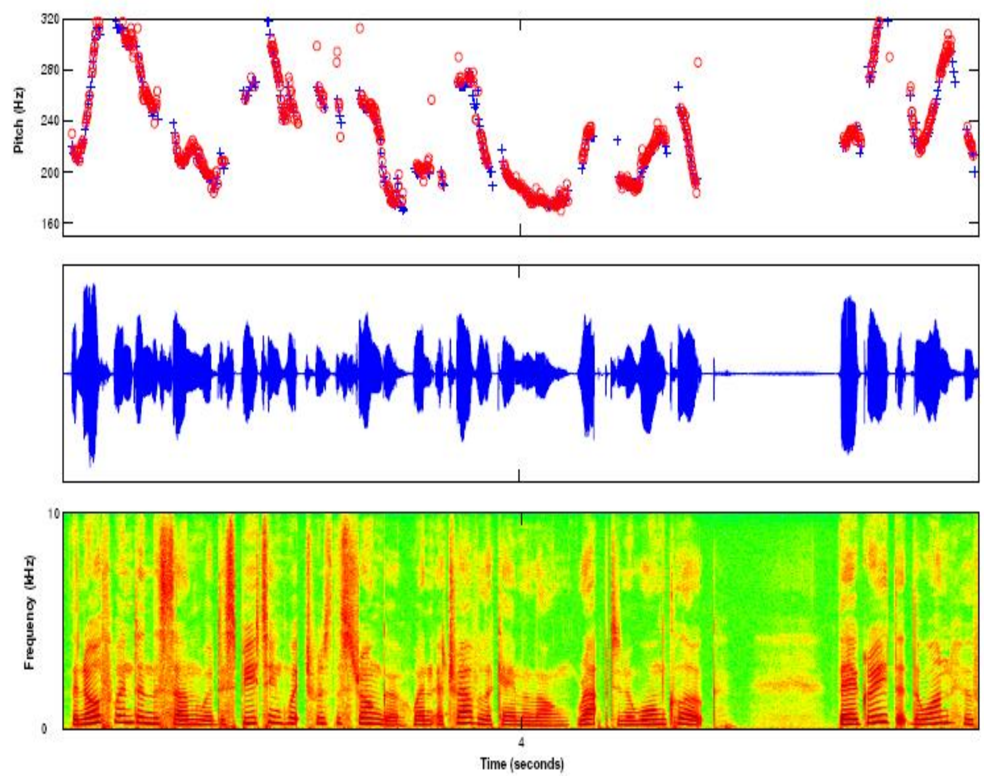
Inferència probabilistica en el temps
Hidden markov model

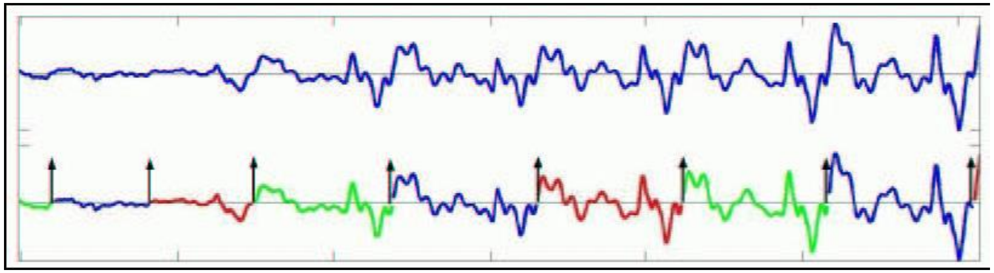
Coneixement, Raonament i Incertesa.

El contingut d’aquest document s’ha derivat de material provinent de Alex Simma, Uni. Berkeley http://www.cs.berkeley.edu/~jordan/courses/294-fall09/lectures/hmm/

Reconeixement de la parla

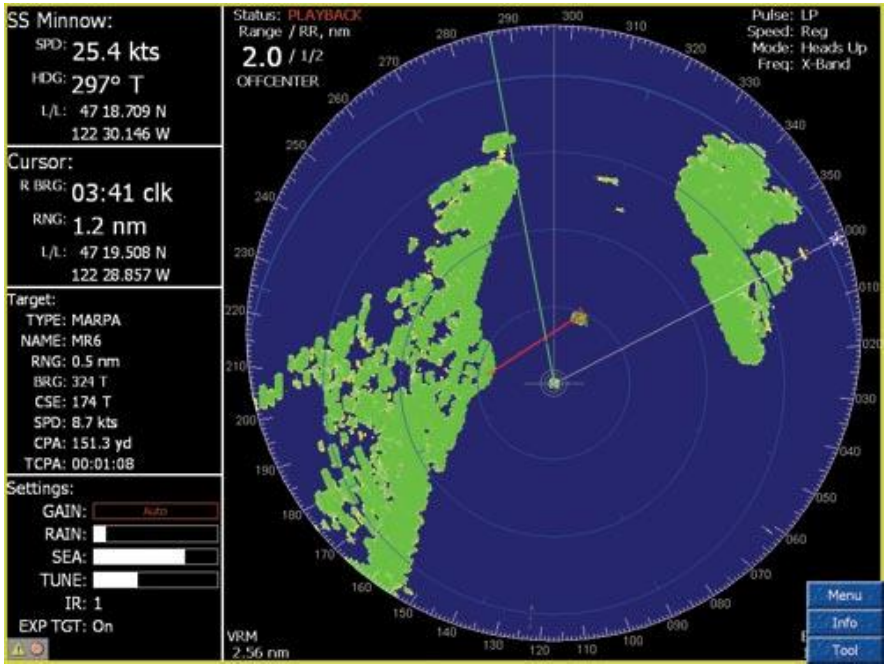
- Donada una ona d'àudio, volem extreure i reconèixer paraules de forma robusta
- Els models estadístics es poden usar per a
  - Donar més robustesa al soroll
  - Adaptar-se a accents diferents
  - Aprendre del entrenament



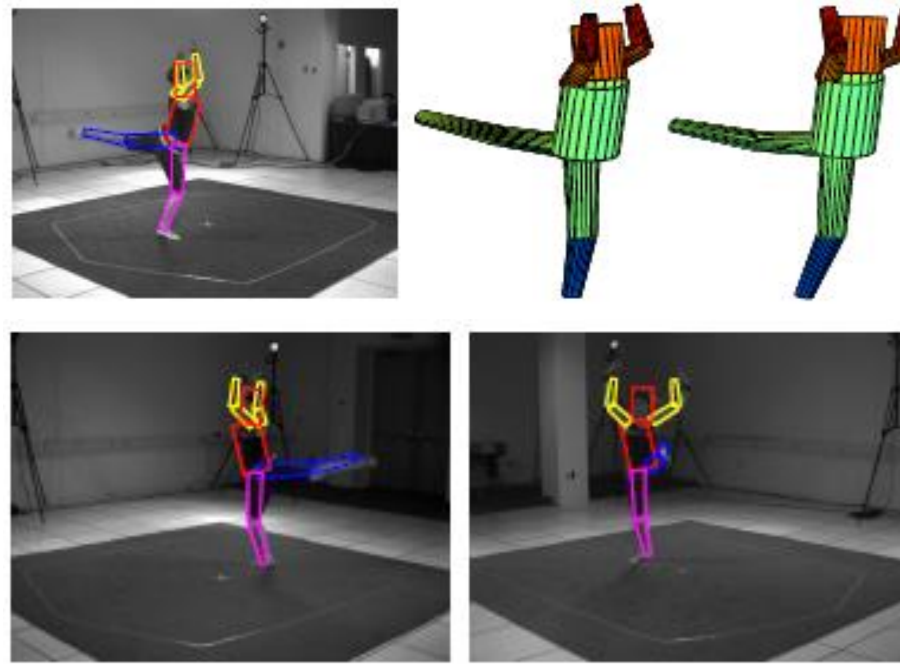


S. Roweis, 2004

Target Tracking



Següiment de múltiples objectius basat en radar

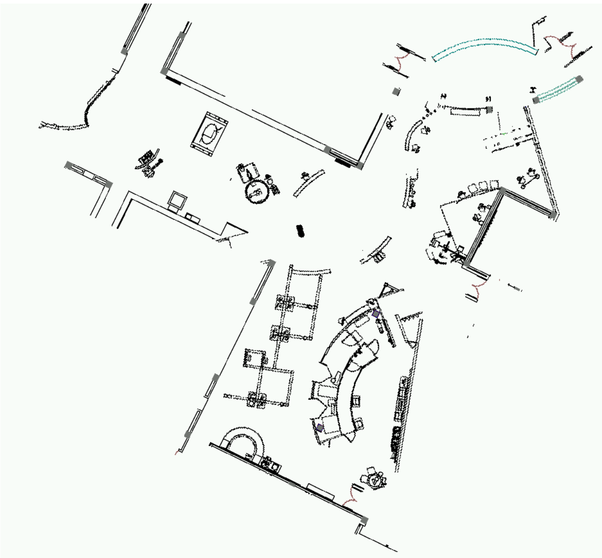


Seguiment visual d'objectes articulats
(L. Sigal et. al., 2006)

- Estimar el moviment de objectes en 3D a partir de mesures indirectes i potencialment sorolloses

Navegació de Robots: *SLAM*

Simultaneous Localization and Mapping



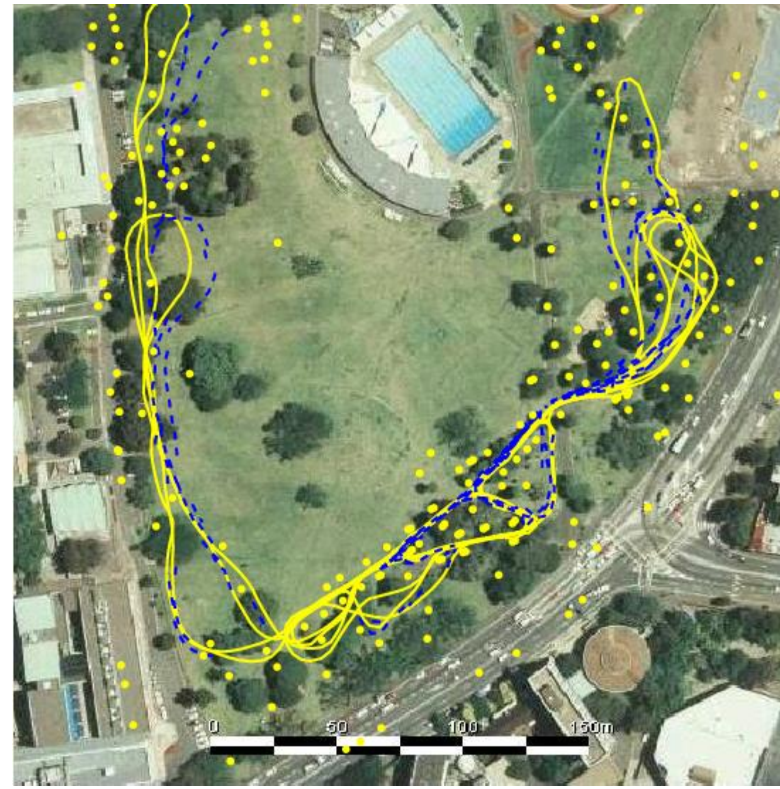
CAD Map

(S. Thrun, San Jose Tech Museum)

Mapa Estimat

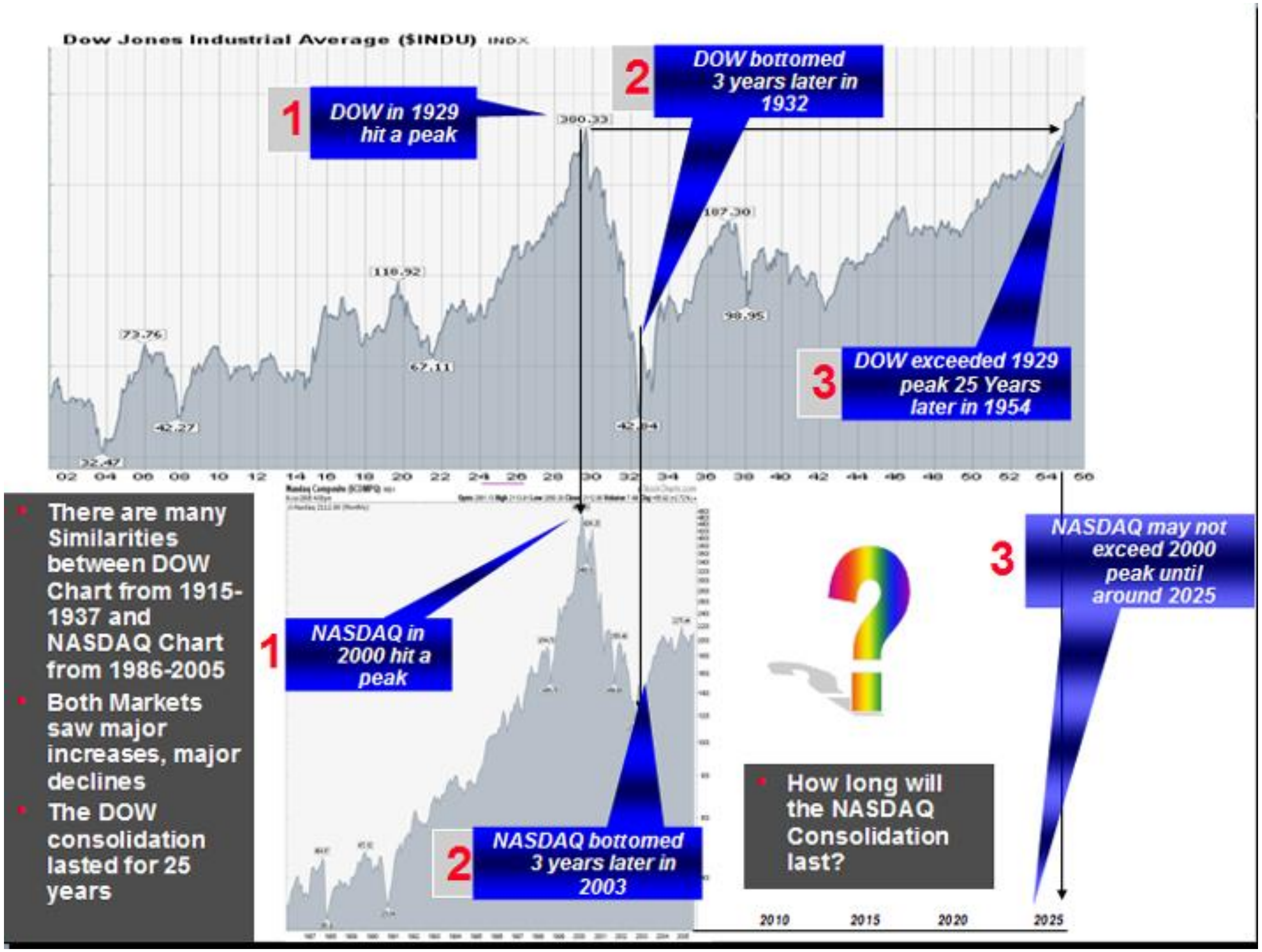
Landmark SLAM

(E. Nebot, Victoria Park)



- Mentre el robot es mou, estimem la posicio i la geometria del món

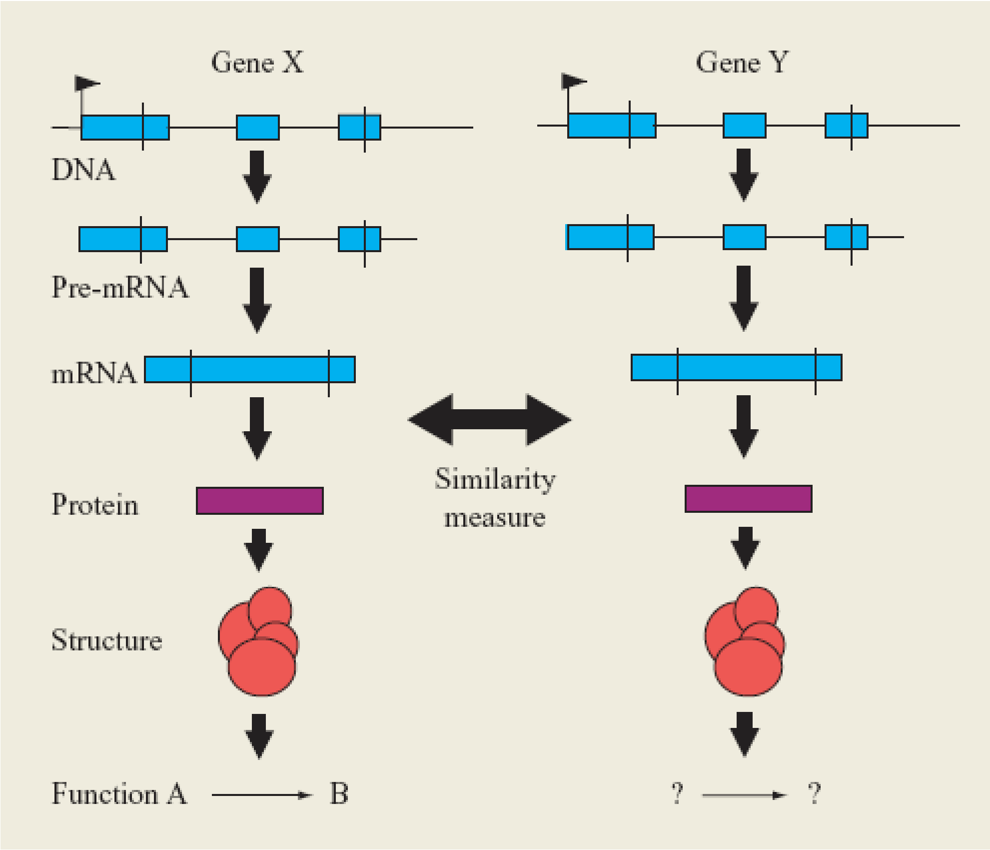
Previsió financera



http://www.steadfastinvestor.com/

- Prediu el comportament futur del mercat a partir de dades històriques, noticies, opinions d'experts, …

Anàlisi de seqüències biològiques



(E. Birney, 2001)

- Els models temporals es poden adaptar a formes més genèriques d’estructures seqüencials, com les que apareixen en seqüències d’ADN

Inferència en el temps

- Hidden Markov Models
- Filtre de Kalman
- Filtres de particules
- Xarxes Bayesianes Dinamiques (DBN)
- …

Processos seqüencials

- Considerem un sistema que pot prendre un dels *N* *estats* o *categories* discrets

$$x_t \in \{1, 2, \ldots, N\} \longrightarrow$$ estat al instant *t*

- Estem interessats en sistemes *estocàstics*, en els que l'evolució del estat segueix una distribució probabilistica
- Qualsevol joint distribution es pot factoritzar en una sèrie de distribucions condicionals (CPD) :

$$p(x_0, x_1, \ldots, x_T) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_0, \ldots, x_{t-1})$$

Processos de Markov

- Per a un procés de *Markov*, el següent estat depèn només del estat actual:

$$p(x_{t+1} \mid x_0, \ldots, x_t) = p(x_{t+1} \mid x_t)$$

- Al mateix temps, implica que

$$p(x_0, \ldots, x_{t-1}, x_{t+1}, \ldots, x_T \mid x_t)$$
$$= p(x_0, \ldots, x_{t-1} \mid x_t) p(x_{t+1}, \ldots, x_T \mid x_t)$$

"Condicionat en el present,
el passat i el futur són independents"

Cadenes de Markov: Model gràfic

$$p(x_0, x_1, \ldots, x_T) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1})$$

$$p(x_0) \quad \begin{tikzcd} x_0 \arrow[r, "p(x_1 \mid x_0)"'] & x_1 \arrow[r, "p(x_2 \mid x_1)"'] & x_2 \arrow[r, "p(x_3 \mid x_2)"'] & x_3 \end{tikzcd}$$

$\bigcirc \Longleftrightarrow$ Valors d'estat en un temps determinat

nodes

$\nearrow \Longleftrightarrow$ Propietats de Markov

arestes

Matrius de transició d'estat

- Una cadena de Markov *estacionaria* amb *N* estats es descriu amb una matriu de transició de *:*

$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

$$q_{ij} \triangleq p(x_{t+1} = i \mid x_t = j)$$

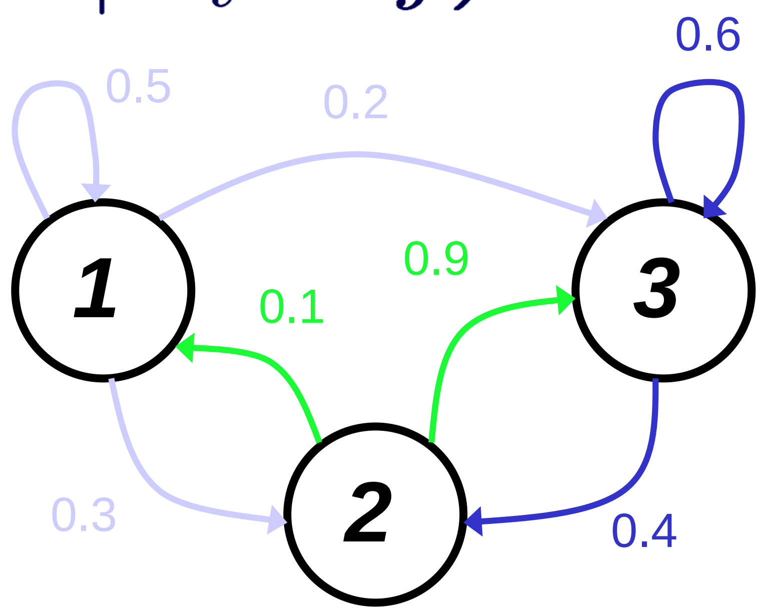
- Restriccions en matrius de transició vàlides:

$$q_{ij} \geq 0 \qquad\qquad \sum_{i=1}^{N} q_{ij} = 1 \quad \text{for all } j$$

Diagrama de transició d'estats

$$q_{ij} \triangleq p(x_{t+1} = i \mid x_t = j)$$

$$Q = \begin{bmatrix} 0.5 & 0.1 & 0.0 \\ 0.3 & 0.0 & 0.4 \\ 0.2 & 0.9 & 0.6 \end{bmatrix}$$

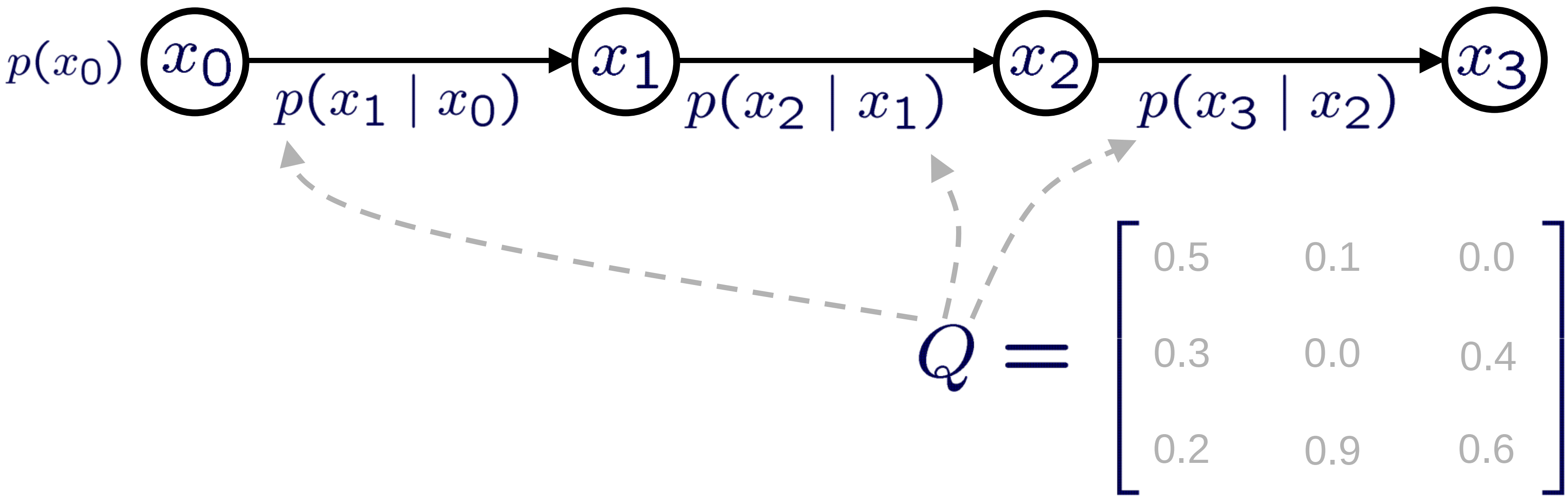


- Pensem en una partícula que segueix una fletxa a l'atzar en cada instant de temps discret.
- Principalment útil quan N és petit i Q *sparse*

http://setosa.io/ev/markov-chains/

Cadenes de Markov: Model gràfic

$$p(x_0, x_1, \ldots, x_T) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1})$$

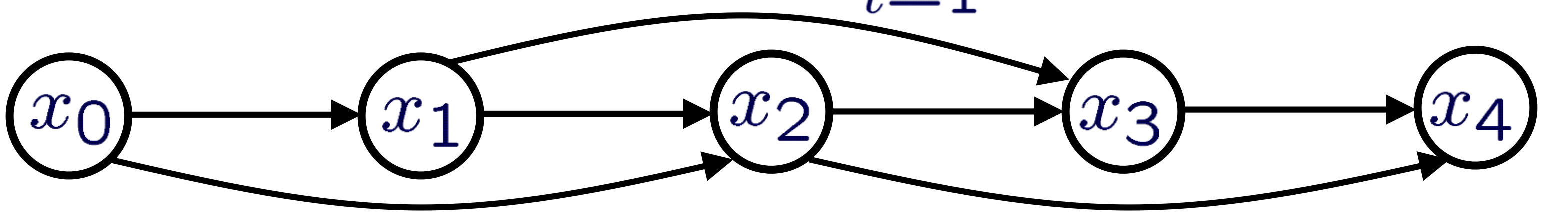


nodes ⟺ Valors d'estat en un temps determinat

arestes ⟺ Propietats de Markov

Cadenes d'ordre superior

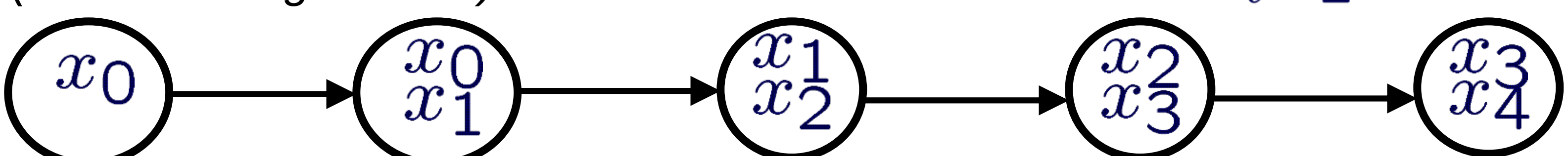
$$p(x_0, x_1, \ldots, x_T) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1}, x_{t-2})$$



- Cada nou estat depèn d'una finestra de longitud fixa dels valors d'estat precedents
- Podem representar-ho com un model de primer odre via augments d'estats:

$$\bar{x}_t \triangleq \{x_t, x_{t-1}\} \qquad p(\bar{x}) = p(\bar{x}_1) \prod_{t=2}^{T} p(\bar{x}_t \mid \bar{x}_{t-1})$$

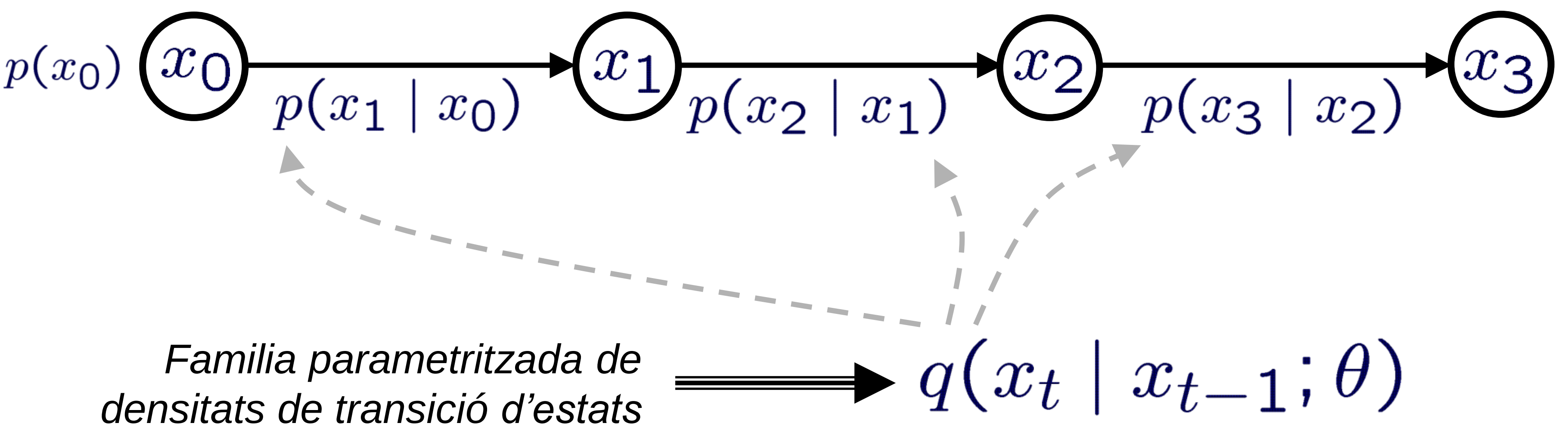
(N^2 estats augmentats)



Procesos d'estat continuus

- En moltes aplicacions, és més natural definir els estats en algun estat continu:

$$x_t \in \mathbb{R}^d$$

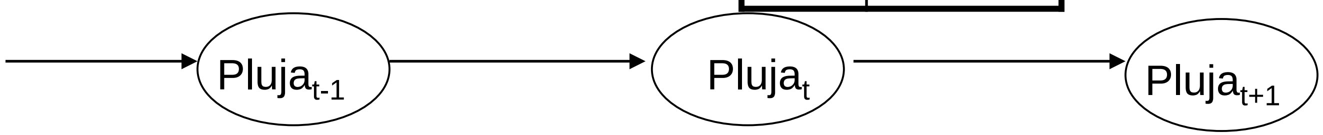


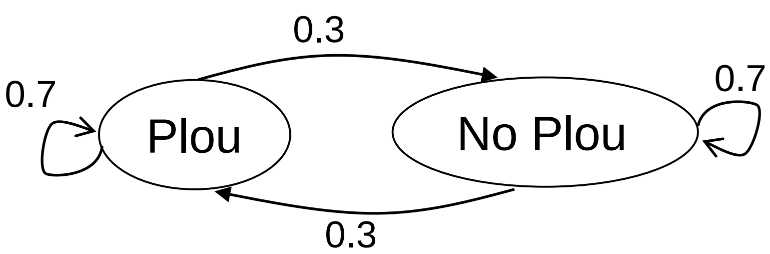
- *Exemples:* preu d'accions, posició d'avions, …

Exemple

P_0	
T	0.5
F	0.5

P_{t-1}	$P(P_t \mid P_{t-1})$
T	0.7
F	0.3





$$Q = \begin{bmatrix} 0{,}7 & 0{,}3 \\ 0{,}3 & 0{,}7 \end{bmatrix}$$

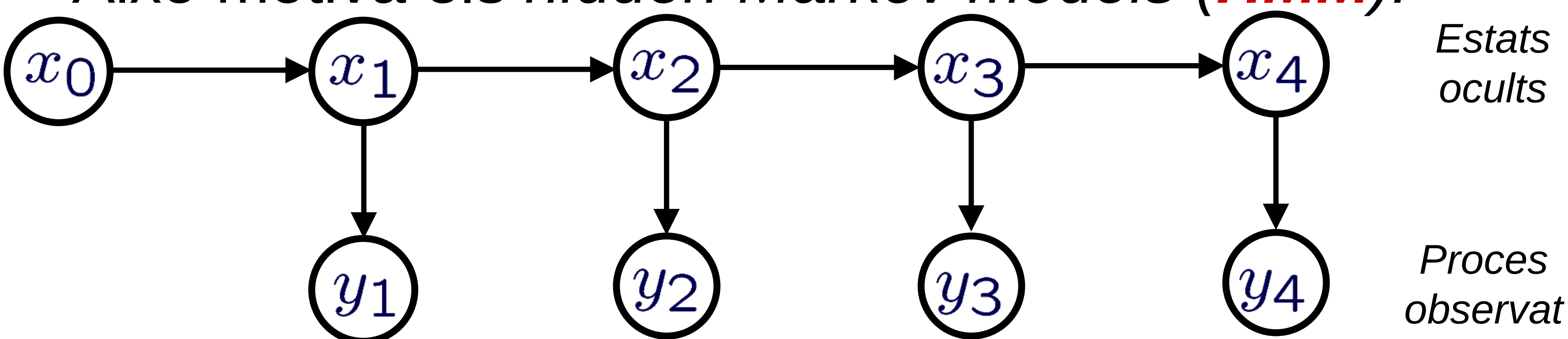
Hidden Markov Models

- Si tenim un model que satisfà les assumpcions dels processos de Markov

"Condicionat en el present,
el passat i el futur són independents"

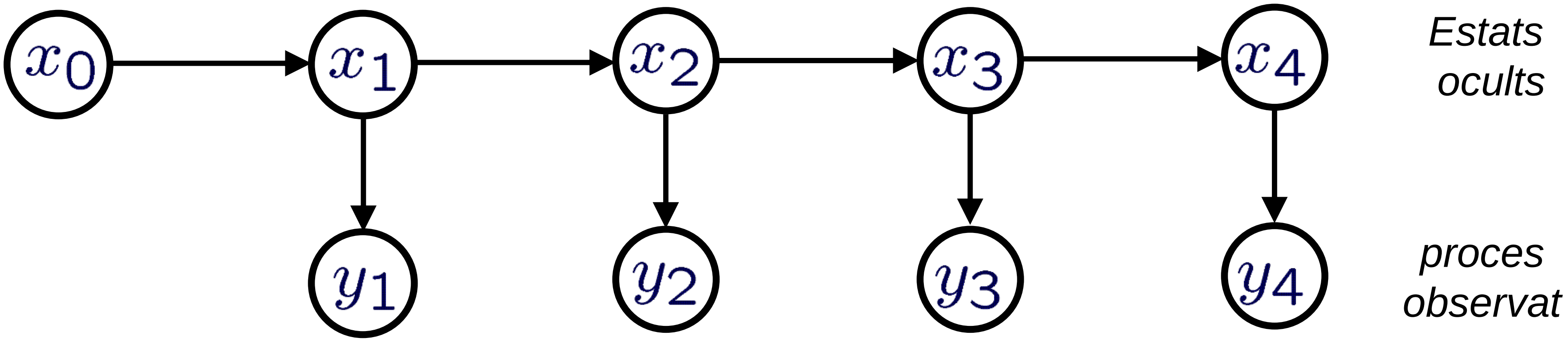
però i si no el podem observar ????

- Això motiva els *hidden Markov models* (***HMM***):



$$p(x_0, x_1, \ldots, x_T) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1}) p(y_t \mid x_t)$$

Estats ocults (hidden states)

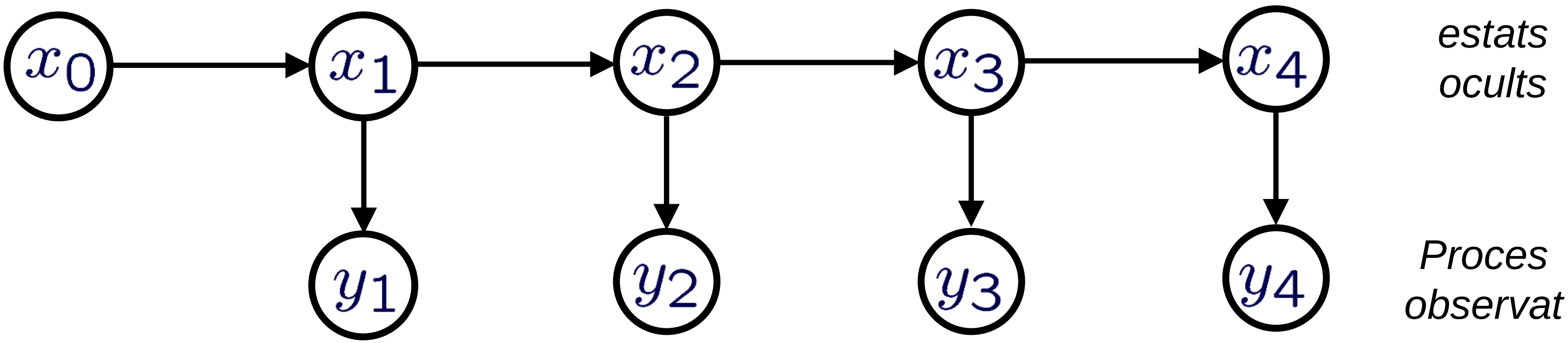


- Donat x_t, les observacions prèvies no proporcionen informació addicional sobre el futur:

$$p(y_t, y_{t+1}, \ldots \mid x_t, y_{t-1}, y_{t-2}, \ldots) = p(y_t, y_{t+1}, \ldots \mid x_t)$$

- Les transformacions de procés sota el que pren la dinàmica prenen forma de primer ordre.

D'on venen els estats?

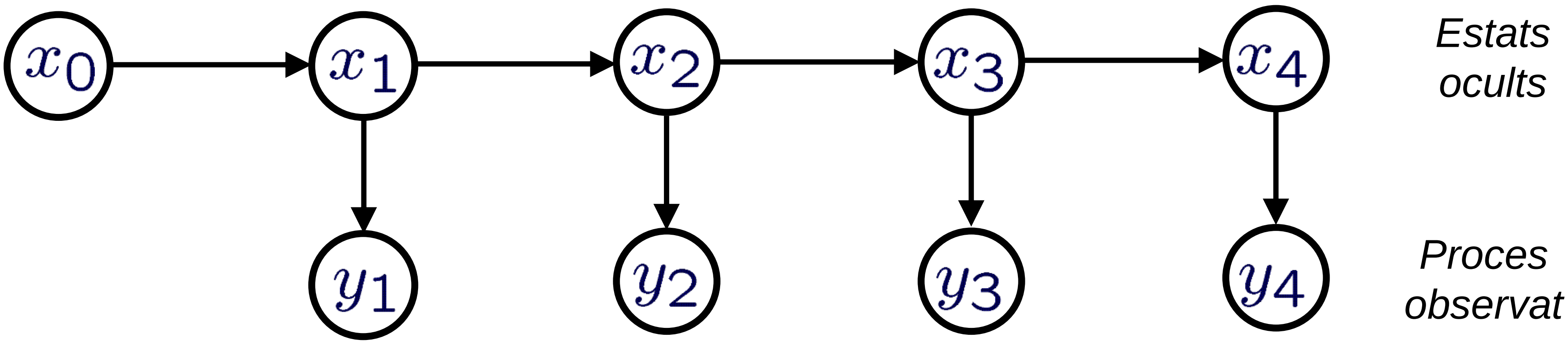


- Anàlisi d'un fenomen físic:
  - Model dinàmics d'un avió o robot
  - Models geofísics d'evolució climàtica
- Obtinguts de dades d'entrenament:
  - Exemples gravats de català parlat
  - Comportament històric dels preus de les accions

http://en.wikipedia.org/wiki/Hidden_Markov_model

HMMs amb estats discrets

$$x_t \in \{1, 2, \ldots, N\}$$



- Associar cada un dels *N* estats ocults amb una distribució d'observació diferent:

$$p(y_t \mid x_t = 1) \qquad p(y_t \mid x_t = 2) \quad \cdots$$

- Les densitats d'observació reflexen el coneixement del domini

HMMs Discrets: Observacions

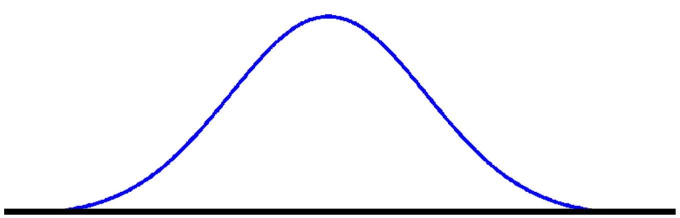
Observacions Discretes

$$y_t \in \{1, 2, \ldots, M\}$$

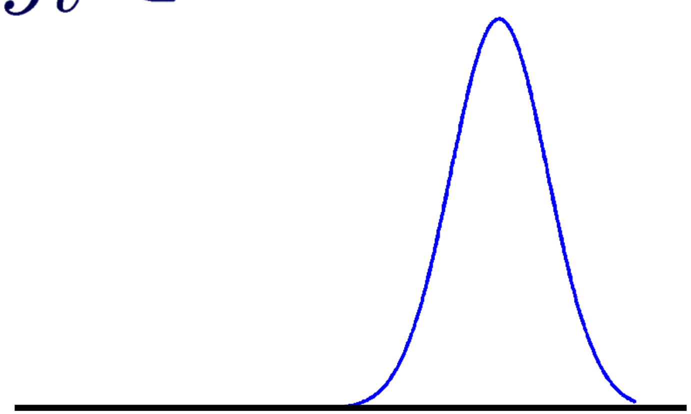
$$p(y_t \mid x_t = 1) = \begin{bmatrix} 0.3 \\ 0.1 \\ 0.5 \\ 0.1 \end{bmatrix} \qquad p(y_t \mid x_t = 2) = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.1 \\ 0.5 \end{bmatrix}$$

Observacions Continues

$$y_t \in \mathbb{R}^k$$

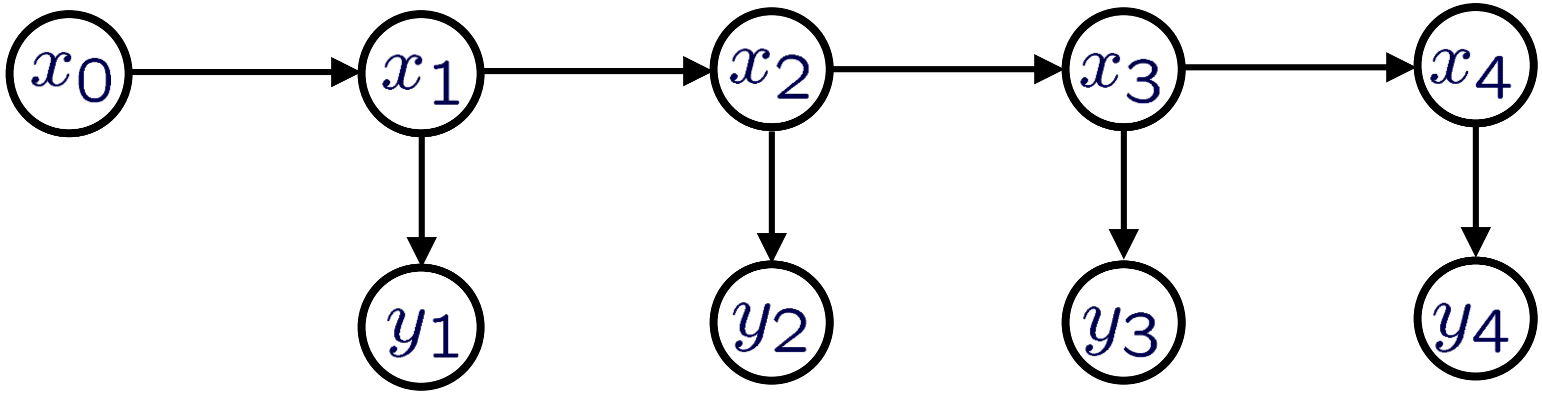


$$p(y_t \mid x_t = 1)$$



$$p(y_t \mid x_t = 2)$$

Especificació d'una HMM



- Model d'observació: π $\quad P(y_i|x_i) \quad m \times n$

Quan tenim l'observació només usem una columna π_i

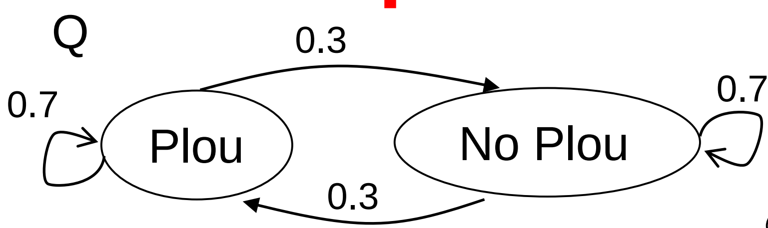
- Model de transició: Q $\quad P(x_i|x_{i-1}) \quad n \times n$

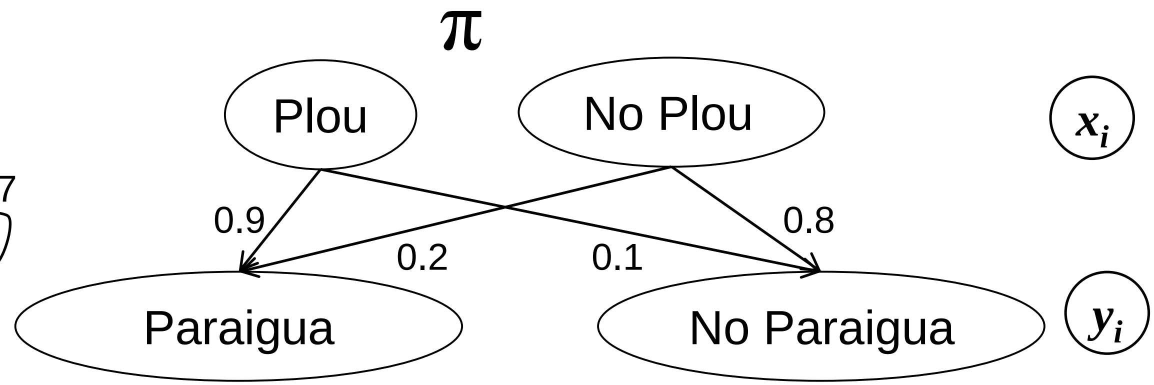
- Distribució d'estat inicial: $\quad P(x_0) \quad n \times 1$

n: número de possibles valors (estats) dels nodes ocults

m: número de possibles valors de les observacions

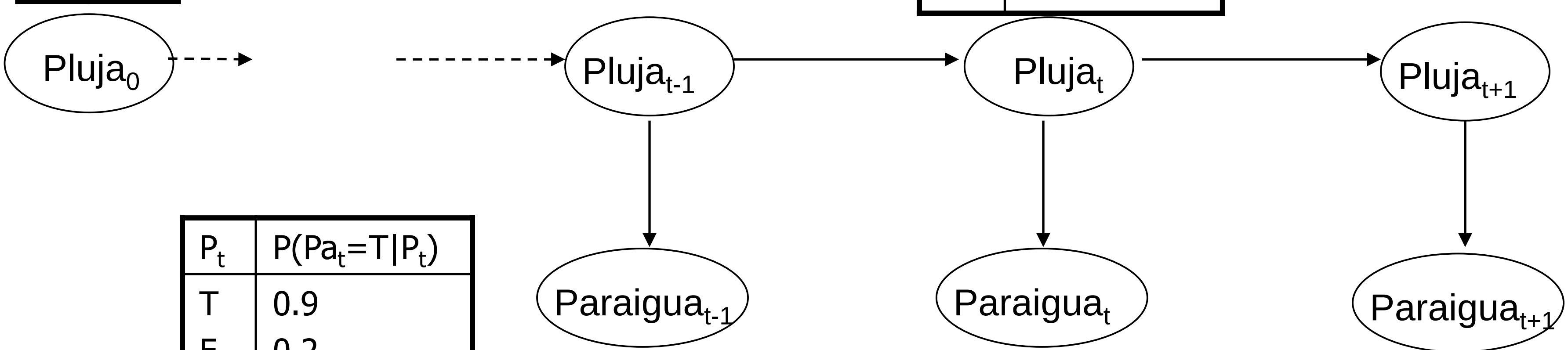
Exemple





P_0	
T	0.5
F	0.5

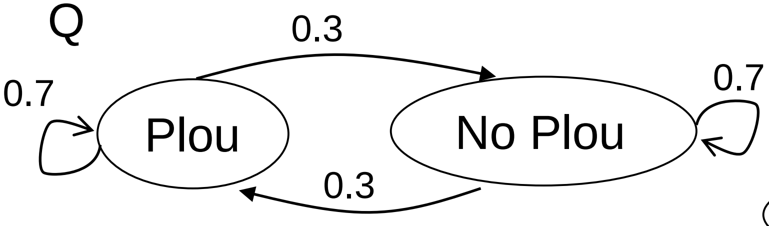
P_{t-1}	$P(P_t=T\|P_{t-1})$
T	0.7
F	0.3

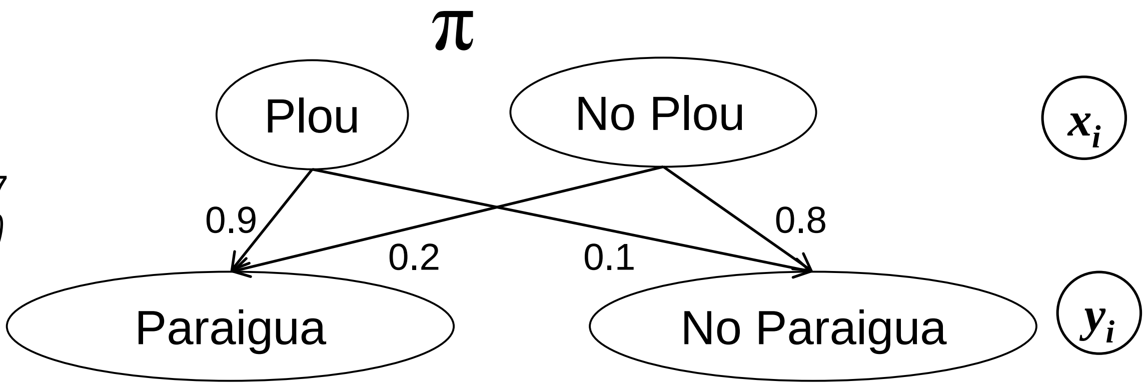


P_t	$P(Pa_t=T\|P_t)$
T	0.9
F	0.2

Quina és la probabilitat de que plogui el dia 2, sabent que el dia 1 i el dia 2 hem vist el paraigua?

Exemple

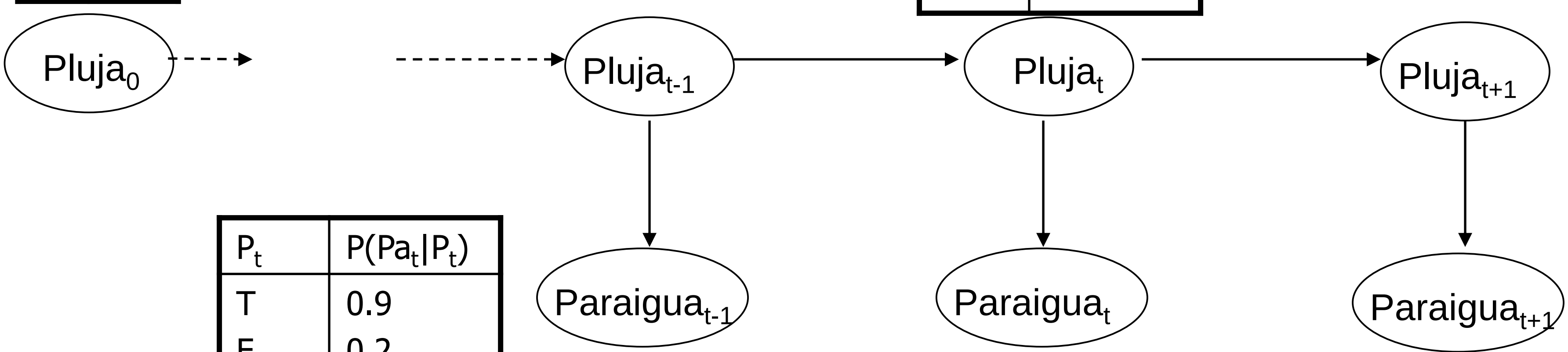




Q	Plou	NoPlou
Plou	0.7	0.8
No Plou	0.3	0.2

P_0	
T	0.5
F	0.5

P_{t-1}	$P(P_t \mid P_{t-1})$
T	0.7
F	0.8



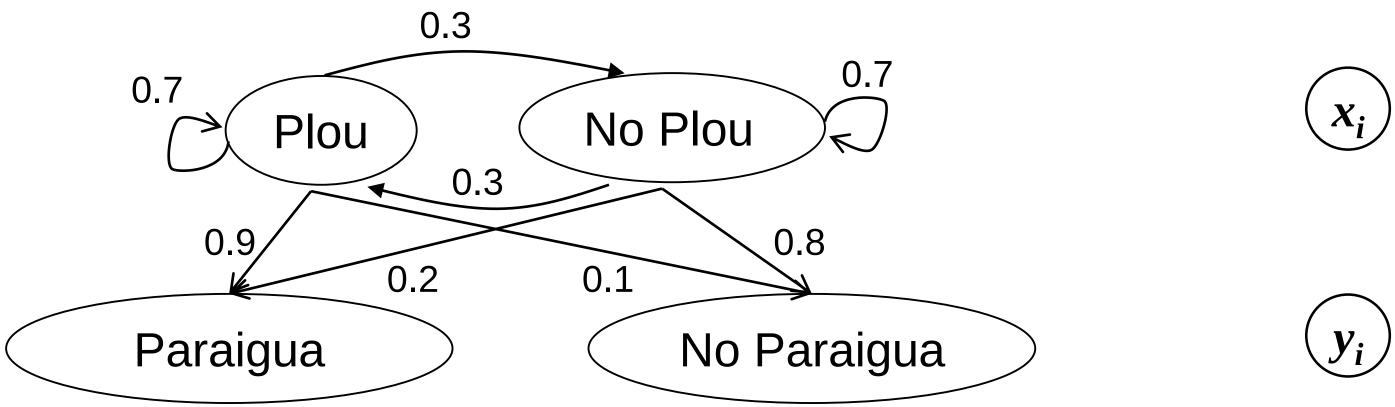
P_t	$P(Pa_t \mid P_t)$
T	0.9
F	0.2

π	plou	noplou
paraigua	0.9	0.2
noparaigua	0.1	0.8

plou, no plou, paraigua, no paraigua

NO són variables, són estats

Exemple



P_0	
T	0.5
F	0.5

Q	Plou	NoPlou
Plou	0.7	0.8
No Plou	0.3	0.2

π	plou	noplou
paraigua	0.9	0.2
noparaigua	0.1	0.8

Tasques d'inferència

- **Filtratge (Filtering o monitoring)**: $P(X_t|y_1,\ldots,y_t)$

  calcular la certesa sobre l'estat actual, donades totes les evidencies fins al moment

- **Predicció**: $P(X_{t+k}|y_1,\ldots,y_t)$

  calcular la probabilitat d'un estat futur

- **Suavitzat (Smoothing)**: $P(X_k|y_1,\ldots,y_t)$ $k<t$

  calcular la probabilitat d'un estat pasat (retrospectiva)

- **Explicació més plausible (Decoding)**: $\arg\max_{x1,..xt} P(x_1,\ldots,x_t|y_1,\ldots,y_t)$

  donada una seqüència d'observacions, trobar la seqüència d'estats més probable que hagi pogut generar les observacions.

Exemples

- **Filtratge:** Quina és la probabilitat que plogui avui, donada totes les observacions de paraigües fins avui?
- **Predicció:** Quina és la probabilitat que plogui passat demà, donades totes les observacions de paraigua fins avui?
- **Smoothing:** Quina és la probabilitat que ahir plogués, donades totes les observacions de paraigua fins avui?
- **Explicació més plausible:** si el paraigua ha aparegut els primers tres dies i no el quart, quin és la seqüència de temps més plausible per a produir les observacions de paraigua?

Filtratge i Suavització (filtering & smoothing)

- Per anàlisi de dades online, busquem l'estimació d'estat filtrada donades les observacions anteriors:

$$p(x_t \mid y_1, y_2, \ldots, y_t) \qquad t = 1, 2, \ldots$$

- En altres casos, volem trobar estimacions suavitzades donades les observacions prèvies i posteriors:

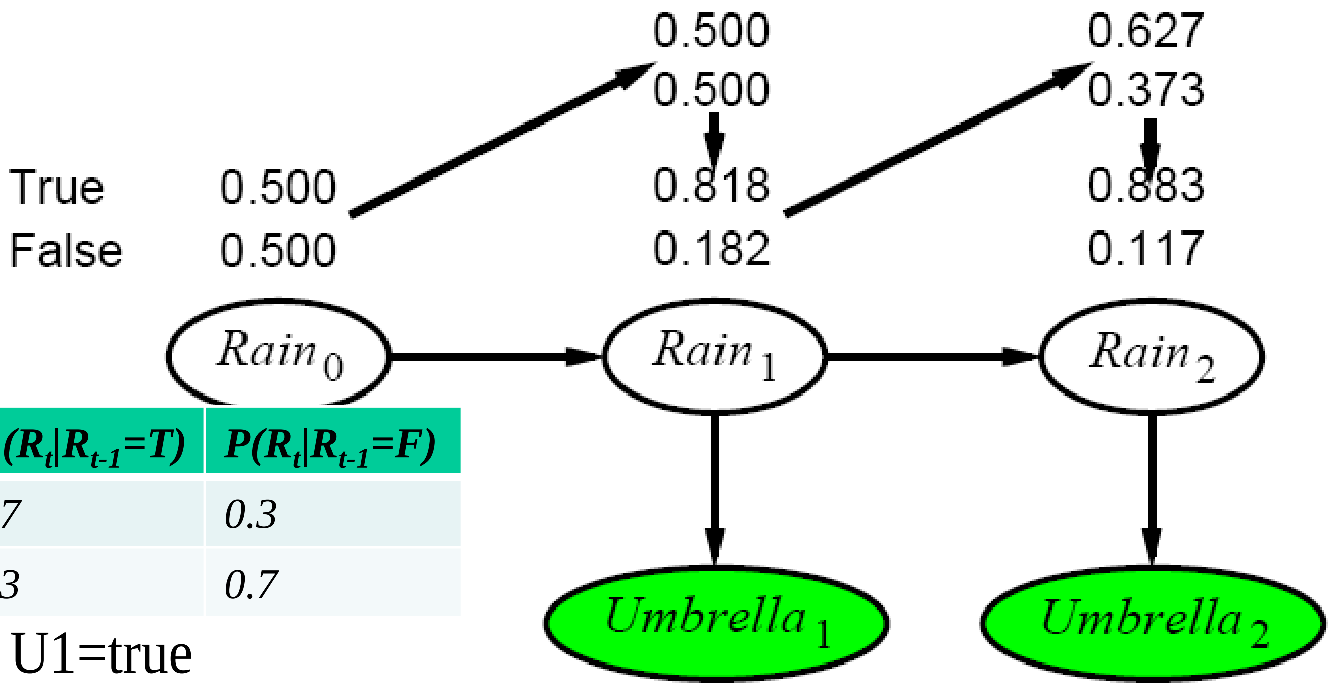
$$p(x_t \mid y_1, y_2, \ldots, y_T) \qquad t = 1, 2, \ldots, T$$

- Depenent com variem T:

fixed-lag smoothing & *predicció*:

$$p(x_t \mid y_1, \ldots, y_{t+\delta}) \qquad p(x_t \mid y_1, \ldots, y_{t-\delta})$$

Exemple Filtering



R_t	$P(R_t \mid R_{t-1}=T)$	$P(R_t \mid R_{t-1}=F)$
T	*0.7*	*0.3*
F	*0.3*	*0.7*

Rt	$P(Ut \mid Rt)$
T	*0.9*
F	*0.2*

Day 1, U1=true

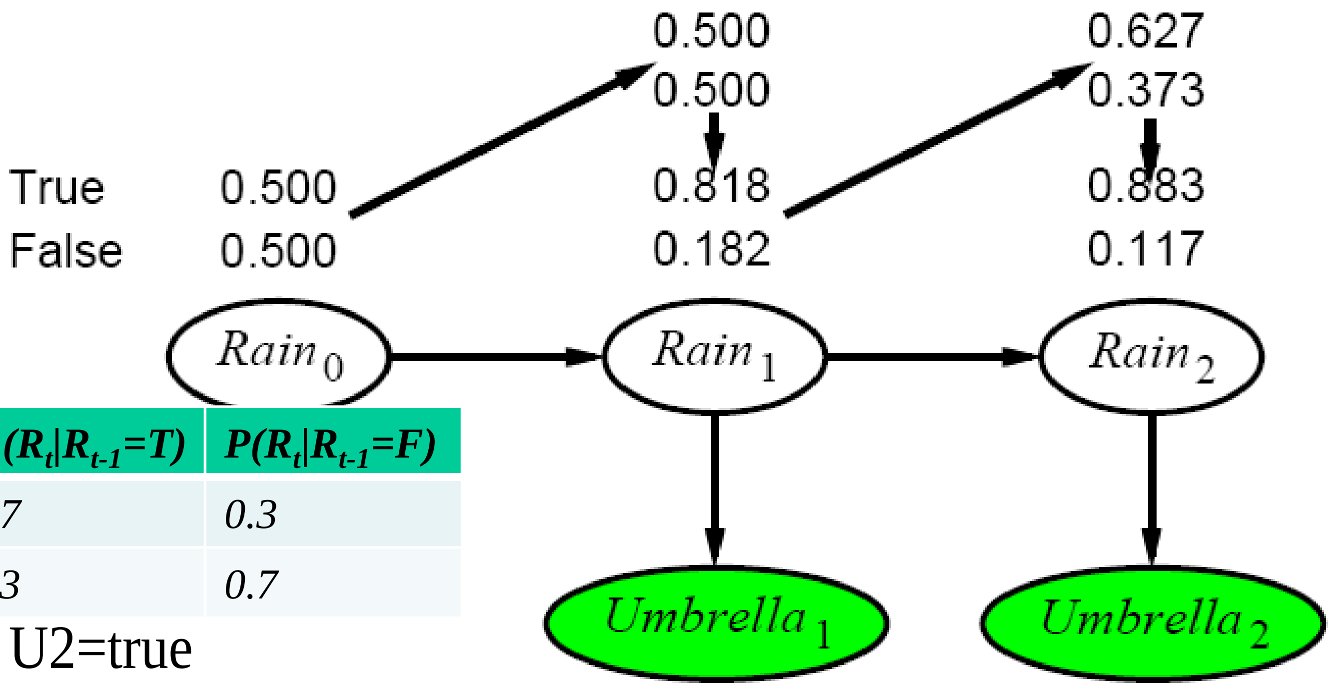
$$p(R_1) = \sum_{r_0} p(R_1 \mid r_0) p(r_0)$$

$$= < 0.7, 0.3 > \times 0.5 + < 0.3, 0.7 > \times 0.5 = < 0.5, 0.5 >$$

$$p(R_1 \mid u_1) = \alpha\, p(u_1 \mid R_1) p(R_1) = \alpha < 0.9, 0.2 > \times < 0.5, 0.5 >$$

$$= \alpha < 0.45, 0.1 > = < 0.818, 0.182 >$$

Exemple Filtering



R_t	$P(R_t \mid R_{t-1}=T)$	$P(R_t \mid R_{t-1}=F)$
T	*0.7*	*0.3*
F	*0.3*	*0.7*

Rt	*P(Ut|Rt)*
T	*0.9*
F	*0.2*

Day 2, U2=true

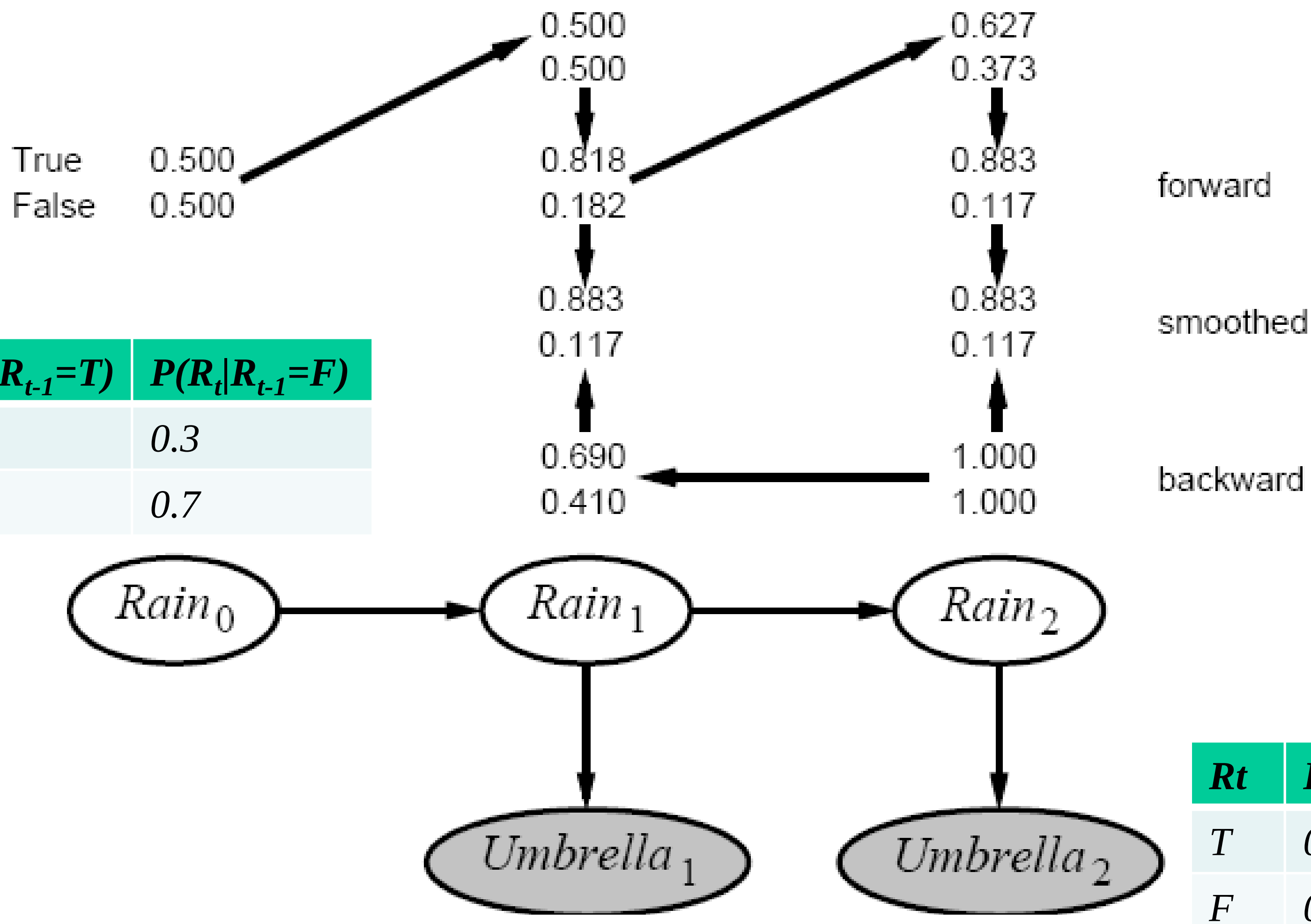
$$p(R_2 \mid u_1) = \sum_{r_1} p(R_2 \mid r_1) p(r_1 \mid u_1)$$

$$= \langle 0.7, 0.3 \rangle \times 0.818 + \langle 0.3, 0.7 \rangle \times 0.182 = \langle 0.627, 0.373 \rangle$$

$$p(R_2 \mid, u_1, u_2) = \alpha\, p(u_2 \mid R_2) p(R_2 \mid u_1) = \alpha \langle 0.9, 0.2 \rangle \times \langle 0.627, 0.373 \rangle$$

$$= \alpha \langle 0.565, 0.075 \rangle = \langle 0.883, 0.117 \rangle$$

Exemple Smoothing



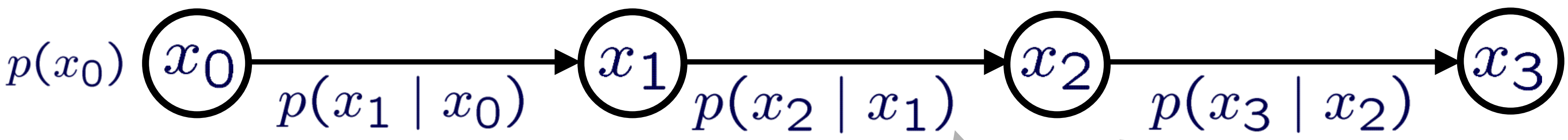
R_t	$P(R_t \mid R_{t-1}=T)$	$P(R_t \mid R_{t-1}=F)$
T	*0.7*	*0.3*
F	*0.3*	*0.7*

Rt	$P(Ut \mid Rt)$
T	*0.9*
F	*0.2*

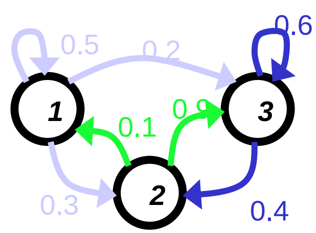
$$p(\mathrm{u}_2 \mid \mathrm{R}_1) = \sum_{r_2} p(u_2 \mid r_2)\, p(\mid r_2)\mathrm{p}(r_2 \mid R_1) =$$

$$= 0.9 \times 1 < 0.7, 0.3 > + 0.2 \times 1 \times < 0.3, 0.7 > = < 0.69, 0.41 >$$

Estadísitica de les cadenes de Markov



$$\alpha_t \triangleq \begin{bmatrix} p(x_t = 1) \\ p(x_t = 2) \\ p(x_t = 3) \end{bmatrix}$$



$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

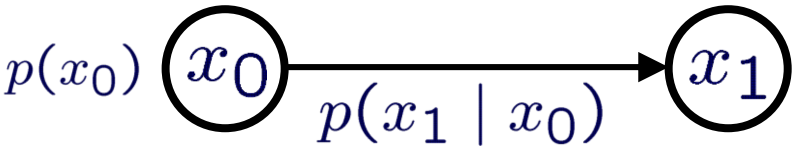
$$q_{ij} \triangleq p(x_{t+1} = i \mid x_t = j)$$

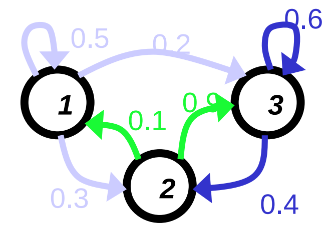
- Per definició de probabilitats condicionals,

$$\alpha_1(i) = \sum_{j=1}^{N} q_{ij}\alpha_0(j)$$

$$\begin{cases} \alpha_1 = Q\alpha_0 \\ \alpha_t = Q^t\alpha_0 \\ \alpha_\infty = ??? \end{cases}$$

Estadísitica de les cadenes de Markov





$$\alpha_t \triangleq \begin{bmatrix} p(x_t = 1) \\ p(x_t = 2) \\ p(x_t = 3) \end{bmatrix}$$

$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

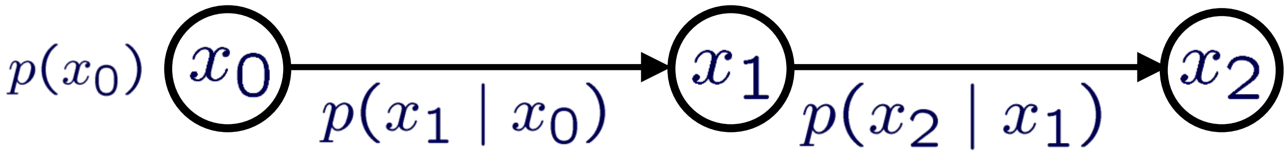
$$q_{ij} \triangleq p(x_{t+1} = i \mid x_t = j)$$

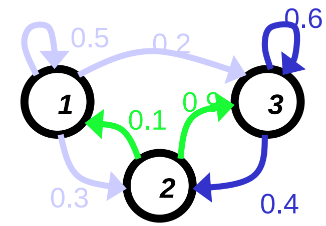
- Per definició de probabilitats condicionals,

$$P(x_1 = i) = \sum_j^n P(x_1 = i, x_0 = j) = \sum_j^n \underbrace{P(x_1 = i | x_0 = j)}_{q_{ij}} \underbrace{P(x_0 = j)}_{\alpha_0(j)} = Q_{i\cdot} \times \alpha_0 = \alpha_1(i)$$

$$\alpha_1 = Q\ \alpha_0$$

Estadísitica de les cadenes de Markov





$$\alpha_t \triangleq \begin{bmatrix} p(x_t = 1) \\ p(x_t = 2) \\ p(x_t = 3) \end{bmatrix}$$

$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

$$q_{ij} \triangleq p(x_{t+1} = i \mid x_t = j)$$

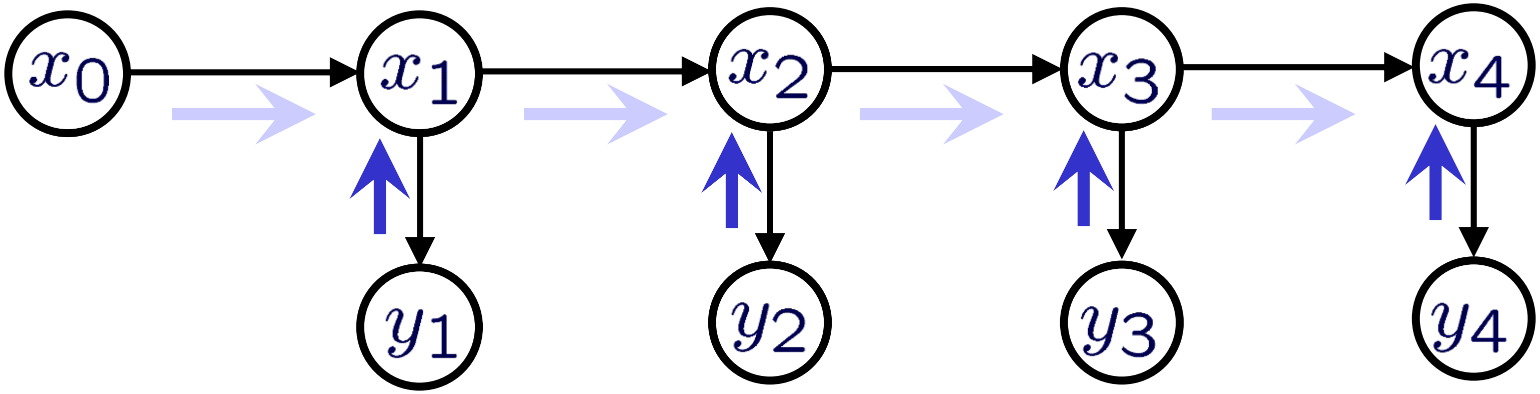
- Per definició de probabilitats condicionals,

$$P(x_2 = i) = \sum_k^n \sum_j^n P(x_2 = i, x_1 = k, x_0 = j) = \sum_k^n \sum_j^n P(x_2 = i | x_1 = k) P(x_1 = i | x_0 = j)\, P(x_0 = j)$$

$$= \sum_k^n P(x_2 = i | x_1 = k) \alpha_1(k) \; = Q_{i\cdot} \times \alpha_1 = \alpha_1(i)$$

$$\alpha_2 = Q\ \alpha_1 = Q\ Q\ \alpha_0 = Q^2\ \alpha_0$$

HMMs discrets: Filtering



$$\alpha_t(x_t) \triangleq p(x_t \mid y_1, \ldots, y_t)$$

$$= \frac{1}{Z_t} \underbrace{p(y_t \mid x_t)}_{\text{actualització}} \underbrace{\sum_{x_{t-1}} p(x_t \mid x_{t-1}) \alpha_{t-1}(x_{t-1})}_{\text{Predicció}}$$

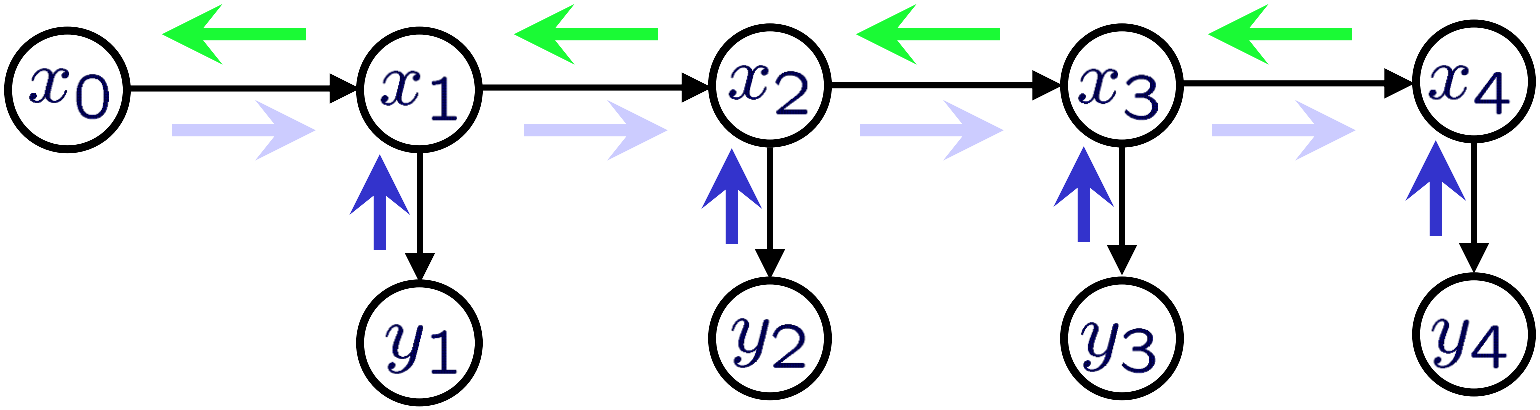
Constant de Normalització

Predicció: $p(x_t \mid y_1, \ldots, y_{t-1})$

actualització: $p(x_t \mid y_1, \ldots, y_t)$

Incorpora T observacions en $\mathcal{O}(TN^2)$ operacions

HMMs discrets: Smoothing



$$p(x_t \mid y) \propto \underbrace{p(x_t \mid y_1, \ldots, y_t)}_{\alpha_t(x_t)} \underbrace{p(y_{t+1}, \ldots, y_T \mid x_t)}_{\beta_t(x_t)}$$

- L'algorisme *forward-backward* actualitza a partir d'una recursió en temps invers:

$$\beta_t(x_t) = \frac{1}{Z_t} \sum_{x_{t+1}} p(x_{t+1} \mid x_t) p(y_{t+1} \mid x_{t+1}) \beta_{t+1}(x_{t+1})$$

resum

$$p(x_t \mid y) \propto \underbrace{p(x_t \mid y_1, \ldots, y_t)}_{\alpha_t(x_t)} \underbrace{p(y_{t+1}, \ldots, y_T \mid x_t)}_{\beta_t(x_t)}$$

$$\alpha_t(x_t) \triangleq \frac{1}{Z_t} p(y_t \mid x_t) \sum_{x_{t-1}} p(x_t \mid x_{t-1}) \alpha_{t-1}(x_{t-1})$$

$$\beta_t(x_t) = \frac{1}{Z_t} \sum_{x_{t+1}} p(x_{t+1} \mid x_t) p(y_{t+1} \mid x_{t+1}) \beta_{t+1}(x_{t+1})$$

- Filtering: $p(x_t | y_1, \ldots, y_t) = \alpha_t(x_t)$
- Smoothing: $p(x_t | y) = \alpha_t(x_t)\beta_t(x_t)$

càlcul

$$p(x_t \mid y) \propto \underbrace{p(x_t \mid y_1, \ldots, y_t)}_{\alpha_t(x_t)} \underbrace{p(y_{t+1}, \ldots, y_T \mid x_t)}_{\beta_t(x_t)}$$

$$\alpha_t(x_t) \triangleq \frac{1}{Z_t} p(y_t \mid x_t) \sum_{x_{t-1}} p(x_t \mid x_{t-1}) \alpha_{t-1}(x_{t-1})$$

- Filtering:

$$p(x_t \mid y_1, \ldots, y_t) = \alpha_t(x_t)$$

$$\alpha_t(x_t) = diag(\pi_{y_t}) \cdot Q^{\mathrm{T}} \cdot \alpha_{t-1}(x_{t-1})$$

càlcul

$$p(x_t \mid y) \propto \underbrace{p(x_t \mid y_1, \ldots, y_t)}_{\alpha_t(x_t)} \underbrace{p(y_{t+1}, \ldots, y_T \mid x_t)}_{\beta_t(x_t)}$$

$$\alpha_t(x_t) \triangleq \frac{1}{Z_t} p(y_t \mid x_t) \sum_{x_{t-1}} p(x_t \mid x_{t-1}) \alpha_{t-1}(x_{t-1})$$

$$\beta_t(x_t) = \frac{1}{Z_t} \sum_{x_{t+1}} p(x_{t+1} \mid x_t) p(y_{t+1} \mid x_{t+1}) \beta_{t+1}(x_{t+1})$$

- smoothing:

$$p(x_t \mid y_1, \ldots, y_T) = \alpha_t(x_t) \cdot \beta_t(x_t)$$

$$\beta_t(x_t) = Q \cdot diag(\pi_{y_{t+1}}) \cdot \beta_{t+1}(x_{t+1})$$

HMMs discrets: Explicació més plausible Algorisme de Viterbi

$$\hat{x} = \arg\max_{x} \; p(x_0, x_1, \ldots, x_T \mid y_1, \ldots, y_T)$$

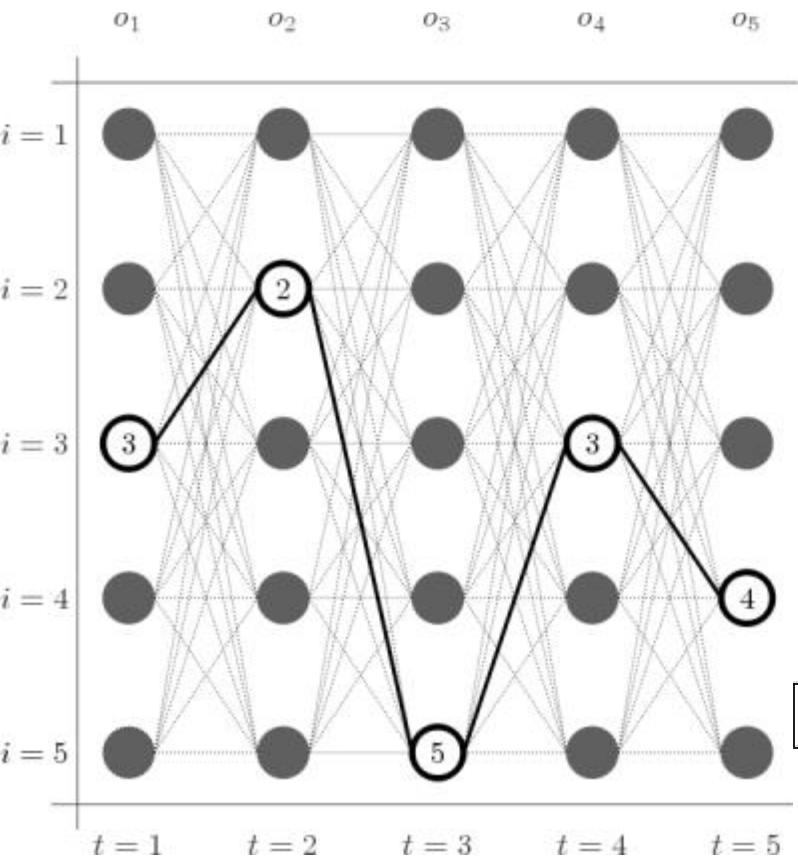
- Usa programació dinàmica per trobar recursivament la probabilitat de la seqüència d'estats més plausible que acaba amb cada un dels $x_t \in \{1, \ldots, N\}$

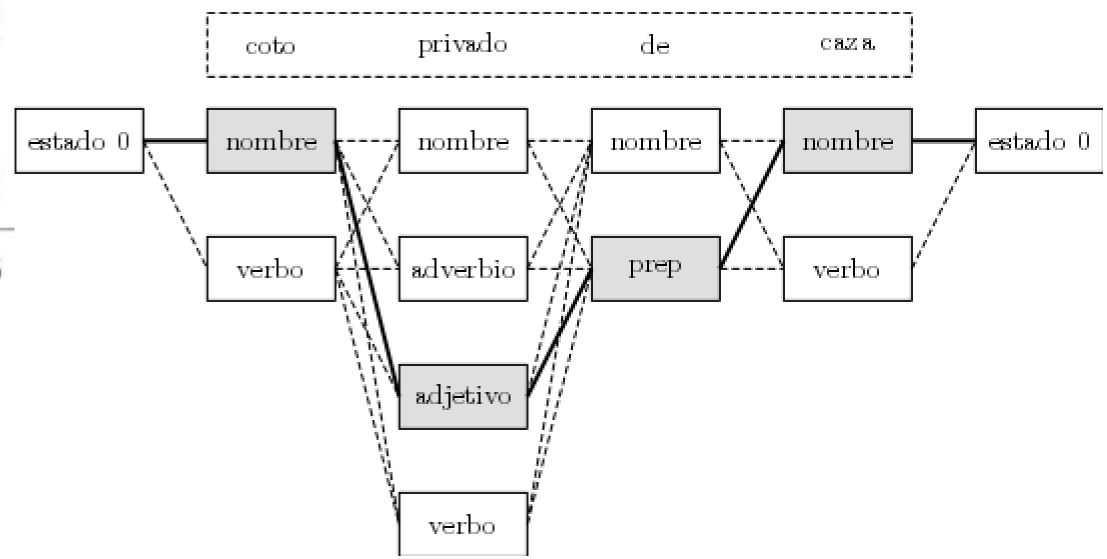
$$\gamma_t(x_t) \triangleq \max_{x_{1,\ldots,x_{t-1}}} p(x_1, \ldots, x_{t-1}, x_t \mid y_1, \ldots, y_t)$$

$$\propto p(y_t \mid x_t) \cdot \left[\max_{x_{t-1}} \; p(x_t \mid x_{t-1}) \gamma_{t-1}(x_{t-1}) \right]$$

$\gamma_t(i)$ És la 'millor' probabilitat d'haver d'estar en l'estat 'i' havent observat les t primeres entrades

- Un procediment de backtracking 'reverse-time' tria la seqüència d'estats màximitzadora

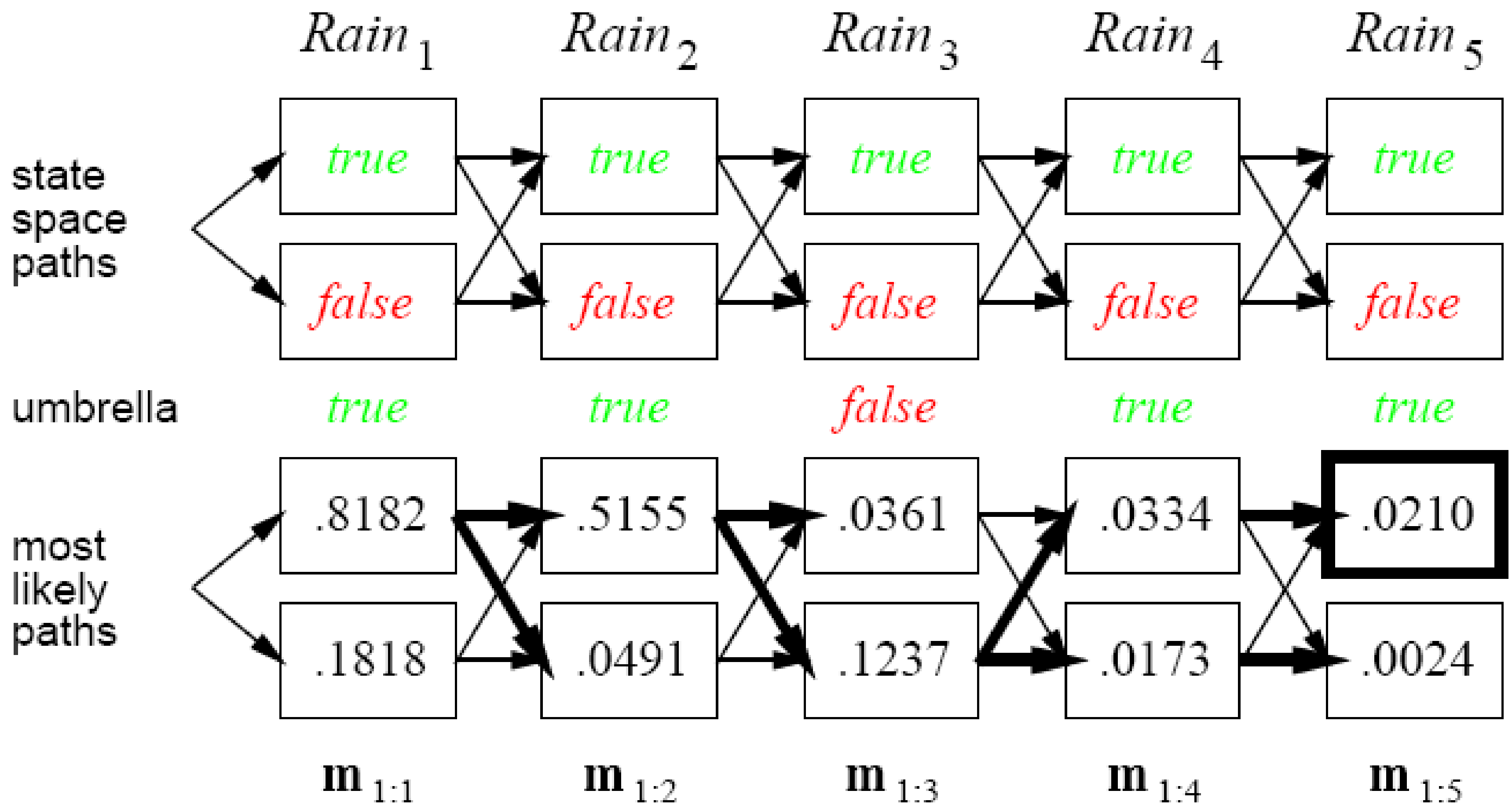

o_1
o_2
o_3
o_4
o_5
$i = 1$
$i = 2$
$i = 3$
$i = 4$
$i = 5$
3
2
5
3
4
$t = 1$
$t = 2$
$t = 3$
$t = 4$
$t = 5$


coto
privado
de
caza
estado 0
nombre
nombre
nombre
nombre
estado 0
verbo
adverbio
prep
verbo
adjetivo
verbo

Exemple Viterbi

π	plou	noplou
paraigua	0.9	0.2
noparaigua	0.1	0.8

Q	Plou	NoPlou
Plou	0.7	0.8
No Plou	0.3	0.2



$$\gamma_t(x_t) \propto p(y_t \mid x_t) \cdot \left[\max_{x_{t-1}} \; p(x_t \mid x_{t-1}) \gamma_{t-1}(x_{t-1}) \right]$$

HMMs discrets: Learning

- Suposem que tenim la sequencia dels estats latents durant el training
- La *maximum likelihood* és

$$(\hat{Q}, \hat{\theta}) = \arg\max_{Q,\theta} \; p(x \mid Q) p(y \mid x, \theta)$$

$$Q = \left[q_{ij}\right] = \left[p(x_{t+1} = i \mid x_t = j)\right]$$

$$\theta = \{\theta_i\}_{i=1}^{N}$$ *(distribucions d'observació)*

- Per simplicitat, assumim que les observacions són gausianes amb variança i mitjana conegudes θ_i

HMMs discrets: Learning

- L'estimador ML de la matriu de transició ve donada per 'comptadors normalitzats' :

$$n(i,j) \triangleq \left\{ \begin{array}{l} \text{Número de vegades} \\ \text{Observat anteriorment} \end{array} \right. \quad \begin{array}{l} x_t = j \\ x_{t+1} = i \end{array}$$

$$\hat{q}_{ij} = \frac{n(i,j)}{\sum_k n(k,j)}$$

- Donat x, independentment estimar la distribució de sortida per a cada estat:

$$\hat{\theta}_i = \frac{1}{|\tau_i|} \sum_{t \in \tau_i} y_t \qquad \tau_i \triangleq \{t \mid x_t = i\}$$

HMMs discrets: algorisme EM

- A la pràctica normalment no coneixem els estats ocults en les seqüències d'entrenament
- L'algorisme EM iterativament maximitza un 'lower bound' en la likelihood de les dades:

E-Step: Usem els parametres actuals per estimar l'estat

$$\widehat{p}^{(s)}(x) = p(x \mid y, \widehat{Q}^{(s-1)}, \widehat{\theta}^{(s-1)})$$

M-Step: Usem estimadors d'estat 'soft' per actualitzar paràmetres

$$(\widehat{Q}^{(s)}, \widehat{\theta}^{(s)}) = \arg\max_{Q,\theta} \sum_{x} \widehat{p}^{(s)}(x) \log p(x \mid Q) p(y \mid x, \theta)$$

Aplicat a HMMs, és equivalent a l'algorisme Baum-Welch

HMMs discrets: algorisme EM

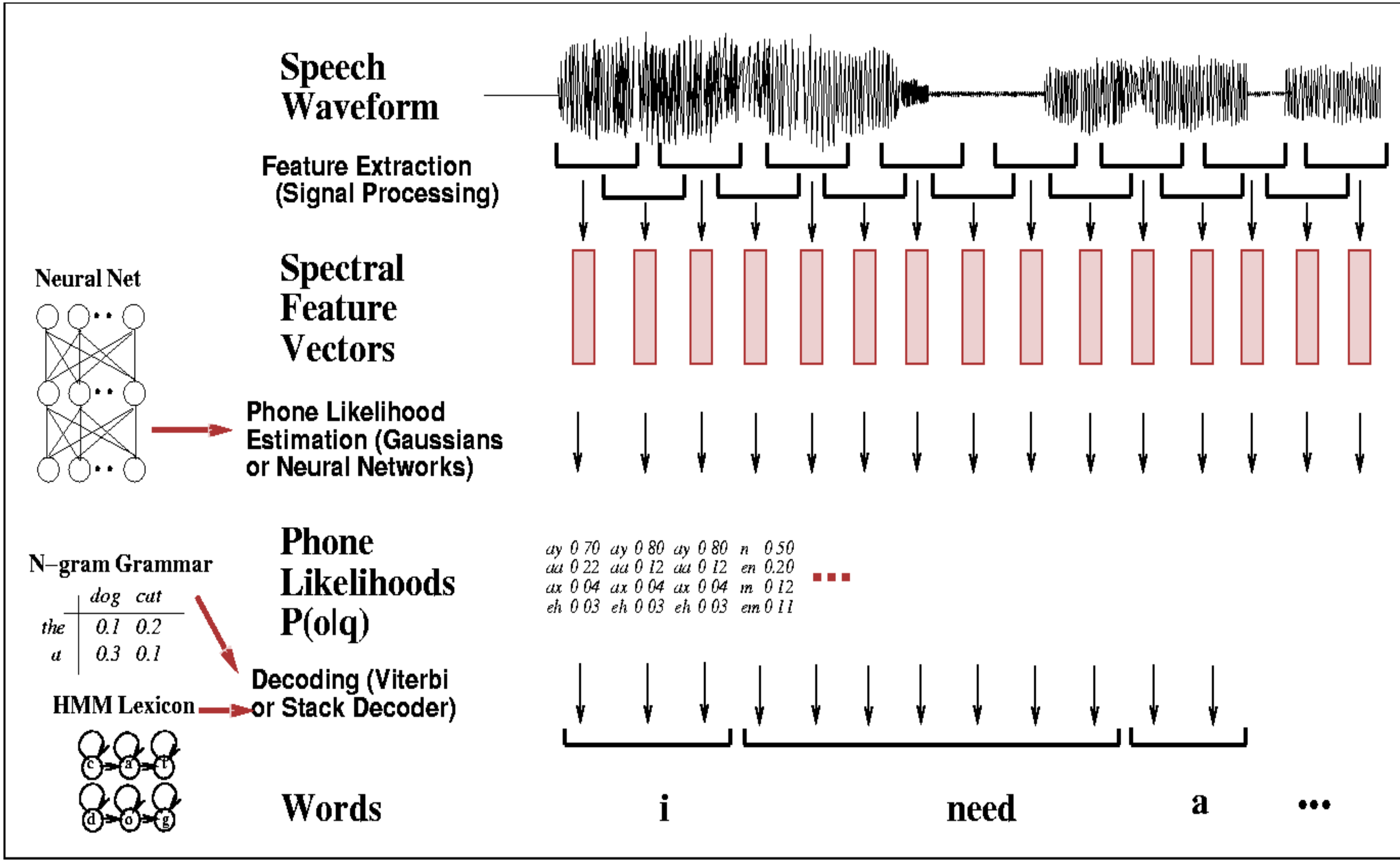
- Donada l'estructura de Markov, l'actualització de paràmetres EM usa els estadistics locals usant l'algorisme *forward-backward (E-step)*
- El pas de maximització (*M-step)* estima la distribució de les observacions via una mitjana 'pesada' (weighted average):

$$\hat{\theta}_i^{(s)} = \frac{\sum_t \hat{p}^{(s)}(x_t = i) y_t}{\sum_t \hat{p}^{(s)}(x_t = i)}$$

- La matriu de transició es calcula de la mateixa manera…

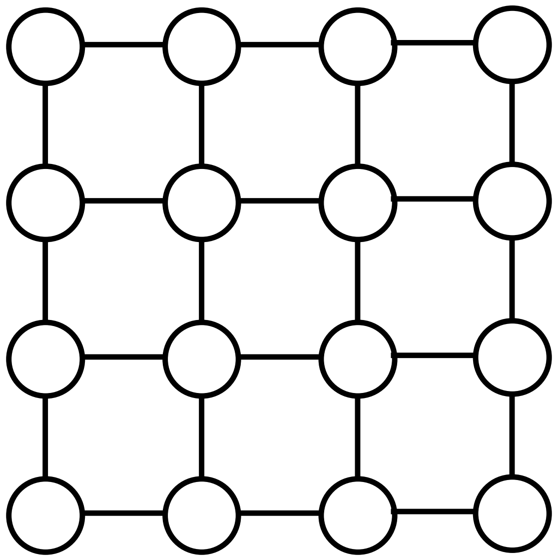
Podem trobar mínims locals, la inicialització és fonamental.

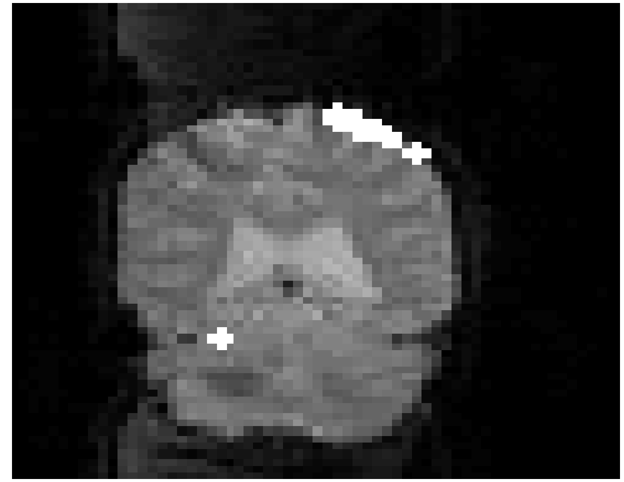
Schematic Architecture for a (simplified) Speech Recognizer



Markov Random Fields en Visió

Idea: pixels veins són similars.





fMRI Analysis *(Kim et. al. 2000)*

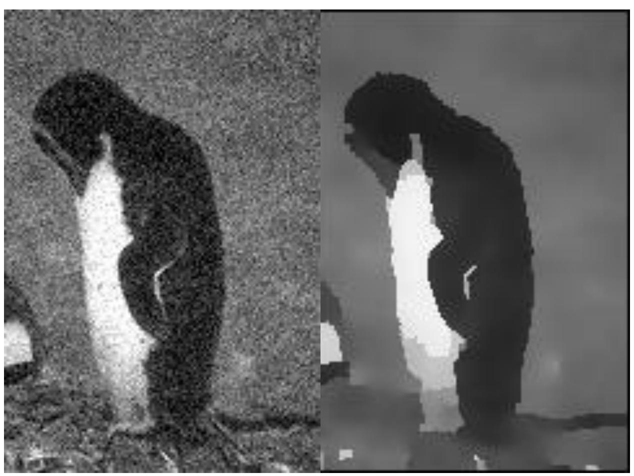
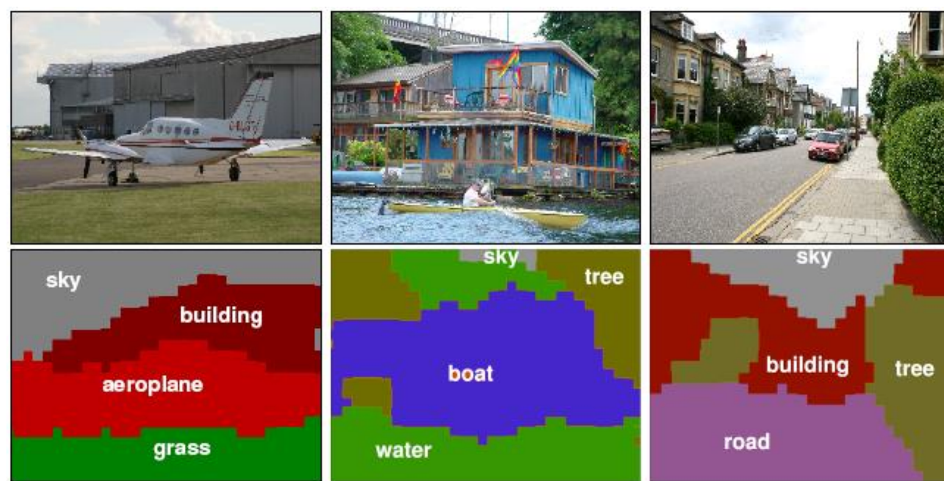


Image Denoising

(Felzenszwalb & Huttenlocher 2004)



Segmentation & Object Recognition

(Verbeek & Triggs 2007)