Sistemas Inteligentes

Escuela Técnica Superior de Informática

Universitat Politècnica de València

Tema B2.T6:
Programación dinámica.
Algoritmo *Forward*.
Algoritmo de Viterbi.

Índice

- 1 *Clasificación sintáctico-estadística* ▷ 1
- 2 Algoritmo Forward ▷ 5
- 3 Algoritmo de Viterbi ▷ 11
- 4 Apéndice: cálculos en clasificación estadística (resumen) ▷ 20

Clasificación sintáctico-estadística

Suponemos C clases de objetos, representados como cadenas de Σ^+. Planteamiento similar al de la clasificación estadística en el caso vectorial:

- ***Probabilidad a priori*** de una clase c: $P(c), 1 \leq c \leq C$
- ***Probabilidad condicional*** de la clase c: $P(y \mid M_c)$, función de probabilidad que modela la distribución de las cadenas de c en Σ^* mediante un modelo de Markov M_c.
- ***Probabilidad a posteriori*** de la clase c: $P(c \mid y)$

$$P(c \mid y) = \frac{P(y \mid M_c)P(c)}{P(y)} \quad \text{donde} \quad P(y) = \sum_{c'=1}^{C} P(y \mid M_{c'})P(c')$$

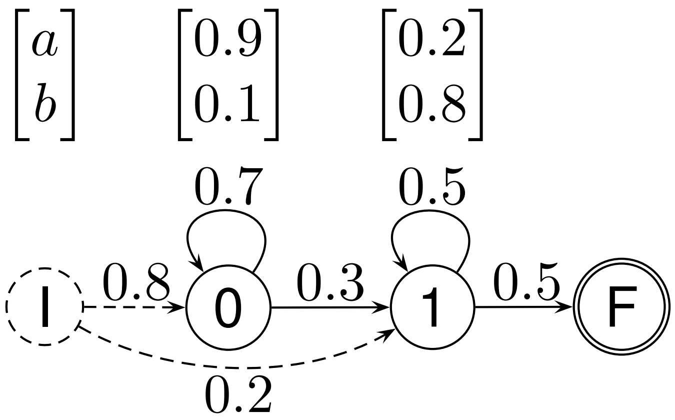
- ***Regla de clasificación***: Una cadena $y \in \Sigma^+$ se asigna a la clase $\hat{c}(y)$:

$$\hat{c}(y) = \underset{1 \leq c \leq C}{\text{argmax}}\, P(c \mid y)$$

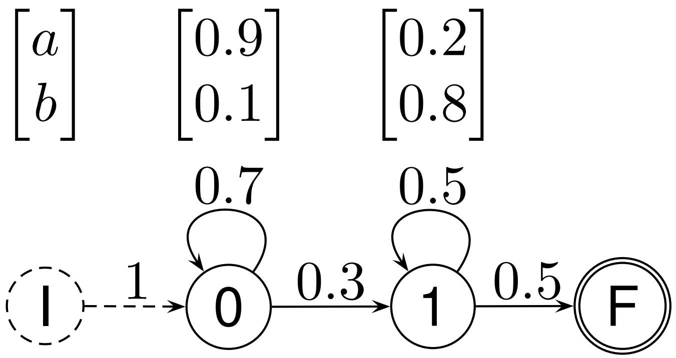
Clasificación sintáctico-estadística: ejercicio

Se tiene un problema de clasificación en dos clases (A y B) de objetos representados mediante cadenas de símbolos en el alfabeto $\Sigma = \{a, b\}$. Las probabilidades a priori de las clases son $P(A) = 0.6$ y $P(B) = 0.4$. Las funciones de probabilidad condicional de las clases vienen caracterizadas por los modelos de Markov:

Modelo M_A: $P(y \mid A) = P(y \mid M_A)$



Modelo M_B: $P(y \mid B) = P(y \mid M_B)$



Sea $y = \texttt{aab}$. Halla $P(y \mid c)$ y $P(c \mid y)$ para $c = A, B$, y clasifica y por mínimo error.

Ejercicio: solución

$$
\begin{aligned}
P(y \mid M_A) \\
&= P(aab, q_1q_2q_3 = 001 \mid A) \\
&+ P(aab, q_1q_2q_3 = 011 \mid A) \\
&+ P(aab, q_1q_2q_3 = 111 \mid A) \\
&= (0.8 \cdot 0.9)\ (0.7 \cdot 0.9)\ (0.3 \cdot 0.8)\ 0.5 \\
&+ (0.8 \cdot 0.9)\ (0.3 \cdot 0.2)\ (0.5 \cdot 0.8)\ 0.5 \\
&+ (0.2 \cdot 0.2)\ (0.5 \cdot 0.2)\ (0.5 \cdot 0.8)\ 0.5 \\
&= 0.0544 + 0.0086 + 0.0008 = \ 0.0638
\end{aligned}
$$

$$
\begin{aligned}
P(y \mid M_B) \\
&= P(aab, q_1q_2q_3 = 001 \mid B) \\
&+ P(aab, q_1q_2q_3 = 011 \mid B) \\
&= (1 \cdot 0.9)\ (0.7 \cdot 0.9)\ (0.3 \cdot 0.8)\ 0.5 \\
&+ (1 \cdot 0.9)\ (0.3 \cdot 0.2)\ (0.5 \cdot 0.8)\ 0.5 \\
&= 0.0680 + 0.0108 \\
&= 0.0788
\end{aligned}
$$

$$
P(A \mid y) = \frac{P(y \mid M_A)\, P(A)}{\sum_{c'} P(y \mid M_{c'})\, P(c')} = \frac{0.0638 \cdot 0.6}{0.0638 \cdot 0.6 + 0.0788 \cdot 0.4} = 0.5484
$$

$$
P(B \mid y) = 1 - P(A \mid y) = 0.4516
$$

$$
\hat{c}(y) = \operatorname*{argmax}_{c=A,B} P(c \mid y) = A
$$

Índice

1 Clasificación sintáctico-estadística ▷ 1

○ 2 *Algoritmo Forward* ▷ 5

3 Algoritmo de Viterbi ▷ 11

4 Apéndice: cálculos en clasificación estadística (resumen) ▷ 20

Algoritmo Forward

Definimos $\alpha(q,t)$ como la probabilidad de que un modelo oculto de Markov M genere el prefijo $y_1 \cdots y_t$, alcanzando el estado q en el instante t:

$$\alpha(q,t) = \sum_{\substack{q_1,\ldots,q_t \\ q_t = q}} P(y_1 \cdots y_t,\, q_1,\ldots,q_t)$$

$\alpha(q,t)$ puede calcularse recursivamente:

$$\begin{aligned}
\alpha(q,t) &= \sum_{\substack{q_1,\ldots,q_t \\ q_t = q}} P(y_1 \cdots y_t,\, q_1,\ldots,q_t) \\
&= \sum_{\substack{q_1,\ldots,q_{t-1} \\ q' \in Q \\ q_{t-1} = q'}} P(y_1 \cdots y_{t-1},\, q_1,\ldots,q_{t-1}) A_{q',q}\, B_{q,y_t} \\
&= \sum_{q' \in Q} \sum_{\substack{q_1,\ldots,q_{t-1} \\ q_{t-1} = q'}} P(y_1 \cdots y_{t-1},\, q_1,\ldots,q_{t-1}) A_{q',q}\, B_{q,y_t} \\
&= \sum_{q' \in Q} \alpha(q', t-1) A_{q',q}\, B_{q,y_t}
\end{aligned}$$

Algoritmo Forward (cont.)

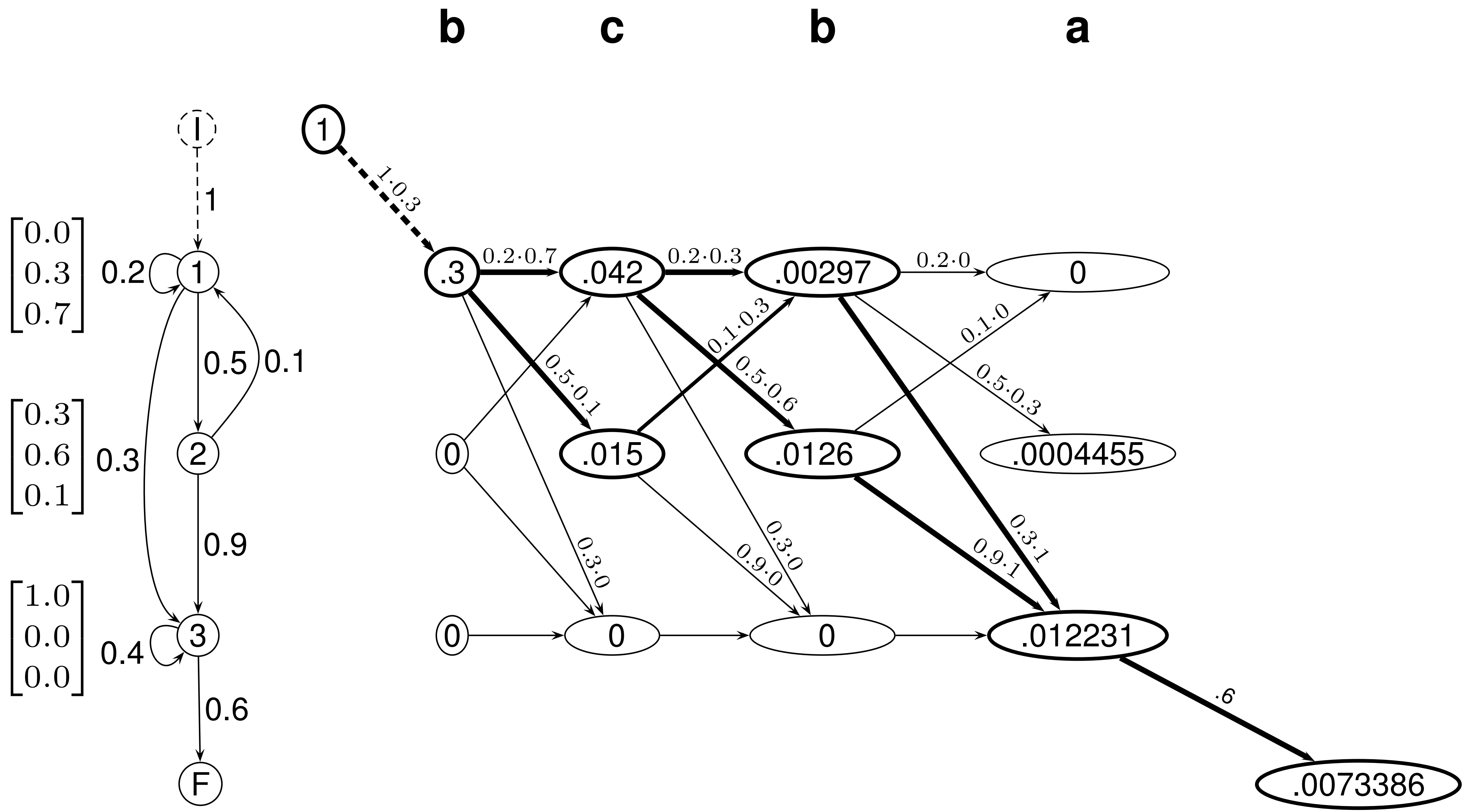
En general: $$\alpha(q,t) = \begin{cases} \pi_q \, B_{q,y_1} & \textbf{\textit{si}}\ t = 1 \\ \sum_{q' \in Q} \alpha(q', t-1) \, A_{q',q} \, B_{q,y_t} & \textbf{\textit{si}}\ t > 1 \end{cases}$$

La probabilidad de la cadena $P(y \mid M)$:

$$P(y \mid M) = \sum_{q \in Q} \alpha(q, |y|) \, A_{q,F}$$

- La función $\alpha()$ puede representarse como una matriz: $\alpha_{q,t} \equiv \alpha(q,t)$.
- Esta matriz define un *grafo multietapa* denominado *trellis* y permite el *cálculo iterativo eficiente* de $\alpha(q,|y|)$ *por Programación Dinámica*.
- Complejidad temporal del algoritmo: $O(mb)$, donde m es la longitud de la cadena y b es el número de transciones entre estados.

Algoritmo Forward: ejemplo



Algoritmo forward: ejercicio

Sea M un modelo con:

$Q = \{1, 2, 3, F\}$

$\Sigma = \{a, b, c\}$

$\pi_1 = \pi_2 = \frac{1}{2}$, $\pi_3 = 0$

A	1	2	3	F
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0
2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
3	0	0	$\frac{1}{2}$	$\frac{1}{2}$

B	a	b	c
1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

1. Aplica el algoritmo forward a la cadena `abc`.

Ejercicio: resolución directa

α	a $t=1$	b $t=2$	c $t=3$	
1	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{3}+$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{5}{144}$	$\frac{5}{144}\cdot\frac{1}{4}\cdot\frac{1}{6}+$ $\frac{1}{24}\cdot\frac{1}{3}\cdot\frac{1}{6}+$ $\frac{5}{96}\cdot 0\cdot\frac{1}{6}=\frac{13}{3456}$	
2	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{4}+$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{4}=\frac{1}{24}$	$\frac{5}{144}\cdot\frac{1}{2}\cdot\frac{1}{2}+$ $\frac{1}{24}\cdot\frac{1}{3}\cdot\frac{1}{2}+$ $\frac{5}{96}\cdot 0\cdot\frac{1}{2}=\frac{1}{76}$	
3		$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{2}+$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{2}=\frac{5}{96}$	$\frac{5}{144}\cdot\frac{1}{4}\cdot\frac{1}{4}+$ $\frac{1}{24}\cdot\frac{1}{3}\cdot\frac{1}{4}+$ $\frac{5}{96}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{7}{576}$	
F				$\frac{13}{3456}\cdot 0+$ $\frac{1}{76}\cdot 0+$ $\frac{7}{576}\cdot\frac{1}{2}=\frac{7}{1152}$

Índice

1 Clasificación sintáctico-estadística ▷ 1

2 Algoritmo Forward ▷ 5

∘ 3 *Algoritmo de Viterbi* ▷ 11

4 Apéndice: cálculos en clasificación estadística (resumen) ▷ 20

Aproximación de Viterbi a $P(y \mid M)$

Dado un modelo oculto de Markov $M = (Q, \Sigma, \pi, A, B)$ con estado final F, y una cadena $y = y_1 \cdots y_m \in \Sigma^+$, la probabilidad de que M genere y es:

$$P(y \mid M) = \sum_{z \in Q^+} P(y, z) = \sum_{q_1, \ldots, q_m \in Q^+} P(y,\, q_1, \ldots, q_m)$$

Una aproximación a $P(y \mid M)$ es la llamada *aproximación de Viterbi*:

$$\tilde{P}(y \mid M) = \max_{q_1, \ldots, q_m \in Q^+} P(y,\, q_1, \ldots, q_m)$$

La correspondiente *secuencia de estados más probable* es:

$$\tilde{q} = (\tilde{q}_1, \ldots, \tilde{q}_m) = \operatorname*{argmax}_{q_1, \ldots, q_m \in Q^+} P(y,\, q_1, \ldots, q_m)$$

Algoritmo de Viterbi

Definimos $V(q,t)$ como la probabilidad máxima de que un modelo oculto de Markov M alcance el estado q en el instante t, emitiendo el prefijo $y_1 \ldots y_t$:

$$V(q,t) = \max_{\substack{q_1,\ldots,q_t \\ q_t=q}} P(y_1 \cdots y_t,\, q_1,\ldots,q_t)$$

$V(q,t)$ puede calcularse recursivamente:

$$V(q,t) = \max_{\substack{q_1,\ldots,q_t \\ q_t=q}} P(y_1 \cdots y_t,\, q_1,\ldots,q_t)$$

$$= \max_{q' \in Q} \max_{\substack{q_1,\ldots,q_{t-1} \\ q_{t-1}=q'}} P(y_1 \cdots y_{t-1},\, q_1,\ldots,q_{t-1}) \cdot A_{q',q}\, B_{q,y_t}$$

$$= \max_{q' \in Q} V(q',t-1) \cdot A_{q',q}\, B_{q,y_t}$$

En general: $$V(q,t) = \begin{cases} \pi_q\, B_{q,y_1} & \textit{\textbf{si}}\ t=1 \\ \max_{q' \in Q} V(q',t-1)\, A_{q',q}\, B_{q,y_t} & \textit{\textbf{si}}\ t>1 \end{cases}$$

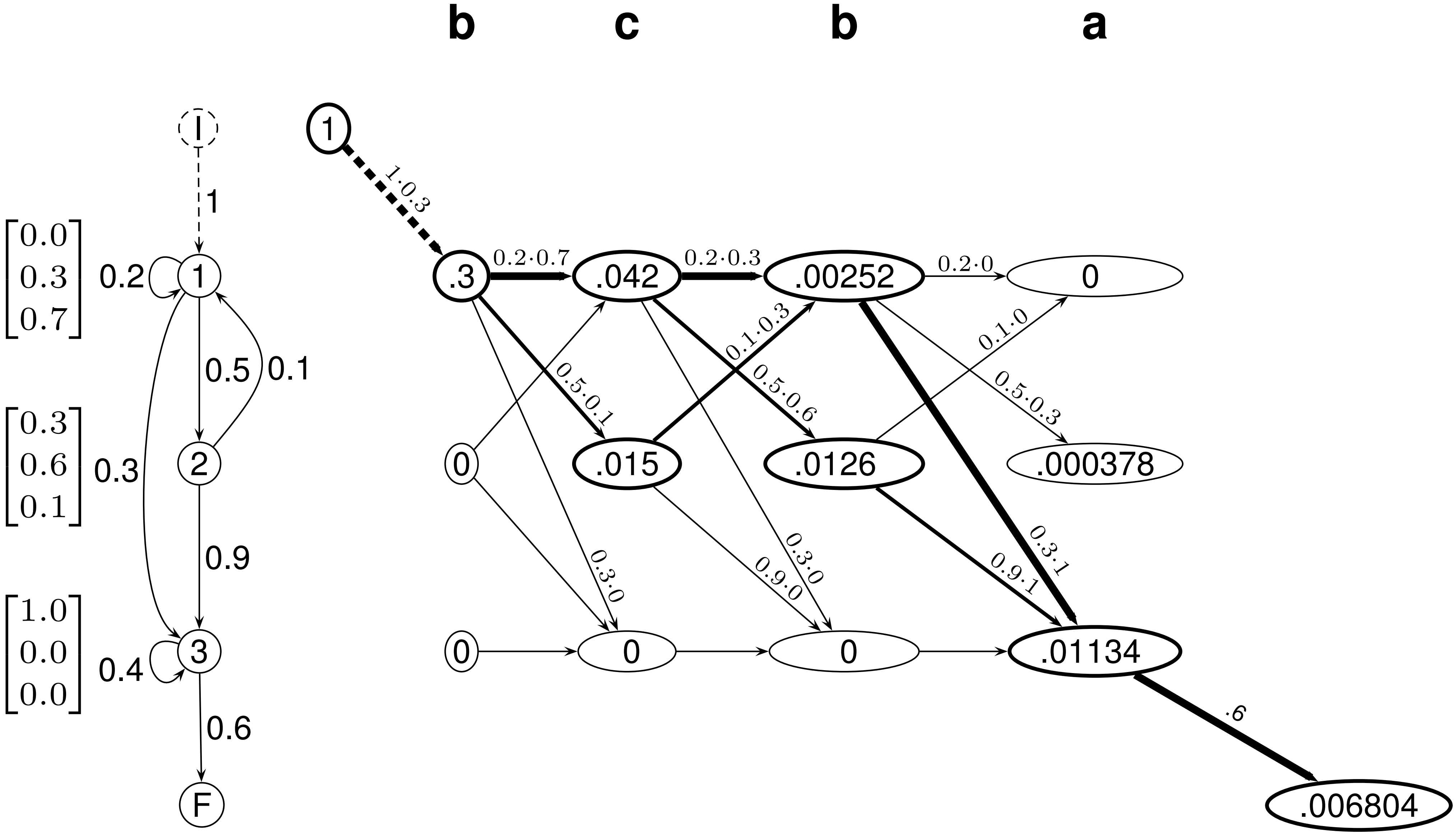
Algoritmo de Viterbi (cont.)

Aproximación de Viterbi a $P(y \mid M)$:

$$\tilde{P}(y \mid M) = \max_{q \in Q} V(q, |y|) \, A_{q,F}$$

- La función $V()$ puede representarse como una matriz: $V_{q,t} \equiv V(q,t)$.
- Esta matriz define un *grafo multietapa* denominado *trellis* y permite el *cálculo iterativo eficiente* de $V(q,|y|)$ *por Programación Dinámica*.
- La correspondiente secuencia óptima de estados, $\tilde{q}$, se calcula recorriendo el *trellis* hacia atrás.
- Complejidad temporal del algoritmo: $O(mb)$, donde m es la longitud de la cadena y b es el número de transiciones entre estados.

Algoritmo de Viterbi: ejemplo



Algoritmo de Viterbi: ejercicio

Sea M un modelo con:

$Q = \{1, 2, 3, F\}$

$\Sigma = \{a, b, c\}$

$\pi_1 = \pi_2 = \frac{1}{2}$, $\pi_3 = 0$

A	1	2	3	F
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0
2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
3	0	0	$\frac{1}{2}$	$\frac{1}{2}$

B	a	b	c
1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

1. Halla el *trellis* para la cadena `abc`.

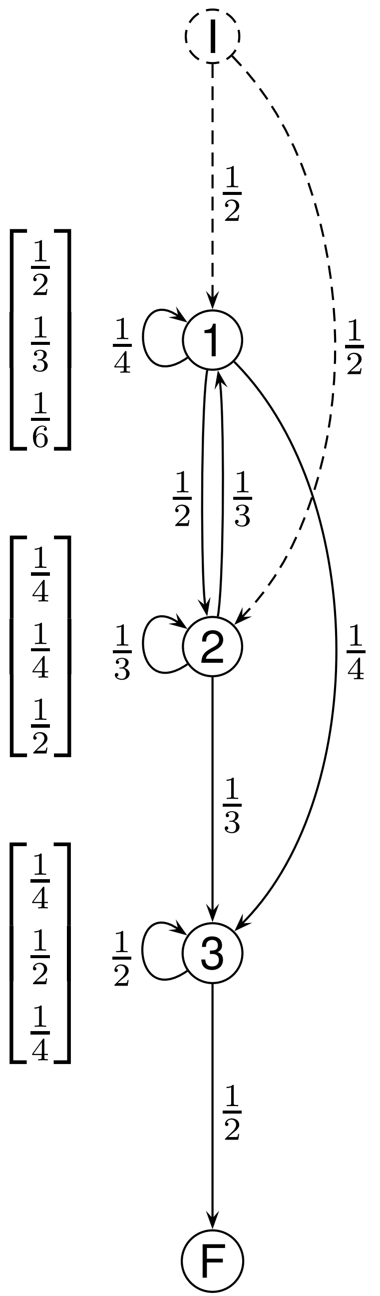
2. Obtén la secuencia óptima de estados asociada.

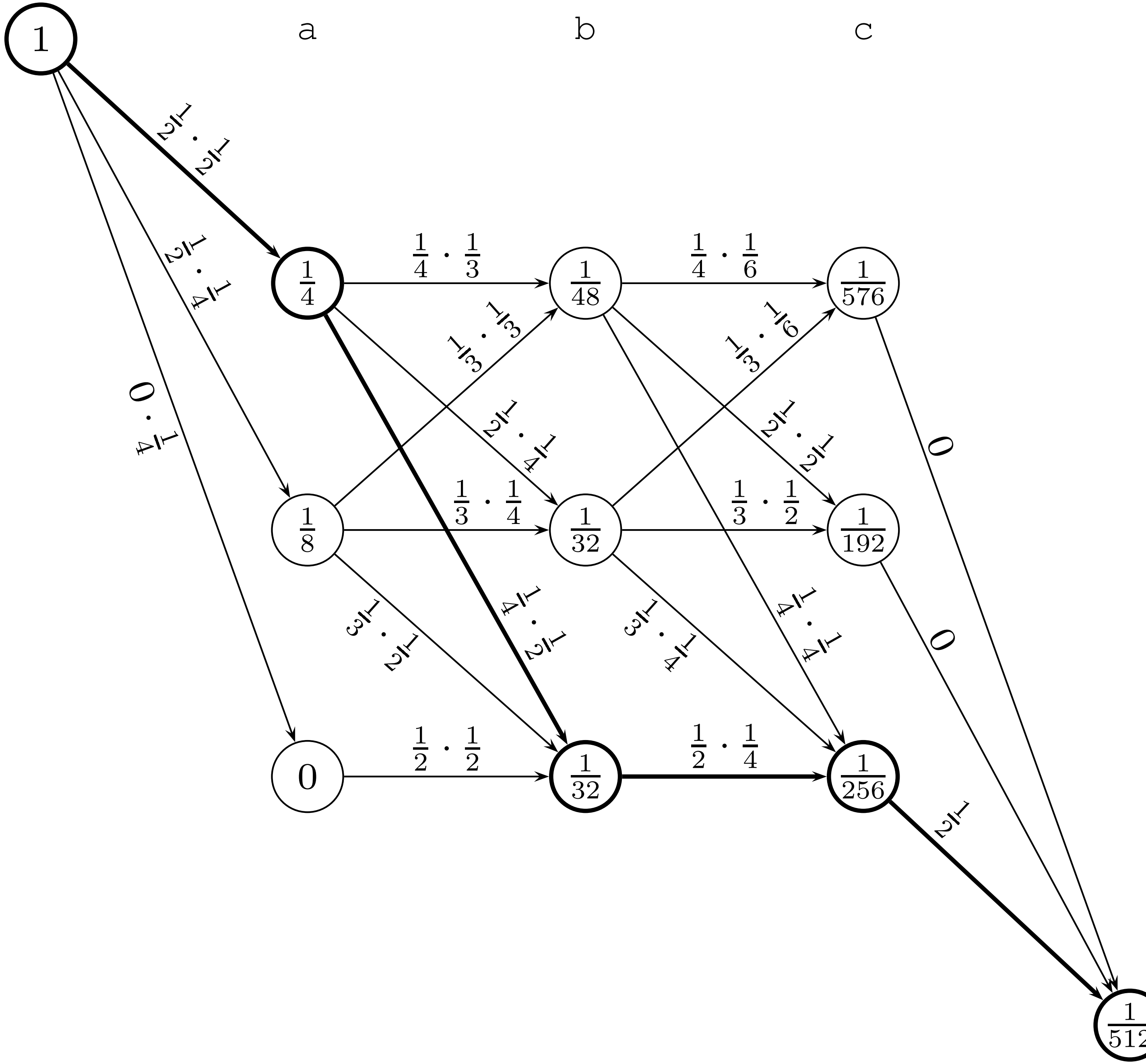
Ejercicio: resolución directa

V	a $t=1$	b $t=2$	c $t=3$	
1	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{3}=\frac{1}{48}$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{72}$	$\frac{1}{48}\cdot\frac{1}{4}\cdot\frac{1}{6}=\frac{1}{1152}$ $\frac{1}{32}\cdot\frac{1}{3}\cdot\frac{1}{6}=\frac{1}{576}$ $\frac{1}{32}\cdot 0\cdot\frac{1}{6}=0$	
2	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{32}$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{4}=\frac{1}{96}$	$\frac{1}{48}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{192}$ $\frac{1}{32}\cdot\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{192}$ $\frac{1}{32}\cdot 0\cdot\frac{1}{2}=0$	
3		$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{2}=\frac{1}{32}$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{48}$	$\frac{1}{48}\cdot\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{768}$ $\frac{1}{32}\cdot\frac{1}{3}\cdot\frac{1}{4}=\frac{1}{384}$ $\frac{1}{32}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{256}$	
F				$\frac{1}{576}\cdot 0=0$ $\frac{1}{192}\cdot 0=0$ $\frac{1}{256}\cdot\frac{1}{2}=\frac{1}{512}$

$$\tilde{Q}=(1,3,3,F)$$

Ejercicio: resolución gráfica





Clasificación sintáctico-estadística mediante Viterbi

En la práctica, las probabilidades condicionales de las clases suelen aproximarse mediante Viterbi. Consideremos el ejercicio de la página 4:

$$\begin{aligned}\tilde{P}(y \mid M_A) \\ &= \max(P(aab, q_1q_2q_3 = 001 \mid A), \\ &\qquad P(aab, q_1q_2q_3 = 011 \mid A), \\ &\qquad P(aab, q_1q_2q_3 = 111 \mid A)) \\ &= \max(0.0544, 0.0086, 0.0008) \\ &= 0.0544\end{aligned}$$

$$\begin{aligned}\tilde{P}(y \mid M_B) \\ &= \max(P(aab, q_1q_2q_3 = 001 \mid B), \\ &\qquad P(aab, q_1q_2q_3 = 011 \mid B)) \\ &= \max(0.0680, 0.0108) \\ &= 0.0680\end{aligned}$$

$$\tilde{P}(A \mid y) = \frac{\tilde{P}(y \mid M_A)\, P(A)}{\sum_{c'} \tilde{P}(y \mid c')\, P(c')} = \frac{0.0544 \cdot 0.6}{0.0544 \cdot 0.6 + 0.0680 \cdot 0.4} = 0.5455$$

$$\tilde{P}(B \mid y) = 1 - \tilde{P}(A \mid y) = 0.4545$$

$$\tilde{c}(y) = \operatorname*{argmax}_{c=A,B} \tilde{P}(c \mid y) = A$$ resultado idéntico al de la página 4

Índice

1 Clasificación sintáctico-estadística ▷ 1

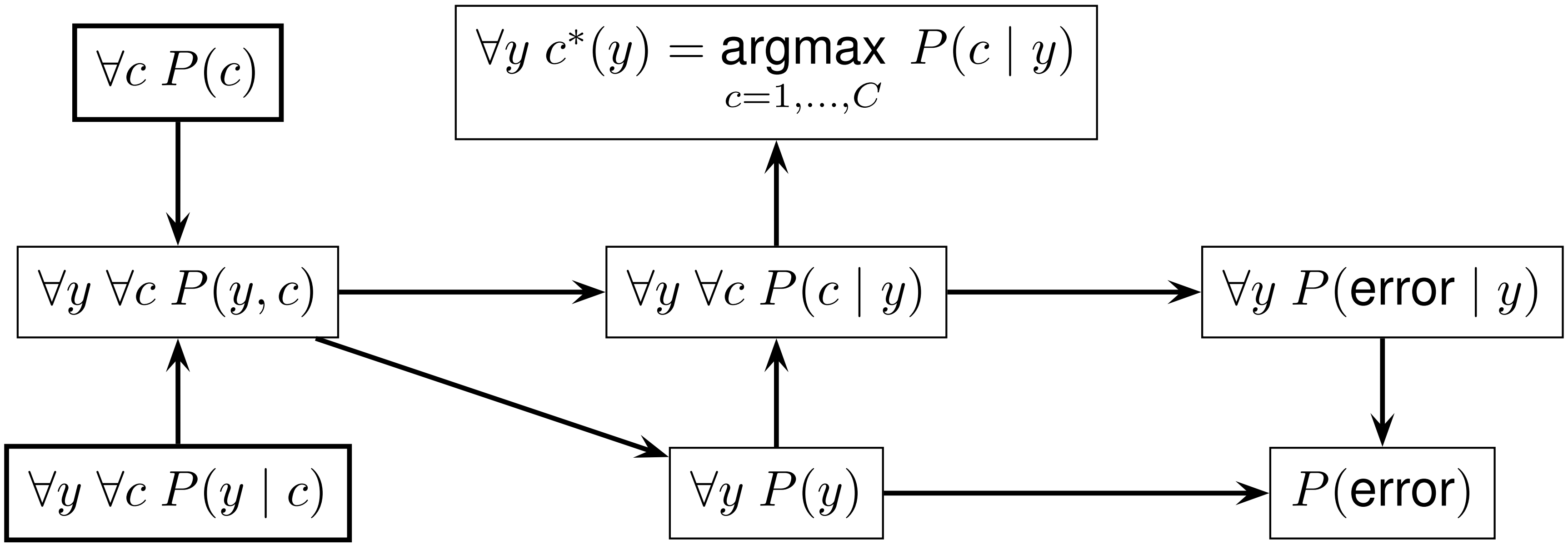
2 Algoritmo Forward ▷ 5

3 Algoritmo de Viterbi ▷ 11

○ 4 *Apéndice: cálculos en clasificación estadística (resumen)* ▷ 20

Apéndice: cálculos en clasificación estadística (resumen)

La aproximación estadística a la clasificación de objetos representados mediante vectores de características es igualmente válida para objetos representados mediante cadenas de símbolos de un alfabeto dado ($y \in \Sigma^+$):



Ejercicio:

- Pon nombres y ecuaciones de cálculo a los nodos del esquema.
- Calcula $P(c)$ a partir de $P(y, c)\ \forall y$.
- Calcula $P(y \mid c)$ a partir de $P(y, c)$ y $P(c)$.
- Calcula $P(y, c)$ a partir de $P(c \mid y)$ y $P(y)$.

Ejercicio: solución

$P(c)$ — ***Probabilidad a priori*** de la clase c

$P(y \mid c)$ — ***(Función de prob.) condicional*** de la clase c

$P(y,c) = P(c)\,P(y \mid c)$ — ***(Función de) probabilidad conjunta***

$P(y) = \sum_{c=1,\dots,C} P(y,c)$ — ***(Función de) probabilidad incondicional***

$P(c \mid y) = \dfrac{P(c)\,P(y \mid c)}{P(y)}$ — ***Probabilidad a posteriori*** de la clase c (para y)

$c^*(y) = \operatorname{argmax}_{c=1,\dots,C} P(c \mid y)$ — ***Clasificador (regla de decisión) de Bayes***

$P(\text{error} \mid y) = 1 - \max_{c=1,\dots,C} P(c \mid y)$ — ***Error de Bayes local (a posteriori)***

$P(\text{error}) = \sum_{y \in \Sigma^+} P(y)\,P(\text{error} \mid y)$ — ***Error de Bayes (global o a priori)***

$$P(c) = \sum_{y \in \Sigma^+} P(y,c) \qquad P(y \mid c) = \frac{P(y,c)}{P(c)} \qquad P(y,c) = P(y)\,P(y \mid c)$$