Clasificadores Bayesianos

En aquellos casos en los cuales los patrones de algunas clases presenten, ya sea una cercanía notable o una dispersión significativa con respecto a su media, en conveniente abandonar la hipótesis determinística a cambio de una más de tipo estadístico.

El primer caso se puede deber, entre otras cosas, al parecido entre patrones o el tipo de rasgos usados para describir dichos patrones.

El Segundo caso puede deberse no solo al parecido entre patrones o el tipo de rasgos usados, sino a su alta variación

El teorema de Bayes

El famoso teorema deBayes se puede expresar de varias maneras. Una de ellas es la siguiente.

Si:

$p(C_i\|x)$:	Representa la probabilidad de que un patrón x pertenezca a la clase C_i
$p(x\|C_i)$	Representa la probabilidad de que dada la clase C_i, el valor de la variable aleatoria sea precisamente x.
$p(C_i)$	Es la probabilidad a priori de que se manifieste un elemento de la clase C_i
$p(x)$	Es la probabilidad a priori de que se manifieste un patrón x.

$$p(C_i|x) = \frac{p(x|C_i)*p(C_i)}{p(x)}$$

Nótese que el elemento p(x) es el mismo para todas las clases, por lo tanto puede eliminarse

Para determinar la clase en la cual debe clasificarse un patrón dado x, se puede usar el término a la izquierda de la ecuación anterior, o el término a la derecha. En el primer caso, dado un patrón x, su clase se puede determinar como sigue:

$$x \rightarrow C_i \text{ si y solo si } p(C_i|x) > p\left(C_j\middle|x\right), \forall\, i \neq j, j = 1,2, \dots, N$$

En otras palabras, el patrón x es asignado a la clase C_i en el caso que su probabilidad a *posteriori* sea máxima

En el segundo caso, dado un patrón x, su clase se puede determinar como:

$$x \rightarrow\ C_i\ sii\ p(C_i|x) * p(C_i) > p\left(C_j\middle|x\right) * p(C_j), \forall\ i\ \neq j, j = 1,2, \dots, N$$

En otras palabras, el patrón x es asignado a la clase C_i para la cual su probabilidad a *priori* sea la máxima.

El clasificador Bayesiano con distribución tipo Gaussiana :

En la mayoría de los casos practicos, para el caso de vectores con *n* rasgos, las distribuciones de probabilidad de cada una de las *N* clases de patrones, presentan un comportamiento normal o Gaussioano, esto es:

$$p(x|C_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|C_i|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right), i = 1,2,\ldots,N$$

Las funciones de discriminación toman pues la forma:

$$p(C_i) * p(x|C_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|C_i|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right), i = 1,2,\ldots,N$$

Al tomar los logaritmos neperianos sobre esta expresión se obtiene:

$$\ln p(C_i) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|c_i| - \frac{1}{2}(x-\mu_i)^T c_i^{-1}(x-\mu_i)$$

Sin pérdida de generalidad y para simplificar la discusión, supondremos que las *N* clases son equivocadas a *priori*. La expresión da cada función de discriminación fd_i, después de eliminar los términos redundantes queda como:

$$fd_i(x) = -\frac{1}{2}x^T c_i^{-1} x + x^T c_i^{-1}\mu_i - \frac{1}{2}\mu_i{}^T c_i^{-1}\mu_i - \frac{1}{2}\ln|c_i|, i = 1,2, \dots, N$$

Enseguida se discuten dos casos que se pueden presentar:

Caso 1: Las matrices de covarianza de las *N* clases son todas diferentes $c_1 \neq c_2 \neq \cdots \neq c_N$, y en general, sin elementos nulos.

En este caso las funciones discriminantes son no literales. Todas las combinaciones cuadráticas posibles de los elementos del vector de rasgos se hacen manifiestas

Caso 2: Las matrices de covarianza de las N clases son todas iguales: $c_1 \neq c_2 \neq \cdots \neq c_N$, y en general, sin elementos nulos.

Este caso es el que más se presenta en la realidad. Se parte del supuesto que el vector de rasgos presenta un comportamiento estadístico, similar para todas las clases de objetos. EN este caso se pueden eliminar todos los términos no lineales de la ecuación anterior, quedando la siguiente función discriminante para cada clase.

$$fd_i(x) = x^T c_i^{-1} \mu_i - \frac{1}{2} \mu_i{}^T c_i^{-1} \mu_i, i = 1,2, \dots, N$$

Estimación de los parámetros de la matriz de covarianza:

El principal problema en el diseño de clasificadores a *priori* es la estimación de las funciones de densidad de probabilidad $p(x|C_i), i = 1,2,\dots,N$, a partir de un conjunto de muestras o realizaciones, uno por clase. P ara una clase dada C_i , y suponiendo que se dispone de *M* muestras o realizaciones, primero se estima la esperanza matemática de la clase a través del cálculo del vector promedio:

$$\mu_i = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \frac{1}{M}\sum_{j=1}^{M} x_{1,j} \\ \frac{1}{M}\sum_{j=1}^{M} x_{2,j} \\ \vdots \\ \frac{1}{M}\sum_{j=1}^{M} x_{n,j} \end{pmatrix}$$

La matriz de covarianza de la clase C_i puede estimarse como:

$$c_i = \frac{1}{M}\sum_{j=1}^{M} x * x_j^T - \mu * \mu^T$$

Ejemplo:

Supóngase el siguiente universo de patrones y sus clases:

$$\left\{\begin{pmatrix}4\\3\end{pmatrix},\begin{pmatrix}5\\4\end{pmatrix},\begin{pmatrix}4\\4\end{pmatrix},\begin{pmatrix}3\\4\end{pmatrix},\begin{pmatrix}5\\3\end{pmatrix}\right\}=C_1$$

$$\left\{\begin{pmatrix}7\\4\end{pmatrix},\begin{pmatrix}7\\5\end{pmatrix},\begin{pmatrix}6\\4\end{pmatrix},\begin{pmatrix}6\\5\end{pmatrix},\begin{pmatrix}5\\5\end{pmatrix}\right\}=C_2$$

Entrenamiento

De acuerdo a lo discutido en el entrenamiento del clasificador a *priori* de Bayes es necesario calcular los vectores de medias y de las matrices de covarianzas de cada clase.

$$\mu_1 = \frac{1}{5}\left[\begin{pmatrix}4\\3\end{pmatrix}+\begin{pmatrix}5\\4\end{pmatrix}+\begin{pmatrix}4\\4\end{pmatrix}+\begin{pmatrix}3\\4\end{pmatrix}+\begin{pmatrix}5\\3\end{pmatrix}\right]=\begin{pmatrix}4.2\\3.6\end{pmatrix}$$

$$\mu_2 = \frac{1}{5}\left[\begin{pmatrix}7\\4\end{pmatrix}+\begin{pmatrix}7\\5\end{pmatrix}+\begin{pmatrix}6\\4\end{pmatrix}+\begin{pmatrix}6\\5\end{pmatrix}+\begin{pmatrix}5\\5\end{pmatrix}\right]=\begin{pmatrix}6.2\\4.6\end{pmatrix}$$

De igual manera:

$$c_1 = \frac{1}{5}\left[\begin{pmatrix}4\\3\end{pmatrix}\begin{pmatrix}4 & 3\end{pmatrix}+\begin{pmatrix}5\\4\end{pmatrix}\begin{pmatrix}5 & 4\end{pmatrix}+\begin{pmatrix}4\\4\end{pmatrix}\begin{pmatrix}4 & 4\end{pmatrix}+\begin{pmatrix}3\\4\end{pmatrix}\begin{pmatrix}3 & 4\end{pmatrix}+\begin{pmatrix}5\\3\end{pmatrix}\begin{pmatrix}5 & 3\end{pmatrix}\right] - \begin{pmatrix}4.2\\3.6\end{pmatrix}\begin{pmatrix}4.2 & 3.6\end{pmatrix} = \begin{pmatrix}18.2 & 15.0\\15.0 & 13.2\end{pmatrix} - \begin{pmatrix}17.64 & 15.12\\15.12 & 12.96\end{pmatrix} = \begin{pmatrix}0.56 & -0.12\\-0.12 & 0.24\end{pmatrix}$$

$$c_2 = \frac{1}{5}\left[\begin{pmatrix}7\\4\end{pmatrix}\begin{pmatrix}7 & 4\end{pmatrix} + \begin{pmatrix}7\\5\end{pmatrix}\begin{pmatrix}7 & 5\end{pmatrix} + \begin{pmatrix}6\\4\end{pmatrix}\begin{pmatrix}6 & 4\end{pmatrix} + \begin{pmatrix}6\\5\end{pmatrix}\begin{pmatrix}6 & 5\end{pmatrix} + \begin{pmatrix}5\\5\end{pmatrix}\begin{pmatrix}5 & 5\end{pmatrix}\right] - \begin{pmatrix}6.2\\4.6\end{pmatrix}\begin{pmatrix}6.2 & 4.6\end{pmatrix} = \begin{pmatrix}39.0 & 28.4\\28.4 & 21.4\end{pmatrix} - \begin{pmatrix}28.44 & 28.52\\28.52 & 21.16\end{pmatrix} = \begin{pmatrix}0.56 & -0.12\\-0.12 & 0.24\end{pmatrix}$$

Note como las matrices de covarianza de las dos clases son iguales. Luego entonces, de acuerdo a lo discutido, las funciones de discriminación pueden ser calculadas en $fd_i(x)$

$$fd_1(x) = \begin{pmatrix}x_1 & x_2\end{pmatrix}\begin{pmatrix}0.56 & -0.12\\-0.12 & 0.24\end{pmatrix}^{-1}\begin{pmatrix}4.2\\3.6\end{pmatrix} - \frac{1}{2}\begin{pmatrix}4.2 & 3.6\end{pmatrix}\begin{pmatrix}0.56 & -0.12\\-0.12 & 0.24\end{pmatrix}^{-1}\begin{pmatrix}4.2\\3.6\end{pmatrix} = 12x_1 + 21x_2 - 63$$

$$fd_2(x) = \begin{pmatrix}x_1 & x_2\end{pmatrix}\begin{pmatrix}0.56 & -0.12\\-0.12 & 0.24\end{pmatrix}^{-1}\begin{pmatrix}6.2\\4.6\end{pmatrix} - \frac{1}{2}\begin{pmatrix}6.2 & 4.6\end{pmatrix}\begin{pmatrix}0.56 & -0.12\\-0.12 & 0.24\end{pmatrix}^{-1}\begin{pmatrix}6.2\\4.6\end{pmatrix} = 17x_1 + 27.667x_2 - 116.334$$

Si se supone que la probabilidad de aparición de un elemento de cualquiera de las dos clases C_1 y C_2 es la misma, esto es si $p(C_1) = p(C_2) = 0.5$, entonces un vector x será clasificado como:

$x \rightarrow C_1$, si y solo si $fd_1(x) > fd_2(x)$, esto es si $p(x|c_1) > p(x|c_2)$

Y

$x \rightarrow C_2$, si y solo si $fd_1(x) < fd_2(x)$, esto es si $p(x|c_1) < p(x|c_2)$

Prueba del clasificador

Supóngase el vector $x = \begin{pmatrix} 5.1 \\ 4.1 \end{pmatrix}$. De acuerdo a lo visto sustituya los valores de x en $fd_n(x)$,

Para cada clase:

$$fd_1(x) = 84.3$$

$$fd_2(x) = 83.8$$

Debido a que $fd_2(x) < fd_1(x)$ entonces el patrón x es clasificado en la clase C_1