Sistemas Inteligentes

Escuela Técnica Superior de Informática

Universitat Politècnica de València

Tema B2T5:
Representación estructurada.
Modelos de Markov.
Algoritmo *Forward*.

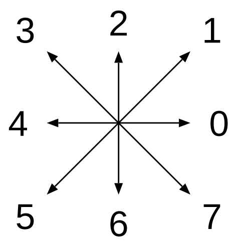
Índice

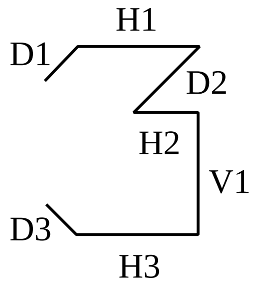
Objetos estructurados en reconocimiento de formas

- La representación de objetos en un espacio vectorial puede suponer una importante pérdida de información en algunos problemas:
  - Reconocimiento del habla
  - Reconocimiento de texto manuscrito
  - Identificación de la lengua
  - Reconocimiento de actitud o predilección en texto o habla
  - Identificación del tema de un documento
  - Reconocimiento de cromosomas
  - Reconocimiento de escenas en imágenes o vídeos
  - ...
- Solución: una representación estructurada:
  - Secuencias de longitud variable de vectores o de símbolos
  - Árboles, grafos, etc.
- Modelización: Modelos estructurales, por ejemplo, gramáticas estocásticas o modelos ocultos de Markov

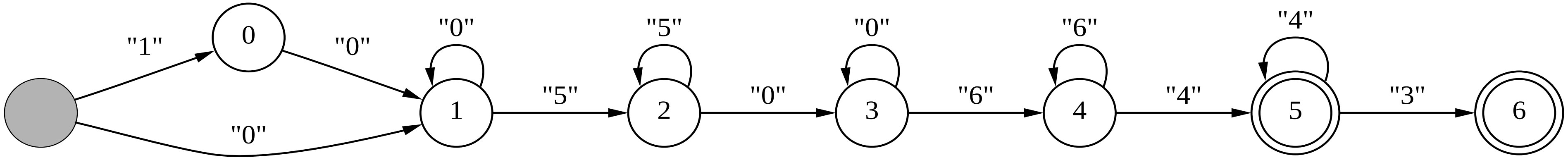
Secuencias de trazos y su modelado gramatical

MODELADO SINTÁCTICO DEL TRAZADO DE UN ”3”

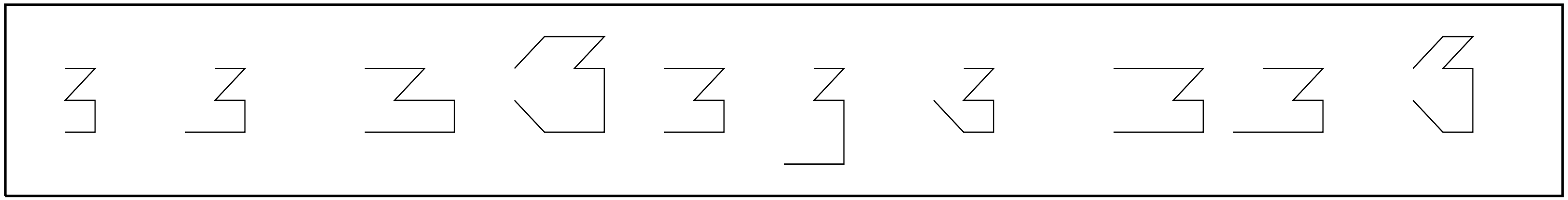




S	=>	D1		
D1	=>	H1	|	"1" H1
H1	=>	"0" H1	|	"0" D2
D2	=>	"5" D2	|	"5" H2
H2	=>	"0" H2	|	"0" V1
V1	=>	"6" V1	|	"6" H3
H3	=>	"4" H3	|	"4" D3
D3	=>	"3"	|	λ



tres.gr

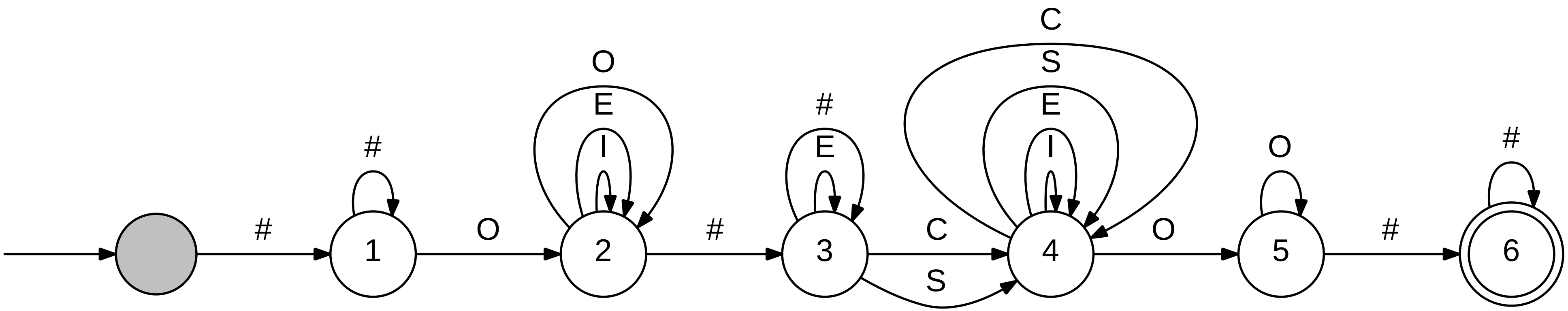


Trazos de treses generados por la gramatica "tres.gr"

Modelado sintáctico de pronunciaciones de palabras aisladas

Ejemplos de pronunciaciones de la palabra “ocho”,
representadas mediante cadenas de símbolos acústicos

```
######OOEIO##ECEEOOOO######
#####OOOEI###ECCEEOOO#####
#####OOOE###SCEOOOO######
######OOOIO###CCEOOOO######
######OOEIO##SCCEOOOO######
#####OOOEIE##ECCIEOOOO######
#####OOEO###CCIEOOOO######
######OOOE###ECCEIEOO#####
#####OOOEO##SCCIEEOO######
######OOOEO##SCSCCIEEOOO######
#####OOOEO##ECSCCEIOOOO#####
          . . .
```

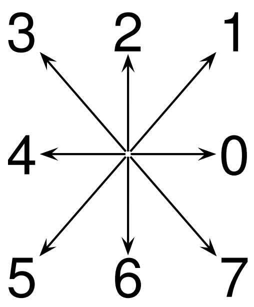


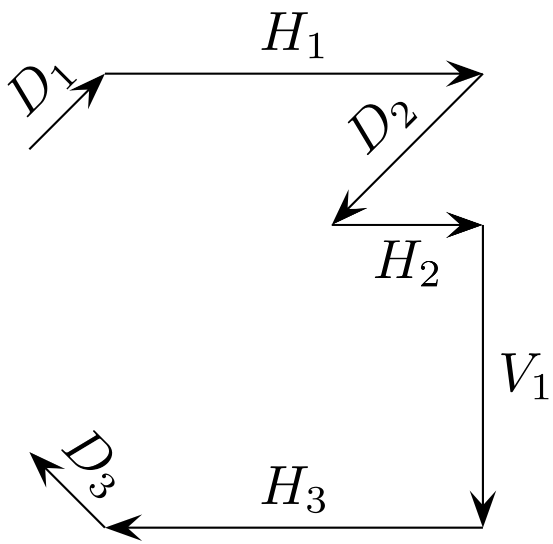
Una gramática que modela estas pronunciaciones

Índice

Un clasificador sintáctico para trazados de dígitos

Volvamos al modelado sintáctico del trazado de un “3”:





$$
\begin{array}{lcl}
S & \Rightarrow & D_1 \\
D_1 & \Rightarrow & 1\,H_1 \mid H_1 \\
H_1 & \Rightarrow & 0\,H_1 \mid 0\,D_2 \\
D_2 & \Rightarrow & 5\,D_2 \mid 5\,H_2 \\
H_2 & \Rightarrow & 0\,H_2 \mid 0\,V_1 \\
V_1 & \Rightarrow & 6\,V_1 \mid 6\,H_3 \\
H_3 & \Rightarrow & 4\,H_3 \mid 4\,D_3 \\
D_3 & \Rightarrow & 3 \mid \lambda
\end{array}
$$

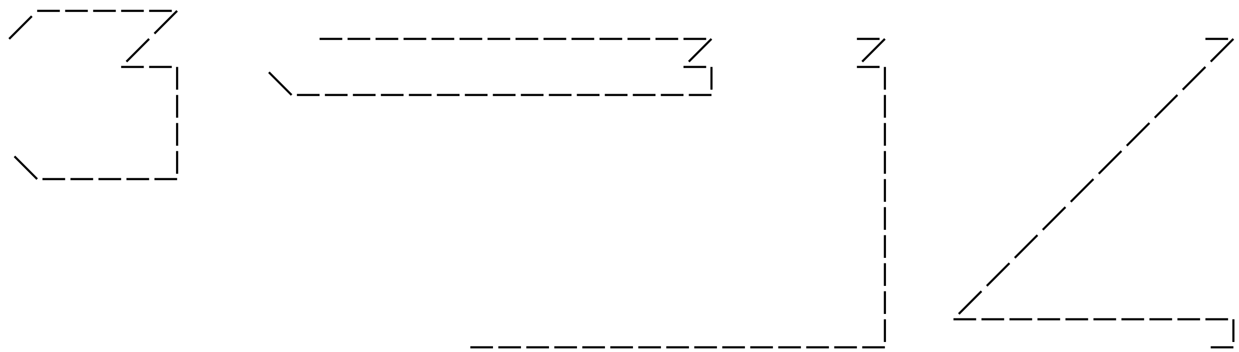
Supongamos que hay una gramática como ésta, G_c, para cada dígito c.

Si $x \in \{0, 1, \ldots, 7\}^+$ representa el trazado de un dígito, podemos decidir a qué dígito corresponde mediante un ***clasificador sintáctico “puro”***:

$$
c(x) = \begin{cases} c & \text{si } G_c \text{ es la única gramática que genera } x \\ \text{“rechazo”} & \text{si ninguna gramática genera } x \\ \text{“duda”} & \text{si hay más de una gramática que genera } x \end{cases}
$$

Incovenientes de las gramáticas convencionales y la clasificación sintáctica “pura”

- La necesidad de incluir las clases “rechazo” y “duda” es un claro incoveniente.
- Otro incoveniente claro es que las gramáticas convencionales generan tanto cadenas naturales como cadenas “indeseable”; p.e.:



La solución generalmente adoptada consiste en introducir probabilidades, es decir, emplear ***gramáticas estocásticas***.

Gramáticas estocásticas

- Una ***Gramática Estocástica*** G' es una gramática G con probabilidades asociadas a sus reglas:

$$G' = (G, p), \quad G = (N, \Sigma, R, S), \quad p : R \to [0, 1]$$

- Una gramática *incontextual* (o *regular*) es ***propia*** si:

$$\forall A \in N \quad \sum_{\forall \beta : A \to \beta \in R} p(A \to \beta) = 1$$

- ***Probabilidad de una cadena*** y generada por G':

$$\forall y \in \Sigma^* \quad p(y|G') = \sum_{d \in D_G(y)} p(d), \qquad p(d) = \prod_{(A \to \beta) \in d} p(A \to \beta)$$

- Una gramática Estocástica G' es ***consistente*** si:

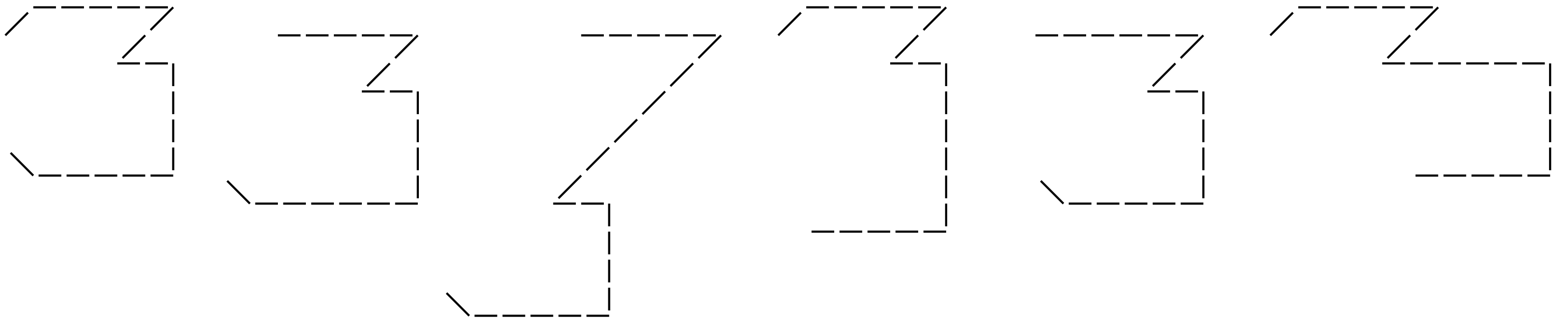
$$\sum_{y \in \Sigma^*} p(y|G') = 1$$

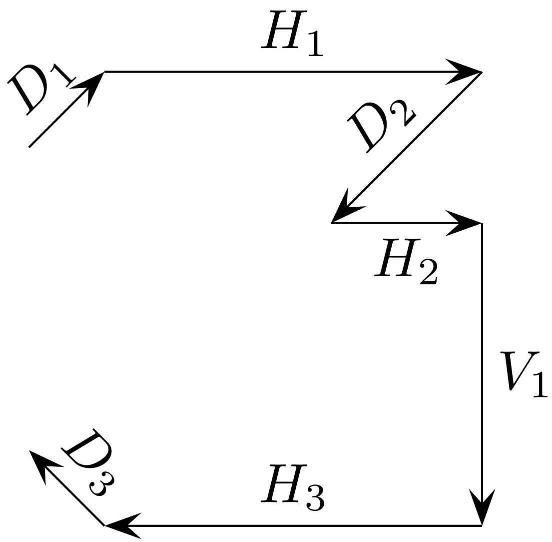
Aprendizaje de gramáticas estocásticas

- Determinación de las *reglas* o su estructura típica (*"topología"*):
  - Generalmente difícil de automatizar
  - Aproximación: topología pre-definida $\Rightarrow$ ***Modelos de Markov***

- Aprendizaje de las probabilidades: *estimación*
  - ***Gramáticas incontextuales (o Regulares) No-ambiguas*** G' **:**
    Estimación por máxima verosimilitud a partir de las frecuencias de uso de las reglas en el análisis de una secuencia de cadenas de entrenamiento supuestamente generadas por G'.
    Estas estimaciones se aproximan a las verdaderas probabilidades cuando el número de cadenas de entrenamiento $\rightarrow \infty$.
  - ***Gramáticas Regulares ambiguas y/o modelos de Markov:***
    Estimación localmente óptima mediante ***"reestimación por Viterbi"*** o, mejor, mediante el algoritmo "Backward-Forward"

Ejemplo de aprendizaje de las probabilidades





$$S \xrightarrow{6/6} D_1$$

$$D_1 \xrightarrow{3/6} 1\,H_1 \qquad D_1 \xrightarrow{3/6} H_1$$

$$H_1 \xrightarrow{25/31} 0\,H_1 \qquad H_1 \xrightarrow{6/31} 0\,D_2$$

$$D_2 \xrightarrow{10/16} 5\,D_2 \qquad D_2 \xrightarrow{6/16} 5\,H_2$$

$$H_2 \xrightarrow{10/16} 0\,H_2 \qquad H_2 \xrightarrow{6/16} 0\,V_1$$

$$V_1 \xrightarrow{20/26} 6\,V_1 \qquad V_1 \xrightarrow{6/26} 6\,H_3$$

$$H_3 \xrightarrow{25/31} 4\,H_3 \qquad H_3 \xrightarrow{6/31} 4\,D_3$$

$$D_3 \xrightarrow{4/6} 3 \qquad D_3 \xrightarrow{2/6} \lambda$$

Índice

Modelos de Markov

Un *modelo de Markov* es una quíntupla $M = (Q, \Sigma, \pi, A, B)$ donde:

- Q es un ***conjunto de estados***
  - En cada instante $t = 1, 2, \ldots,$ M está en uno de sus estados, denotado q_t
  - Q incluye un *estado final* F

- Σ es un ***conjunto de símbolos "observables"***
  En cada instante $t = 1, 2, \ldots,$ M emite un símbolo, que se denota con y_t

- $\pi \in \mathbb{R}^Q$ es un ***vector de probabilidades iniciales***:
  M elige q_1 según π: $\pi_q = P(q_1 = q)$

- $A \in \mathbb{R}^{Q \times Q}$ es una ***matriz de probabilidades de transición (entre estados)***:
  M elige q_{t+1} basándose en q_t y A: $A_{q,q'} = P(q_{t+1} = q' | q_t = q)$

- $B \in \mathbb{R}^{Q \times \Sigma}$ es una ***matriz de probabilidades de emisión (de símbolos)***:
  M elige y_t basándose en q_t y B: $B_{q,\sigma} = P(y_t = \sigma \mid q_t = q)$

Modelos de Markov (cont.)

Condiciones de normalización para π, A, B:

- Probabilidad de estado inicial:

$$0 \leq \pi_q \leq 1, \quad \sum_{q \in Q} \pi_q = 1, \quad \pi_F = 0$$

- Probabilidades de Transición entre estados:

$$0 \leq A_{q,q'} \leq 1, \quad \sum_{q' \in Q} A_{q,q'} = 1, \quad A_{F,q} = 0$$

- Probabilidades de emisión de observables:

$$0 \leq B_{q,\sigma} \leq 1, \quad \sum_{\sigma \in \Sigma} B_{q,\sigma} = 1, \quad B_{F,\sigma} = 0$$

Modelos de Markov: ejemplo

Q={1,2,3,F}; Σ = {a,b,c}; $\pi_1 = 1; \quad \pi_2 = \pi_3 = \pi_F = 0$

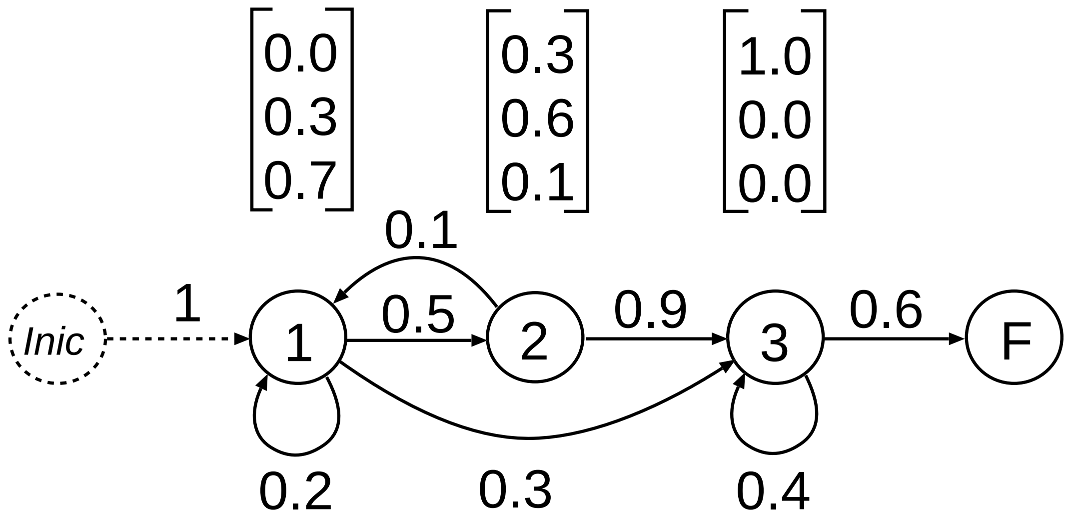
$p(q'|q) \equiv A(q,q')$ →

	1	2	3	F
1	0.2	0.5	0.3	0.0
2	0.1	0.0	0.9	0.0
3	0.0	0.0	0.4	0.6

$p(\sigma|q) \equiv B(q,\sigma)$ →

	a	b	c
1	0.0	0.3	0.7
2	0.3	0.6	0.1
3	1.0	0.0	0.0

Representación Gráfica Equivalente:



Proceso Markoviano generador de cadenas

Sea $M = (Q, \Sigma, \pi, A, B)$ un modelo de Markov con estado final F

1. *Elegir un estado inicial* $q \in Q$ según $P(q) \equiv \pi_q$
2. *Seleccionar una observación* $\sigma \in \Sigma$ según $P(\sigma|q) \equiv B_{q,\sigma}$; emitir σ
3. *Elegir un estado siguiente* $q' \in Q$ según $P(q'|q) \equiv A_{q,q'}$
4. Si $q = F$ *terminar*; sino, ***ir a paso 2***

Sean:

- $y = y_1, y_2, \ldots, y_m \in \Sigma^+$: secuencia de observaciones producida por M
- $z = q_1, q_2, q_m, \ldots, F \in Q^+$: secuencia de estados que genera a y

La probabilidad de que M produzca y mediante z es:

$$P(y, z) = P(z) \cdot P(y \mid z) = P(q_1) \prod_{t=2}^{m} P(q_t \mid q_{t-1}) \, P(F \mid q_m) \cdot \prod_{t=1}^{m} P(y_t \mid q_t)$$

Probabilidad de generar una cadena con un modelo de Markov

Probabilidad de que M genere la cadena $y = y_1 \dots y_m \in \Sigma^+$:

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

$$= \sum_{q_1,\dots,q_m \in Q^+} P(q_1) \prod_{t=2}^{m} P(q_t \mid q_{t-1}) \;\; P(F \mid q_m) \cdot \prod_{t=1}^{m} P(y_t \mid q_t)$$

$$= \sum_{q_1,\dots,q_m \in Q^+} \pi_{q_1} \, B_{q_1,y_1} \left(\prod_{t=2}^{m} A_{q_{t-1},q_t} \, B_{q_t,y_t} \right) A_{q_m,F}$$

Se cumple: $0 \leq P(y|M) \leq 1, \qquad \sum_{y \in \Sigma^+} P(y|M) = 1$

Probabilidades calculadas con el modelo del ejemplo

$$
\begin{aligned}
P(\mathtt{cba}|M) &= (\pi_1 \cdot B_{1,\mathtt{c}}) \; (A_{1,2} \cdot B_{2,\mathtt{b}}) \; (A_{2,3} \cdot B_{3,\mathtt{a}}) \; A_{3,F} \\
&+ (\pi_1 \cdot B_{1,\mathtt{c}}) \; (A_{1,1} \cdot B_{1,\mathtt{b}}) \; (A_{1,3} \cdot B_{3,\mathtt{a}}) \; A_{3,F} \\
&= (1 \cdot 0.7) \; (0.5 \cdot 0.6) \; (0.9 \cdot 1) \; 0.6 \\
&+ (1 \cdot 0.7) \; (0.2 \cdot 0.3) \; (0.3 \cdot 1) \; 0.6 \; = 0.1134 + 0.00756 \\
&\approx 0.12
\end{aligned}
$$

$$P(\mathtt{bcbaa}|M) = P(y, z_1) + P(y, z_2) + P(y, z_3) + P(y, z_4) + P(y, z_5)$$

$$
\begin{array}{rcccccccl}
y = & \mathtt{b} & \mathtt{c} & \mathtt{b} & \mathtt{a} & \mathtt{a} & & & \\
z_1 = & 1 & 1 & 1 & 2 & 3 & F & & \\
P(y, z_1) = & (1 \cdot 0.3) & (0.2 \cdot 0.7) & (0.2 \cdot 0.3) & (0.5 \cdot 0.3) & (0.9 \cdot 1) & 0.6 & & = 0.000204 \\
z_2 = & 1 & 1 & 1 & 3 & 3 & F & & \\
P(y, z_2) = & (1 \cdot 0.3) & (0.2 \cdot 0.7) & (0.2 \cdot 0.3) & (0.3 \cdot 1) & (0.4 \cdot 1) & 0.6 & & = 0.000181 \\
z_3 = & 1 & 1 & 2 & 3 & 3 & F & & \\
P(y, z_3) = & (1 \cdot 0.3) & (0.2 \cdot 0.7) & (0.5 \cdot 0.6) & (0.9 \cdot 1) & (0.4 \cdot 1) & 0.6 & & = 0.002722 \\
z_4 = & 1 & 2 & 1 & 2 & 3 & F & & \\
P(y, z_4) = & (1 \cdot 0.3) & (0.5 \cdot 0.1) & (0.1 \cdot 0.3) & (0.5 \cdot 0.3) & (0.9 \cdot 1) & 0.6 & & = 0.000036 \\
z_5 = & 1 & 2 & 1 & 3 & 3 & F & & \\
P(y, z_5) = & (1 \cdot 0.3) & (0.5 \cdot 0.1) & (0.1 \cdot 0.3) & (0.3 \cdot 1) & (0.4 \cdot 1) & 0.6 & & = 0.000032 \\
 & & & & & & & P(y|M) & = 0.003175
\end{array}
$$

Equivalencia entre modelos de Markov y gramáticas regulares estocásticas

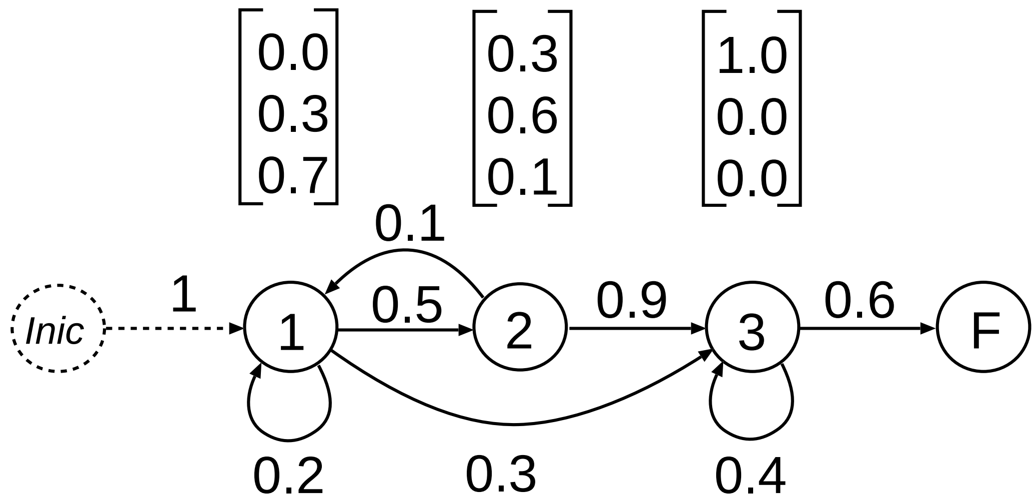
Dado un modelo de Markov M, existe una gramática regular estocástica G tal que $P(y|M) = P(y|G)\ \forall y \in \Sigma^*$

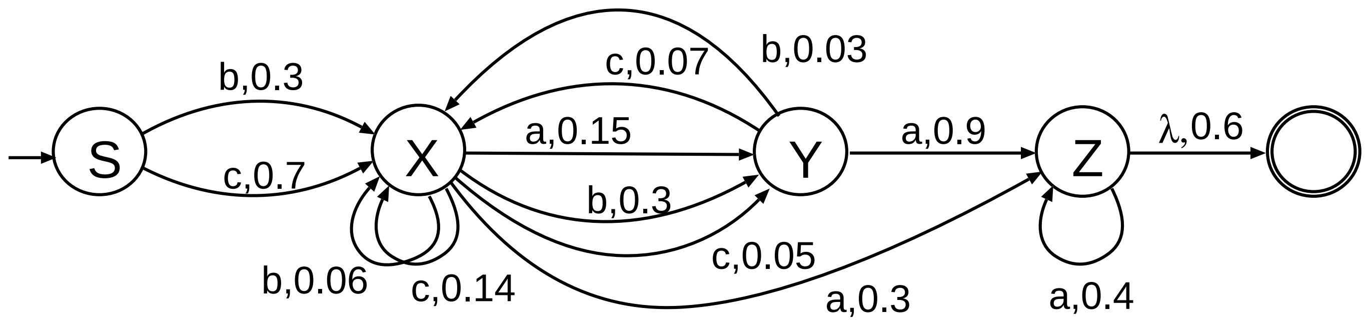
Se demuestra fácilmente por construcción.

Dada una gramática regular estocástica G, existe un modelo de Markov M tal que $P(y|M) = P(y|G)\ \forall y \in \Sigma^*$ (salvo casos degenerados)

Se demuestra por construcción (algo más compleja que la anterior).

Equivalencia entre modelos de Markov y gramáticas estocásticas





0.3: S -> bX	0.06: X -> bX	0.90: Y -> aZ	0.4: Z -> aZ
0.7: S -> cX	0.14: X -> cX	0.03: Y -> bX	0.6: Z -> λ
	0.15: X -> aY	0.07: Y -> cX	
	0.30: X -> bY		
	0.05: X -> cY		
	0.30: X -> aZ		

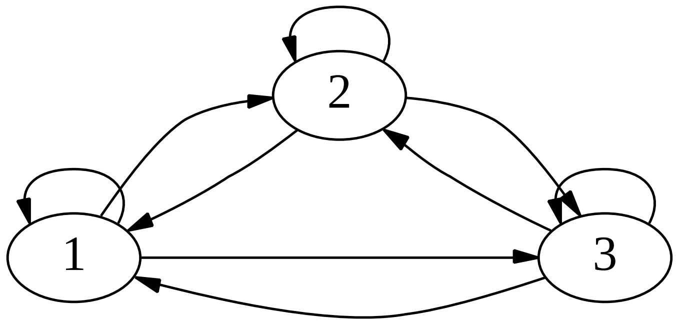
Estructura o “topología” de un modelo de Markov

La Topología de un Modelo de Markov: es la forma del grafo subyacente. Viene determinada por la estructura (número y ubicación de ceros) de la matriz de transición entre estados A. Topologías más comunes:

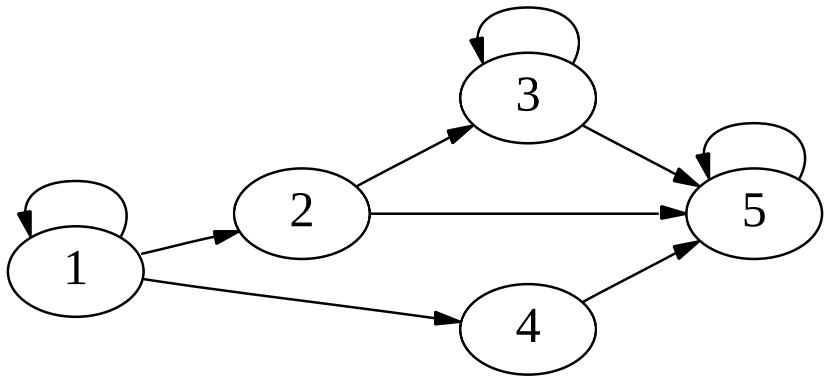
- ***Ergódica:*** grafo completo, no hay ceros en A.
- ***Izquierda-derecha:*** el grafo es *dirigido y acíclico* (DAG), aunque pueden haber bucles individuales en los estados. A es triangular.
- ***Lineal:*** el grafo es un DAG (con posibles bucles en estados) restringido en el que las transiciones que salen del estado $i-$ésimo sólo pueden alcanzar a los estados $i+1, \ldots, i+k$. Los elemetos no nulos de A están en $k+1$ diagonales adyacentes. Estas transiciones se denominan *“saltos”* o *“skips”*.
- ***Estrictamente Lineal:*** el grafo es una concatenación de estados, (con posibles bucles en estados). Los elemetos no nulos de A están en dos diagonales adyacentes.

Ejemplos de topologías de modelos de Markov

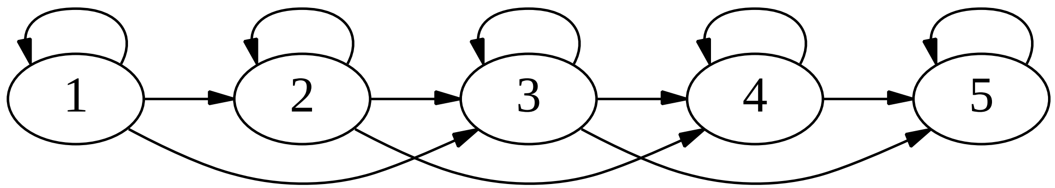
Ergódica



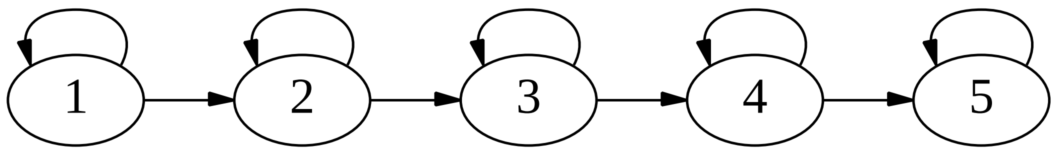
Izquierda-Derecha



Lineal



Estrictamente Lineal



Ejercicio

Sea M un modelo de estados $Q = \{1, 2, 3, 4, 5, F\}$; alfabeto $\Sigma = \{a, b\}$; probabilidades iniciales $\pi_1 = 1$, $\pi_2 = \pi_3 = \pi_4 = \pi_5 = 0$; y probabilidades de transición y de emisión:

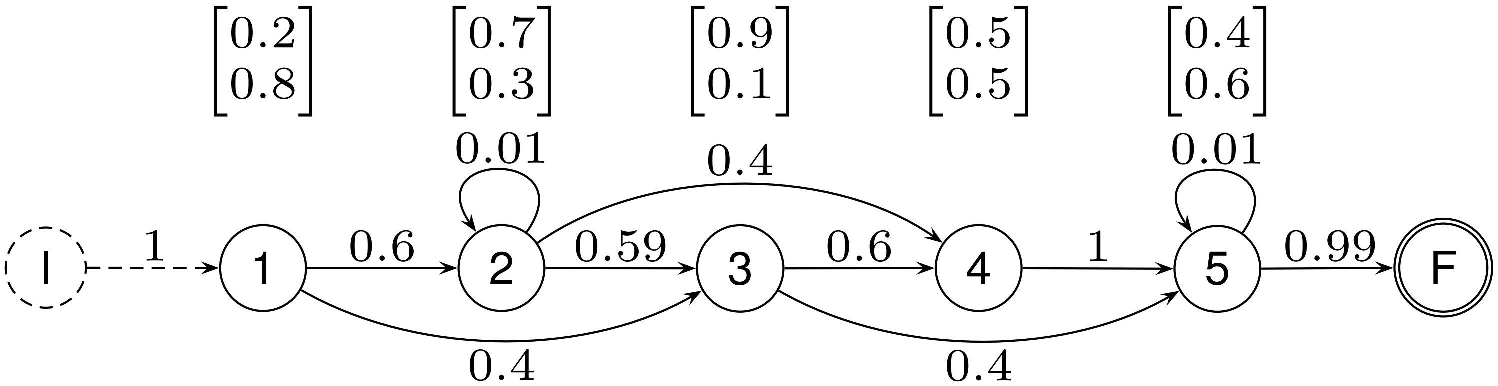
A	1	2	3	4	5	F
1		0.6	0.4			
2		0.01	0.59	0.4		
3				0.6	0.4	
4					1.0	
5					0.01	0.99

B	a	b
1	0.2	0.8
2	0.7	0.3
3	0.9	0.1
4	0.5	0.5
5	0.4	0.6

1. Representa gráficamente este modelo.
2. Calcula la probabilidad de que M genere una cadena de 3 símbolos.
3. ¿Puede decirse que las cadenas más probables tienden a ser las cadenes de longitudes comprendidas entre 3 y 5 símbolos?

Ejercicio (solución)

1)



2) Hay 8 cadenas de longitud $L = 3$: $\{a, b\}^3 = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$. Cada una de ellas se genera mediante una única secuencia de estados: $1, 3, 5, F$.

$$\begin{aligned}
P(L = 3 \mid M) &= \sum_{y\in\{a,b\}^3} P(y \mid M) = \sum_{y\in\{a,b\}^3} \pi_1\, B_{1,y_1} \left(A_{1,3}\, B_{3,y_3}\ A_{3,5}\, B_{5,y_5}\right) A_{5,F} \\
&= \pi_1\, A_{1,3}\, A_{3,5}\, A_{5,F} \sum_{y\in\{a,b\}^3} (B_{1,y_1}\, B_{3,y_3}\, B_{5,y_5}) \\
&= 0.1584\, (B_{1,a}\, B_{3,a}\, B_{5,a} + B_{1,a}\, B_{3,a}\, B_{5,b} + \cdots + B_{1,b}\, B_{3,b}\, B_{5,b}) \\
&= 0.1584\, (B_{1,a} + B_{1,b})\, (B_{3,a} + B_{3,b})\, (B_{5,a} + B_{5,b}) \\
&= 0.1584 \cdot 1 \cdot 1 \cdot 1 \ = \ \mathbf{0.1584}
\end{aligned}$$

3) ***Sí***. Los productos que incluyen $A_{2,2}$ o $A_{5,5}$ (ambas $= 0.01$) influyen de forma despreciable en las sumas. a) $P(L = l \mid M) = \mathbf{0}\ \forall l < 3$; b) $P(L = 4 \mid M) = 1\cdot 0.6\cdot 0.4\cdot 1\cdot 0.99 + 1\cdot 0.4\cdot 0.6\cdot 1\cdot 0.99 + 1\cdot 0.6\cdot 0.59\cdot 0.4\cdot 0.99 + \ldots \approx \mathbf{0.62}$; c) $P(L = 5 \mid M) = 1\cdot 0.6\cdot 0.59\cdot 0.6\cdot 1\cdot 0.99 + \ldots \approx \mathbf{0.21}$; d) $\forall l \geq 6,\ P(L = l \mid M) < \mathbf{0.005}$ y disminuye exponencialmente con l de forma muy acusada.

Índice

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 $\hat{c}(y)$:

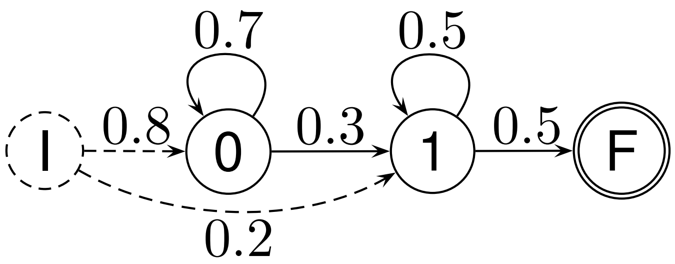
$$\hat{c}(y) = \operatorname*{argmax}_{1 \leq c \leq C} P(c \mid y)$$

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:

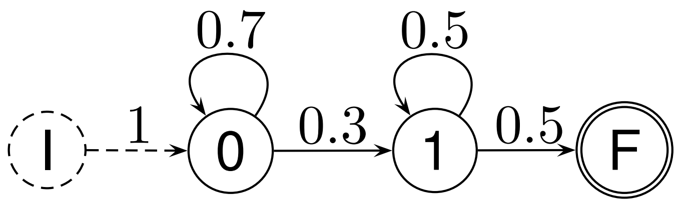
Modelo M_A: $P(y \mid A) = P(y \mid M_A)$

$$\begin{bmatrix} a \\ b \end{bmatrix} \quad \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix} \quad \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$



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

$$\begin{bmatrix} a \\ b \end{bmatrix} \quad \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix} \quad \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$



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

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:

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

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

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

Algoritmo Forward (cont.)

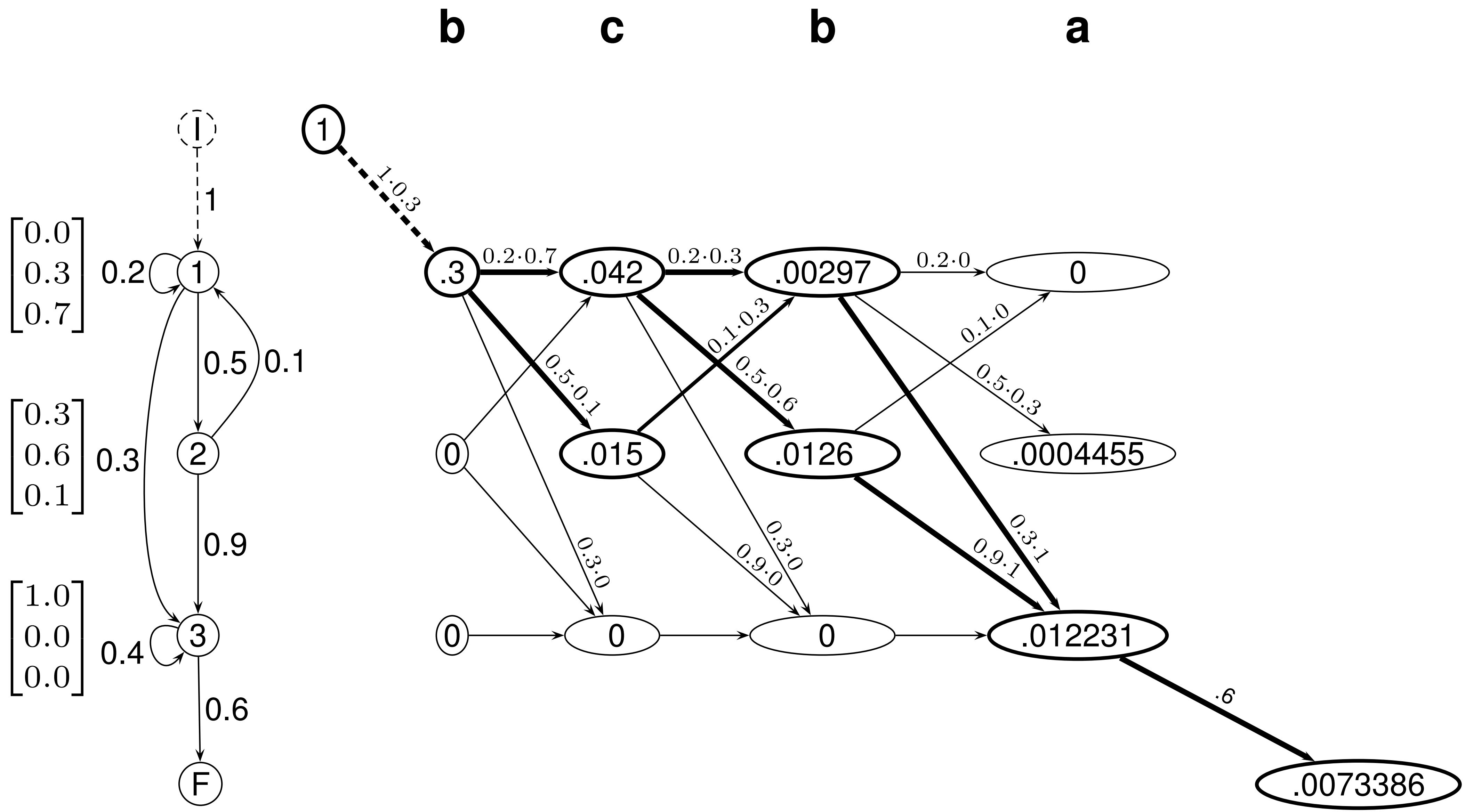
En general: $$\alpha(q,t) = \begin{cases} \pi_q \, B_{q,y_1} & \textbf{\textit{si}}\ t = 1 \\ \sum_{q' \in Q} \alpha(q', t-1)\, A_{q',q}\, B_{q,y_t} & \textbf{\textit{si}}\ t > 1 \end{cases}$$

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

Gramáticas

- ***Monoide Libre*** Σ^****:*** Dado un conjunto finito Σ, Σ^+ es el conjunto de todas las cadenas de longitud finita formadas por elementos de Σ. Además, $\Sigma^* = \Sigma^+ \cup \{\lambda\}$ (la *cadena vacia*).

- ***Gramática:*** $G = (N, \Sigma, R, S)$
  - N: Conjunto finito de *No-Terminales*
  - Σ: Conjunto finito de *Terminales o Primitivas*
  - $S \in N$: Símbolo No-Terminal inicial o *.Axioma"*
  - $R \subset (N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$: conjunto de *Reglas*.

    A regla se escribe como:

    $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}, \quad \alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*, \;\; \beta \in (N \cup \Sigma)^*$

    Si varias reglas comparten su parte izquierda, pueden abreviarse como:

    $\boldsymbol{\alpha} \rightarrow \; \boldsymbol{\beta_1} \mid \boldsymbol{\beta_2} \mid \cdots$

Gramáticas y lenguajes

- ***Derivación Elemental :*** $\underset{G}{\Longrightarrow}$**:**

  $\mu\,\alpha\,\delta \underset{G}{\Longrightarrow} \mu\,\beta\,\delta$ sii $\exists(\alpha \to \beta) \in R, \ \ \mu, \delta \in (N \cup \Sigma)^*$

- ***Derivación*** $\overset{*}{\underset{G}{\Longrightarrow}}$**:**

  Es una *secuencia finita de derivaciones elementales*. Una derivación d puede escribirse como la secuencia correspondiente de reglas de G.

  El *conjunto de derivaciones* de $y \in \Sigma^*$ (tales que $S \overset{*}{\underset{G}{\Longrightarrow}} y$) se denota como $D_G(y)$.

  Una gramática G es *ambigua* si $\exists y \in \Sigma^*$ tal que $|D_G(y)| > 1$

- ***Lenguaje generado por una gramática*** G, $\mathcal{L}(G)$**:**

$$\mathcal{L}(G) \ = \ \{\, y \in \Sigma^* \ \mid \ S \overset{*}{\underset{G}{\Longrightarrow}} y \,\}$$

Tipos de gramáticas y lenguajes

Jerarquía de Chomsky para lenguajes recursivos

0: No-restringidos

1: Contextuales

$$\alpha \rightarrow \beta \, , \qquad |\alpha| \; \leq \; |\beta|$$

2: Incontextuales

$$B \rightarrow \beta \, , \qquad B \in N$$

3: ***Regulares o de "Estados-Finitos"***

$$A \rightarrow aB \;\; \text{o} \;\; A \rightarrow a \, , \quad A, B \in N, \;\; a \in \Sigma \cup \{\lambda\}$$

Gramáticas regulares y autómatas finitos

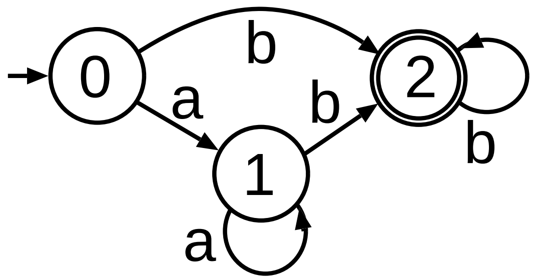
- ***Gramáticas Regulares:*** $G = (N, \Sigma, R, S)$,
  Reglas de R de la forma: $A \to aB \;\vee\; A \to a, \quad A, B \in N, \quad a \in \Sigma$

- ***Autómatas Finitos:*** $\mathcal{A} = (Q, \Sigma, \delta, q_0, F), \quad q_0 \in Q, \quad F \subseteq Q, \quad \delta : Q \times \Sigma \to 2^Q$

- ***Equivalencia:*** Para cada Gramática Regular existe un Autómatas Finito que reconoce el mismo lenguaje. (*¡Ojo: la inversa no es siempre cierta en el caso de lenguajes* **estocásticos***!*).

Ejemplo:

$$G = (N, \Sigma, R, S);$$
$$\Sigma = \{a, b\};\; N = \{S, A_1, A_2\};$$
$$R = \{\; S \to aA_1 \mid bA_2 \mid b,$$
$$A_1 \to aA_1 \mid bA_2 \mid b,$$
$$A_2 \to bA_2 \mid b\}$$



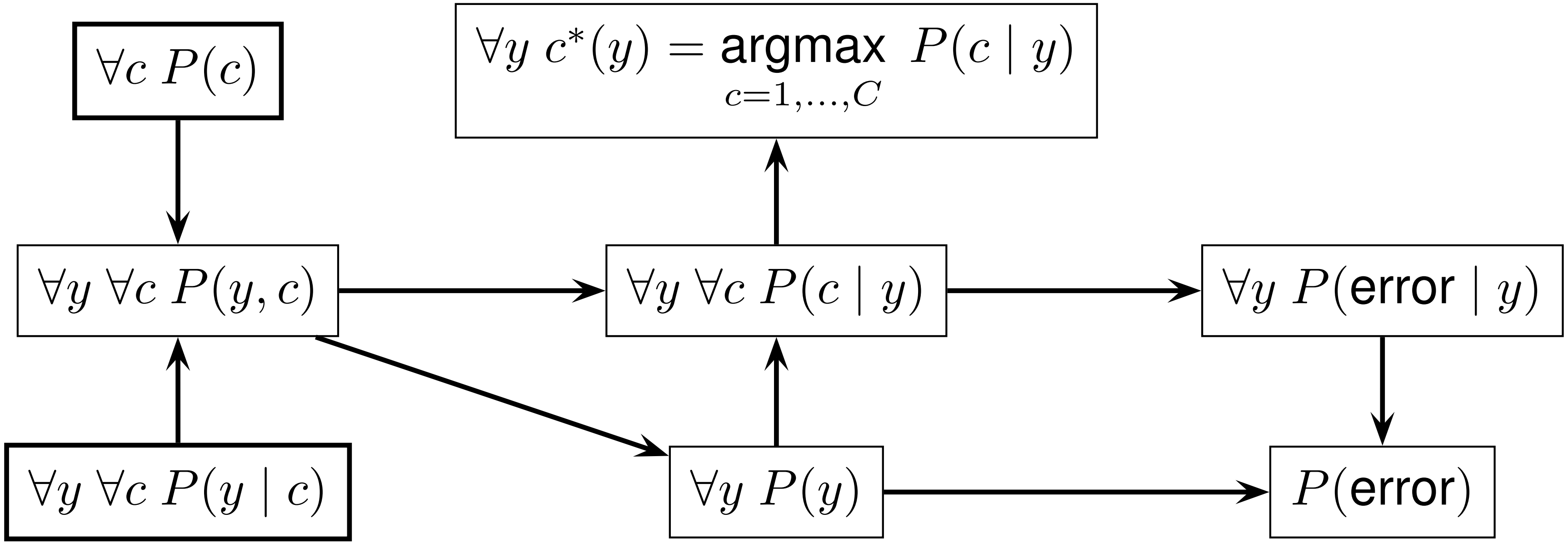
$$\mathcal{A} = \{Q, \Sigma, \delta, q_0, F\};$$
$$Q = \{0, 1, 2\},$$
$$\Sigma = \{a, b\},$$
$$q_0 = 0,\; F = \{2\}$$

$$\mathcal{L}(G) = \{b, ab, bb, aab, abb, bbb, \ldots, aaabbbb, \ldots\} = \mathcal{L}(\mathcal{A})$$

Índice

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***

$P(c \mid y) = \frac{P(c)\, P(y \mid c)}{P(y)}$ ***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)***

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