida es el número de clusters de este problema.

El método de clasificación que se verá a continuación es conocido como “ c -means”, y permite ordenar los objetos en c clusters. El “ c -means” necesita que uno

determine previamente en cuántos clusters se debe clasificar la información. Veremos tanto una versión “crisp” como una difusa de dicho método.

Ejemplo 2: Para introducir este procedimiento, veremos un ejemplo muy utilizado en la literatura, “la mariposa”. Nos interesa proponer métodos de clasificación para los puntos que se observan en la Figura 8.3.

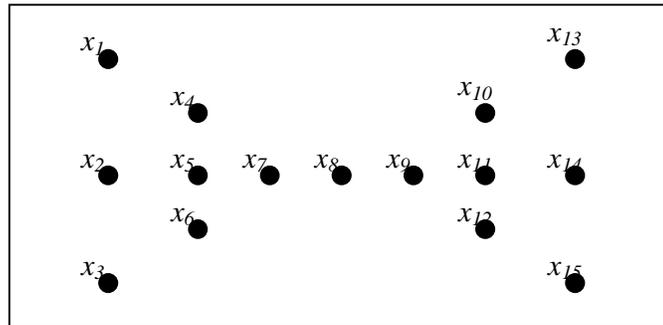


Figura 8.3. La Mariposa

Supongamos que consideramos como lo más adecuado clasificar los puntos que se observan en la Figura 8.3. en dos clusters. Una posible solución sería la que se indica en la Figura 8.4. Esta sería una clasificación “crisp”, es decir, cada punto pertenece totalmente a uno o a otro cluster.

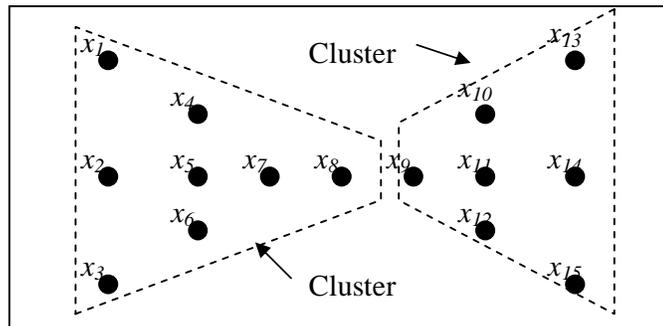


Figura 8.4. Clasificación “crisp” de la mariposa

Como se ve en la Figura 8.4, la mariposa es simétrica en torno al punto x_8 . Este debe asignarse, bajo este algoritmo de clasificación, a uno u otro cluster. Nos podemos preguntar ¿es la pertenencia absoluta a uno u otro cluster una forma adecuada de clasificar a x_8 ? Si consideramos que la función de pertenencia a un cluster sólo toma valores 1 o 0, ¿el valor de pertenencia 1 al cluster 1 es un valor que representa la pertenencia de x_8 a dicho cluster?

Una solución más acorde con la estructura de los datos la da una clasificación difusa. En ella los valores de pertenencia a uno u otro cluster estarían dentro del rango $[0,1]$. En la Figura 8.5 se observa una posible solución difusa para el problema de la mariposa, en lo que se refiere a los valores de pertenencia al cluster 1.

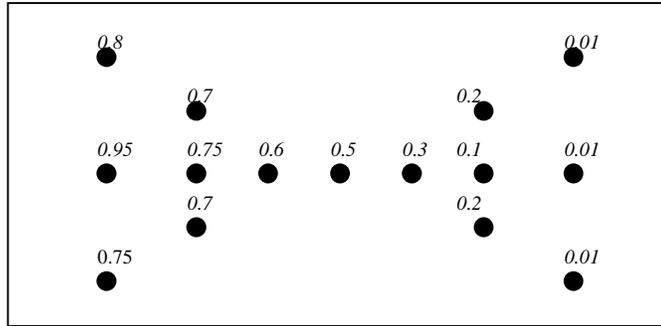


Figura 8.5. Clasificación difusa de la mariposa, caso cluster 1

Podemos plantear el problema de clasificación de la siguiente manera: Sea el conjunto de datos $X = \{x_1, \dots, x_n\} \subseteq \mathfrak{R}^p$

- Para el caso "crisp": Queremos clasificar estos n elementos en c subconjuntos "crisp" $\{S_j\}_{j=1}^c$. La pertenencia de x_i a S_j queda caracterizada por la función indicador $u_{S_j}(x) : X \rightarrow \{0,1\}$.
- Para el caso difuso: Queremos clasificar estos n elementos en c subconjuntos difusos $\{F_j\}_{j=1}^c$. La pertenencia de x_i a F_j queda indicada por la función de membresía $\mu_{F_j}(x) : X \rightarrow [0,1]$.

Clasificar los elementos en c clusters es en el fondo hacer una partición adecuada de X en c subconjuntos. Luego, para una clasificación "crisp" se define la matriz $U = \{a_{ij}\}$ como una *partición- c* "crisp" si esta cumple las propiedades descritas en las expresiones 1, 2 y 3:

$$a_{ij} = u_{S_j}(x_i) \quad \forall j=1\dots n, \forall i=1\dots c \quad (\text{exp 1})$$

$$\sum_{i=1}^c a_{ij} = 1 \quad \forall j=1\dots n \quad (\text{exp 2})$$

$$0 \leq \sum_{j=1}^n a_{ij} \leq n \quad \forall i=1\dots c \quad (\text{exp 3})$$

Ejemplo 3: Sea $X = \{x_1, x_2, x_3\}$. Podemos definir una partición-2 de X como los siguientes dos subconjuntos: $X_1 = \{x_1, x_2\}$ y $X_2 = \{x_3\}$. Esta partición se puede representar con la matriz que se ve en la expresión 4. Otra forma de efectuar una partición-2 la podemos ver en la expresión 5.

$$U = \begin{bmatrix} x_1 & x_2 & x_3 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{exp 4})$$

$$U = \begin{matrix} & x_1 & x_2 & x_3 \\ \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{matrix} \end{matrix} \quad (\text{exp 5})$$

De igual manera podemos definir una *partición-c* difusa, $\tilde{U} = \{\tilde{a}_{ij}\}$, como aquella que cumple con las propiedades definidas en las expresiones 6, 7 y 8.

$$\tilde{a}_{ij} = \mu_{S_j}(x_i) \quad \forall i=1\dots c, \forall j=1\dots n \quad (\text{exp 6})$$

$$\sum_{i=1}^c \tilde{a}_{ij} = 1 \quad \forall j=1\dots n \quad (\text{exp 7})$$

$$0 \leq \sum_{j=1}^n \tilde{a}_{ij} \leq n \quad \forall i=1\dots c \quad (\text{exp 8})$$

Ejemplo 4: Tomando el mismo conjunto X del ejemplo 3, una forma de hacer una *partición-2* difusa de este conjunto es la siguiente:

$$\tilde{U} = \begin{matrix} & x_1 & x_2 & x_3 \\ \begin{matrix} 1 & 0,8 & 0,3 \\ 0 & 0,2 & 0,7 \end{matrix} \end{matrix} \quad (\text{exp 9})$$

Es muy importante dentro de este proceso de clasificación definir a qué llamaremos el centro del cluster. Este se define como el objeto alrededor del cual los elementos del cluster se concentran. En general no corresponde a ninguno de los elementos de X ni se conoce su valor. Para efectos de este algoritmo los designaremos como $\{v_i\}_{i=1}^c$.

Podemos medir la similitud entre los elementos de X y los centros de los cluster a través de una norma vectorial general. Sea esta $\|\cdot\|_G$. Con ella es posible plantear el proceso de clasificación como **un problema de optimización, en el cual buscamos la partición que minimice esta medida de distancia entre los elementos x_k y los centros de los clusters**. La idea es que los elementos que pertenecen a un determinado cluster deben estar lo más cerca posible de los centros de estos clusters, y lo más alejados del centro de otros clusters. De ahí que ubicarlos en un determinado cluster es un problema de escoger la partición adecuada. Luego, podemos enfrentar el problema como un proceso de optimización en búsqueda de la partición óptima.

Veremos cómo se plantea este tipo de proceso de clasificación tanto para el caso difuso como para el caso crisp.

1) Caso Crisp:

Podemos calcular la medida de distancia acumulada entre los elementos de la partición y los centros de los clusters $v_j, j = 1, \dots, c$. Sea esta d , y se calcula como se ve en la expresión 10.

$$d = \sum_{j=1}^c \sum_{x_k \in S_j} \|x_k - v_j\|_G \quad (\text{exp 10})$$

Luego, podemos plantear el mencionado problema de optimización como la minimización de esta distancia, lo que en el fondo será determinar la partición U que minimice d .

$$\min d(U, \{v_j\}_{j=1}^c) \quad (\text{exp 11})$$

Como una forma de estimar el valor de los centros de los clusters se calcula cada v_j como el promedio de los valores de los elementos que pertenecen al cluster j . Para obtener este promedio sumamos los valores de los elementos que pertenecen al cluster y lo dividimos por el número de elementos del cluster, es decir, por la cardinalidad del subconjunto S_j , como se ve en la expresión 12.

$$v_j = \frac{1}{|S_j|} \sum_{x_k \in S_j} x_k \quad \text{para } j = 1, \dots, c \quad (\text{exp 12})$$

En este caso crisp, podemos expresar la pertenencia o no pertenencia a un determinado cluster a través de los elementos de la matriz de particiones $U = \{a_{ij}\}$. Así, la distancia acumulada d , los valores v_j y el problema de optimización quedan como se ven en las expresiones 13 y 14 respectivamente.

$$\min d(U, \{v_j\}_{j=1}^c) = \sum_{j=1}^c \sum_{i=1}^n a_{ij} \|x_i - v_j\|_G \quad (\text{exp 13})$$

$$v_j = \frac{1}{\sum_{i=1}^n a_{ij}} \sum_{i=1}^n a_{ij} x_i \quad \text{para } j = 1 \dots c \quad (\text{exp 14})$$

Escojamos una norma general para resolver este proceso de clasificación. Tomemos aquella definida en la expresión 15. Para ello consideremos a Y un vector en \mathfrak{R}^p y a G como una matriz en $\mathfrak{R}^p \times \mathfrak{R}^p$ simétrica definida positiva.

$$\|Y\|_G^2 = Y^T \cdot G \cdot Y \quad (\text{exp 15})$$

Cuando se escoge $G=I$, la identidad, esta norma general coincide con la norma euclidiana. Consideremos esta norma como el valor particular de la norma general para el problema de optimización descrito en las expresiones 13 y 14.

Ahora, si analizamos este problema, vemos la presencia de variables discretas dentro de las expresiones 13 y 14. Estas son los valores de la matriz U . Como los a_{ij} solo toman valores en el conjunto $\{0,1\}$, para encontrar la partición que minimice d es necesario resolver un problema de optimización combinatorial muy difícil. Por ejemplo, para $c=10$ y $n=25$, existen del orden de 10^{18} particiones diferentes que deben ser exploradas para poder encontrar el óptimo.

2) Caso Difuso:

En el caso difuso, la pertenencia de un elemento a un cluster viene dada por los valores que tome la función de membresía del cluster, los cuales serían nuestro objetivo de búsqueda en el proceso de optimización. Estos valores de pertenencia corresponden a los elementos de la matriz de particiones $\tilde{U} = \{\tilde{a}_{ij}\}$. Siguiendo un proceso análogo al definido para el caso crisp, y eligiendo la misma norma general que en el caso anterior, el problema de optimización queda como se puede ver en las expresiones 16 y 17.

$$\min d(\tilde{U}, \{v_j\}_{j=1}^c) = \sum_{j=1}^c \sum_{i=1}^n (\tilde{a}_{ij})^m \|x_i - v_j\|_G \quad (\text{exp 16})$$

$$v_j = \frac{1}{\sum_{i=1}^n \tilde{a}_{ij}} \sum_{i=1}^n (\tilde{a}_{ij})^m x_k \quad \text{para } j = 1 \dots c \quad (\text{exp 17})$$

El coeficiente $m \geq 1$ que aparece en las expresiones 16 y 17 es un valor opcional que permite aumentar la convergencia de los métodos iterativos que se utilizan para solucionar este problema. Permite aumentar la influencia de los elementos x_k que tengan valores de membresía cercanos a 1, y disminuir la de aquellos que tengan valores de membresía cercanos a 0.

Como los valores \tilde{a}_{ij} pueden tomar valores continuos dentro del rango $[0,1]$ (ver la expresión 6), este problema de optimización es mucho más sencillo de resolver. Derivando d con respecto a sus variables de optimización, se obtiene la expresión 18.

$$\tilde{a}_{ij} = \frac{\left(\frac{1}{\|x_i - v_j\|_G} \right)^{\frac{1}{m-1}}}{\sum_{j=1}^c \left(\frac{1}{\|x_i - v_j\|_G} \right)^{\frac{1}{m-1}}} \quad \text{para } j = 1 \dots c, i = 1 \dots n \quad (\text{exp 18})$$

Con la expresión 18, que representa $c \times n$ ecuaciones, más la expresión 17, que representa otras c ecuaciones, componemos un sistema de ecuaciones con las cuales podemos determinar los valores de la matriz \tilde{U} y de los $\{v_j\}_{j=1}^c$. Sin

embargo, este sistema no tiene solución analítica. Para encontrar estos valores, se propone el siguiente algoritmo iterativo:

a) Elegir valores para las siguientes variables:

c , el número de clusters, que debe cumplir $2 \leq c \leq n$

m , el coeficiente de aceleración de la convergencia, que debe cumplir $1 < m < \infty$

G , una matriz de $p \times p$, que debe cumplir con ser simétrica y definida positiva.

U_0 , una matriz de partición difusa cualquiera de $c \times n$.

k , un contador, que inicializamos en 0.

ε , el parámetro que determina el criterio de detención del algoritmo, debe cumplir $\varepsilon > 0$

b) Calcular $\{v_j\}_{j=1}^c$ con los valores de \tilde{U}_k , utilizando la expresión 17. Sean estos valores $\{v_j^{(k)}\}_{j=1}^c$.

c) Calcular \tilde{U}_{k+1} con los valores de $\{v_j^{(k)}\}_{j=1}^c$ y las ecuaciones de la expresión 18,

siempre que $x_i \neq v_j^{(k)}$. Si $x_i = v_j^{(k)}$, entonces $\tilde{\alpha}_{ij} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$

d) Con una norma matricial $\| \cdot \|_M$, calcular $\Delta = \|\tilde{U}_{k+1} - \tilde{U}_k\|_M$. Si $\Delta > \varepsilon$, se asigna $k = k + 1$ y se vuelve al punto b) de este proceso. Si $\Delta < \varepsilon$, se detiene el proceso. Siendo ε el criterio definido para detener el proceso.

Ejemplo 5: El siguiente ejemplo fue extraído del libro de H.J. Zimmermann "Fuzzy Sets Theory", publicado en 1996 por Allied Publishers Limited. Tomando el mismo ejemplo de la mariposa, se definió la partición inicial que se ve en la expresión 19.

$$\tilde{U}_0 = \begin{bmatrix} 0,854 & 0,146 & 0,854 & \dots & 0,854 \\ 0,146 & 0,854 & 0,146 & \dots & 0,146 \end{bmatrix}_{2 \times 15} \quad (\text{exp 19})$$

En el problema de la mariposa, se distinguen dos clusters, por lo cual $c=2$. Se tomó G como la matriz identidad, para así trabajar con la norma euclidiana, se escogió $\varepsilon=0.01$ y $m=1.25$. Con estos valores, tras 6 iteraciones, se obtuvieron los resultados, para un cluster, que se pueden ver en la Figura 8.6. También se efectuó este proceso iterativo con $m=2$, y el resultado de este proceso se observa en la Figura 8.7.

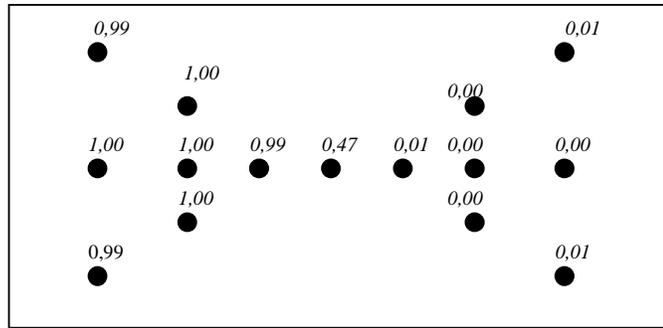


Figura 8.6. Resultados del proceso de clasificación con $m=1.25$.

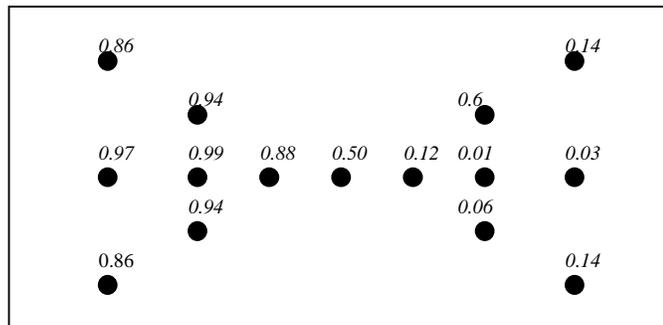


Figura 8.7. Resultados del proceso de clasificación con $m=2$.