UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Razonamiento probabilístico: representación y inferencia

Albert Sanchis

DSIC

Departament de Sistemes Informàtics i Computació

1. Problema

Clasificar tres tipos de flores: *Setosa, Versicolor* y *Virgínica*.

Variables aleatorias de interés:

Tipo de flor : $C \in \{SETO, VERS, VIRG\}$

Número de Pétalos : $N \in \{3, 4, 5\}$

Distribución conjunta:

T	N	P
$SETO$	3	0,13
$SETO$	4	0,13
$SETO$	5	0,07
$VERS$	3	0,27
$VERS$	4	0,07
$VERS$	5	0,00
$VIRG$	3	0,00
$VIRG$	4	0,20
$VIRG$	5	0,13
Suma:		1,000

2. Inferencia probabilística

A partir de la distribución conjunta podemos calcular la probabilidad de cualquier *suceso* (*proposición*) mediante aplicación de:

La regla suma:

$$P(x) = \sum_{y} P(x, y)$$

La regla producto:

$$P(x, y) = P(x)\, P(y \mid x)$$

En general no es necesario conocer la tabla completa de probabilidades conjuntas para calcular la probabilidad de un suceso dado.

3. Teorema de Bayes

El *teorema de Bayes* permite actualizar nuestro conocimiento sobre una hipótesis y tras observar una nueva evidencia x:

$$P(y \mid x) \;=\; \frac{P(x, y)}{P(x)} \;=\; P(y)\,\frac{P(x \mid y)}{P(x)}$$

De otra forma: $P(y \mid x)$ es la probabilidad de que se produzca el efecto y tras observar que se ha producido la causa x.

Inferencia probabilística: problema de las flores

	SETO	VERS	VIRG	3	4	5
$P(x)$:						

$$P(y \mid x) = \frac{P(x,y)}{P(x)} = P(y)\,\frac{P(x|y)}{P(x)}$$

x	3	4	5
SETO			
VERS			
VIRG			

x	SETO	VERS	VIRG
3			
4			
5			

Distribución conjunta

T	N	P
$SETO$	3	0,13
$SETO$	4	0,13
$SETO$	5	0,07
$VERS$	3	0,27
$VERS$	4	0,07
$VERS$	5	0,00
$VIRG$	3	0,00
$VIRG$	4	0,20
$VIRG$	5	0,13
Suma:		1,000

Inferencia probabilística: problema de las flores (solución)

	SETO	VERS	VIRG	3	4	5
$P(x)$:	0.33	0.34	0.33	0.40	0.40	0.20

$$P(y \mid x) = \frac{P(x,y)}{P(x)} = P(y)\,\frac{P(x|y)}{P(x)}$$

x	3	4	5
SETO	0.394	0.394	0.212
VERS	0.794	0.206	0.000
VIRG	0.000	0.606	0.394

x	SETO	VERS	VIRG
3	0.325	0.675	0.000
4	0.325	0.175	0.500
5	0.350	0.000	0.650

Distribución conjunta

T	N	P
$SETO$	3	0,13
$SETO$	4	0,13
$SETO$	5	0,07
$VERS$	3	0,27
$VERS$	4	0,07
$VERS$	5	0,00
$VIRG$	3	0,00
$VIRG$	4	0,20
$VIRG$	5	0,13
Suma:		1,000

4. La regla de decisión de Bayes

La ***regla de decisión de Bayes*** predice el efecto que producirá una causa x eligiendo, entre un conjunto de efectos posibles $\mathcal{C}$, uno de màxima ***probabilidad a posteriori*** (de la observación de la causa):

$$c^*(x) = \arg\max_{c \in \mathcal{C}} \; P(c \mid x)$$

Probabilidad de error (efecto predicho distinto del realmente producido)

$$P(\text{error} \mid x) = 1 - P(c^*(x) \mid x)$$

Ninguna otra elección mejoraría esta probabilidad de error!

La regla de decisión de Bayes: Problema de las flores

	SETO	VERS	VIRG	3	4	5
$P(x)$:	0.33	0.34	0.33	0.40	0.40	0.20

$$P(y \mid x) = \frac{P(x,y)}{P(x)} = P(y)\,\frac{P(x|y)}{P(x)}$$

x	SETO	VERS	VIRG
3	0.325	0.675	0.000
4	0.325	0.175	0.500
5	0.350	0.000	0.650
x	3	4	5
$c^\star(x)$			
$P(\text{error} \mid x)$			

Distribución conjunta

T	N	P
$SETO$	3	$0{,}13$
$SETO$	4	$0{,}13$
$SETO$	5	$0{,}07$
$VERS$	3	$0{,}27$
$VERS$	4	$0{,}07$
$VERS$	5	$0{,}00$
$VIRG$	3	$0{,}00$
$VIRG$	4	$0{,}20$
$VIRG$	5	$0{,}13$
Suma:		$1{,}000$

La regla de decisión de Bayes: Problema de las flores (Solución)

	SETO	VERS	VIRG	3	4	5
$P(x)$:	0.33	0.34	0.33	0.40	0.40	0.20

$$P(y \mid x) = \frac{P(x,y)}{P(x)} = P(y)\,\frac{P(x|y)}{P(x)}$$

x	SETO	VERS	VIRG
3	0.325	0.675	0.000
4	0.325	0.175	0.500
5	0.350	0.000	0.650
x	3	4	5
$c^\star(x)$	VERS	VIRG	VIRG
$P($error $\mid x)$	0.325	0.500	0.350

Distribución conjunta

T	N	P
$SETO$	3	0,13
$SETO$	4	0,13
$SETO$	5	0,07
$VERS$	3	0,27
$VERS$	4	0,07
$VERS$	5	0,00
$VIRG$	3	0,00
$VIRG$	4	0,20
$VIRG$	5	0,13
Suma:		1,000

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Regla de Bayes: otra versión

Por el teorema de Bayes, la regla de Bayes se puede reescribir como:

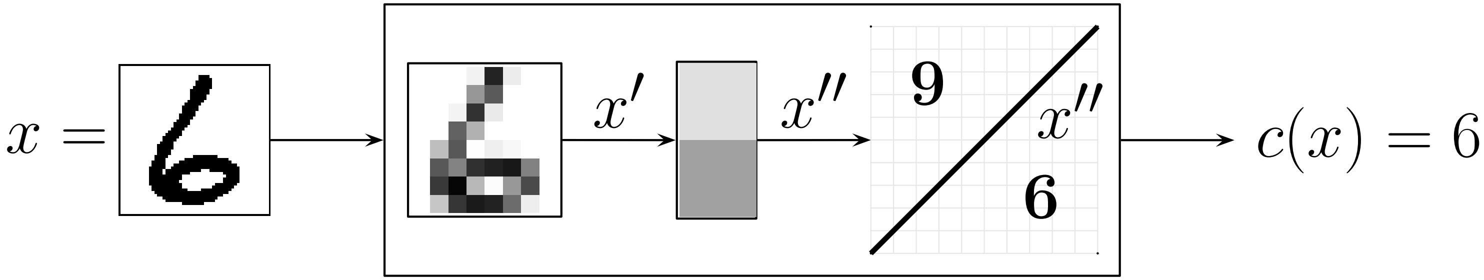
$$
\begin{aligned}
c^*(x) &= \arg\max_{c \in \mathcal{C}} \; P(c \mid x) \\
&= \arg\max_{c \in \mathcal{C}} \; P(c) \, \frac{p(x \mid c)}{p(x)} \\
&= \arg\max_{c \in \mathcal{C}} \; P(c) \, p(x \mid c)
\end{aligned}
$$

donde $P(c)$ es la ***probabilidad a priori*** del efecto c; y $p(x \mid c)$ es la ***(densidad de) probabilidad*** de que x sea la causa del efecto c.

5. Reconocimiento de Formas y Apr. Automático

El ***Reconocimiento de Formas*** y ***Aprendizaje Automático*** estudian sistemas capaces de aprender y predecir a partir de datos.

Un problema clásico es la construcción de clasificadores para objetos percibidos con sensores apropiados; p.e. un OCR de 6 o 9:



La aproximación convencional se basa en la regla de Bayes:

$$c^*(x) = \arg\max_{c \in \mathcal{C}} \; P(c \mid x) = \arg\max_{c \in \mathcal{C}} \; P(c)\, p(x \mid c)$$

donde $P(c \mid x)$, o $P(c)$ i $p(x \mid c)$, se aprenden a partir de ejemplos.

6. Ejercicio

Un problema clásico de decisión consiste en clasificar flores de la familia *Iris* en tres clases; *setosa, versicolor* y *virgínica*, en base a los tamaños de sus pétalos y sépalos (x).
Para ello se han calculado sendos histogramas de las superficies de los pétalos de una muestra de 50 flores de cada clase. Normalizando estos histogramas, se ha estimado la siguiente distribución de tamaños de pétalos para cada clase (c)

	tamaño de los pétalos en cm^2											
$P(x \mid c)$	$<$1	1	2	3	4	5	6	7	8	9	10	$>$10
SETO	0.90	0.10	0	0	0	0	0	0	0	0	0	0
VERS	0	0	0	0.20	0.30	0.32	0.12	0.06	0	0	0	0
VIRG	0	0	0	0	0	0	0.08	0.12	0.24	0.14	0.20	0.22

Asumiendo que las clases son equiprobables, calcular:

1. Las probabilidades a posteriori $P(c \mid x),\ c \in \{\text{SETO}, \text{VERS}, \text{VIRG}\}$, para una flor cuyo tamaño de pétalos es $x =$7 cm^2
2. La decisión óptima de clasificación de esta flor y la probabilidad de que dicha decisión sea errónea
3. Repetir los calculos anteriores, asumiendo que las probabilidades a priori son: $P(\text{c=SETO}) = 0{,}3,\ P(\text{c=VERS}) = 0{,}5,\ P(\text{c=VIRG}) = 0{,}2$

Soluciones: 1. 0.0, 0.33, 0.67; 2. VIRG, 0.33; 3. 0.0, 0.56, 0.44;