SEMINARIO DEL 24/04

Ruggieri Andrea

Stranieri Francesco

MAD Lab

IMPLEMENTAZIONE EM SOFT

Correzione errore inferenza esatta

Versione finale

EVOLUZIONE DEL PROGETTO

Ultimo meeting

Algoritmo EM funziona correttamente sull'**esempio** presentato

Algoritmo funziona su BN dove **ogni nodo è binario** e **ogni nodo ha al più un genitore**

Algoritmo funziona bene su **strutture BN complesse**. **Tutti i nodi sono binari**

Algoritmo funziona bene su **strutture BN complesse** con **dati discreti**

Ok test

Tuttavia, durante l'esecuzione di un test, un bug è stato riscontrato all'interno della fase di Expectation. All'interno del dato erano presenti tante variabili nascoste causando una computazione errata dell'inferenza esatta. Il bug è stato identificato e corretto attraverso la definizione di una nuova matrice. La logica è stata riscritta in modo da essere più chiara

Risolto

L'errore era stato sollevato all'interno del test 3

INFERENZA ESATTA

$$P(X \mid e) = \alpha\, P(X,e) = \alpha \sum_{y} P(X,e,y)$$

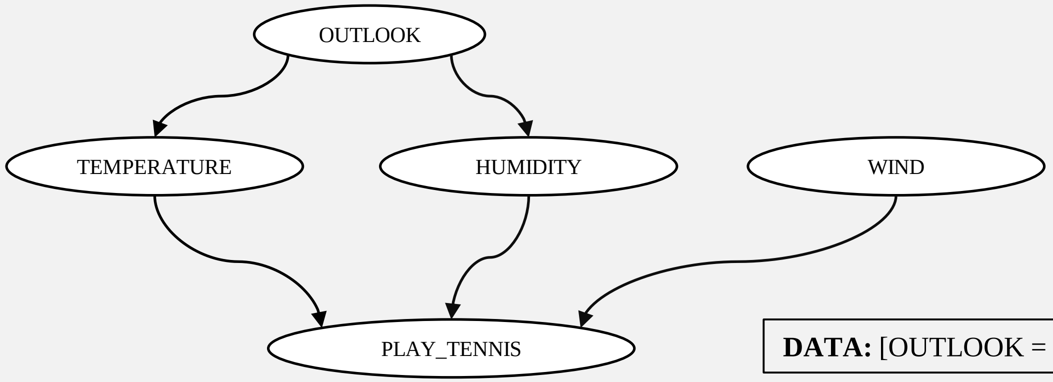
Somma su tutti i possibili valori y delle variabili nascoste

Per poter computare queste sommatorie (nel caso in cui il numero di variabili nascoste sia maggiore di 1), è stato definito un metodo chiamato `get_not_observed_node`

Tale metodo restituisce una matrice con tutte le combinazioni tra i missing values in modo da implementare tutte le possibili sommatorie, prima calcolate erroneamente

N.B:. La modifica è stata effettuata nel passo di Expectation durante la computazione delle probabilità a posteriori

IDEA: MATRIX_NODE_NOT_OBSERVED



OUTLOOK = {0,1,2}
TEMPERATURE = {0,1,2}
HUMIDITY = {0,1,2}
WIND = {0,1}
PLAY_TENNIS = {0,1}

DATA: [OUTLOOK = **NA**, TEMPERATURE = **NA**, HUMIDITY = 2, WIND = 1, PLAY_TENNIS = **NA**]

GOAL: Calcolare P(TEMPERATURE | e)

$$P(T|H,W) = \alpha P(T,H,W) = \alpha \sum_{O} \sum_{P} P(T,H,W,o,p)$$

Matrix_node_not_observed	
OUTLOOK	PLAY TENNIS
0	0
0	1
1	0
1	1
2	0
2	1

EM SOFT VERSIONE FINALE - RIASSUNTO

Inizializzazione uniforme della rete

Expectation step: Computazione delle probabilità a posteriori dei dati mancanti. Si tratta dell'applicazione di **inferenza esatta per enumerazione**

$$P(X|e) = \alpha P(X,e) = \alpha \sum_{y} P(x,e,y)$$

Maximisation step: Stima dei nuovi parametri sulla base delle probabilità a posteriori calcolati nel passo di Expectation

$$\theta = \frac{1}{n}\sum_{x_i} 1\mathrm{I}_{O} + 1\mathrm{I}_{M}\pi_{x_i^M}$$

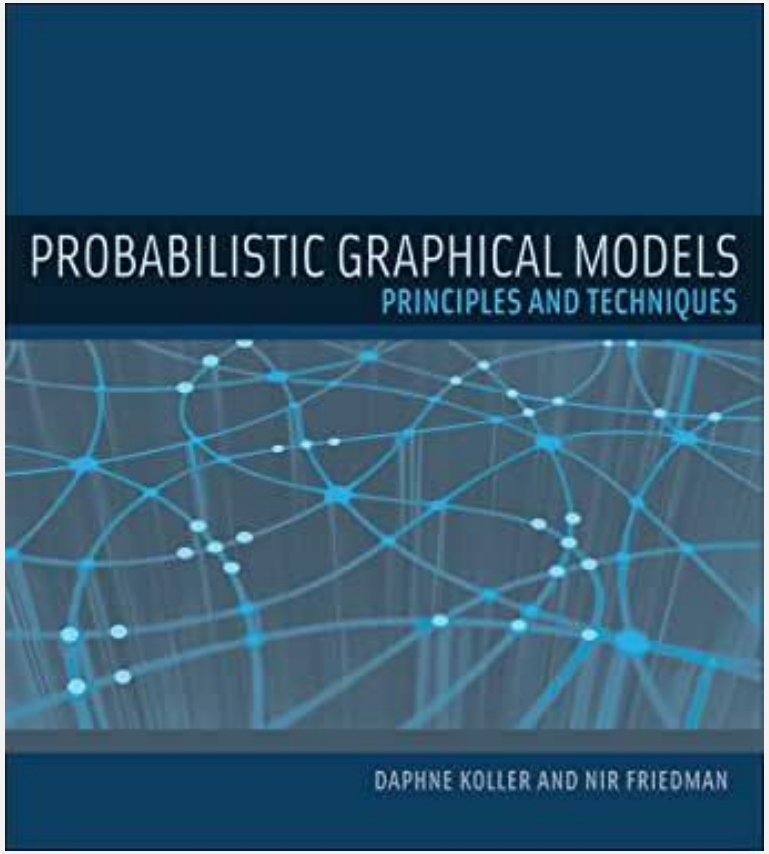
Updating: Imputazione del valore più probabile al dato mancante calcolato sulla base dell'Expectation step

Stopping criteria: Confronto tra θ^{t-1} con θ^{t}. Se l'algoritmo arriva a convergere (differenza in valore assoluto) è minore di un iperparametro α, l'algoritmo si arresta

EM HARD

Introduzione EM
Differenza tra EM HARD e EM SOFT

REFERENZE



A tutorial on the EM algorithm for Bayesian networks: *Serge Romaric Tembo Mouafo, Sandrine Vaton, Jean-Luc Courant, Stephane Gosselin* https://hal.archives-ouvertes.fr/hal-01394337/document

The Bayesian Structural EM Algorithm: *Nir Friedman*

Learning Bayesian Networks: http://pages.cs.wisc.edu/~dpage/cs760/BNall.pdf

NOTAZIONI:

θ	Parametri (CPT) della rete bayesiana
$< x^0, y^0 >$	Istanza di un dato
$<?, y^0 >$	Istanza di un dato con un valore mancante
$M[x^0, y^0]$	Conteggio delle volte in cui su presenta quell'istanza
$H[m]$	variabili che hanno valori mancanti nella istanza di dato o[m]
$\xi[m]$	m-esimo esempio di training
U	Insieme dei parenti per il nodo X
I	Variabile di indicatore che può assumere valore 1 o 0

PROBLEMA MISSINGNESS

Precedentemente abbiamo già discusso riguardo ai problemi della missingness

- Quando si impara da dati completi, trovare **statistiche sufficienti** è facile. Tuttavia, quando abbiamo a che fare con **dati mancanti**, non abbiamo accesso a statistiche complete.
- Ci sono diversi modi per rimpiazzare i valori mancanti e i dati mancanti stessi si possono distinguere in **3 categorie**.
- I metodi di imputazione classici hanno il problema che i dati rimpiazzati sono **condizionatamente indipendenti** dai valori delle altre variabili e i metodi di imputazione **non permettono di apprendere dipendenze** tra variabili nascoste e tutte le altre variabili
- Quando noi apprendiamo con dei dati mancanti, stiamo provando a risolvere due problemi in una volta sola: **Apprendere i parametri** θ e **ipotizzare i valori** per la variabile non osservata. Purtroppo, risolvere questo problema non è semplice.

ALGORITMO EM:

- L'algoritmo EM soft inizia con una **configurazione iniziale** dei parametri θ^0

- **Expectation step**: l'algoritmo usa i parametri correnti θ^t per computare le **expected sufficient statistics**
  - Per ogni dato ***o[m]*** e per ogni famiglia X, U, computa tutte le probabilità marginali:
$$Q(X,U) = P(X,U|o,\theta)$$
  - Calcolo delle expected sufficient statistics per ogni x, u come:
$$\bar{M}_\theta[u] = \sum_{m=1}^{M} \sum_{h[m]\epsilon Val(H[m])} Q(h[m])\mathrm{I}\{\xi[m] < Y >= y\}$$

- **Maximisation step:** Tratta le expected sufficient statistics come osservate ed esegue la **maximum likelihood estimation**
$$\theta_{x|u}^{t+1} = \frac{\bar{M}_{\theta^t}\ [x,u]}{\bar{M}_{\theta^t}\ [u]}$$

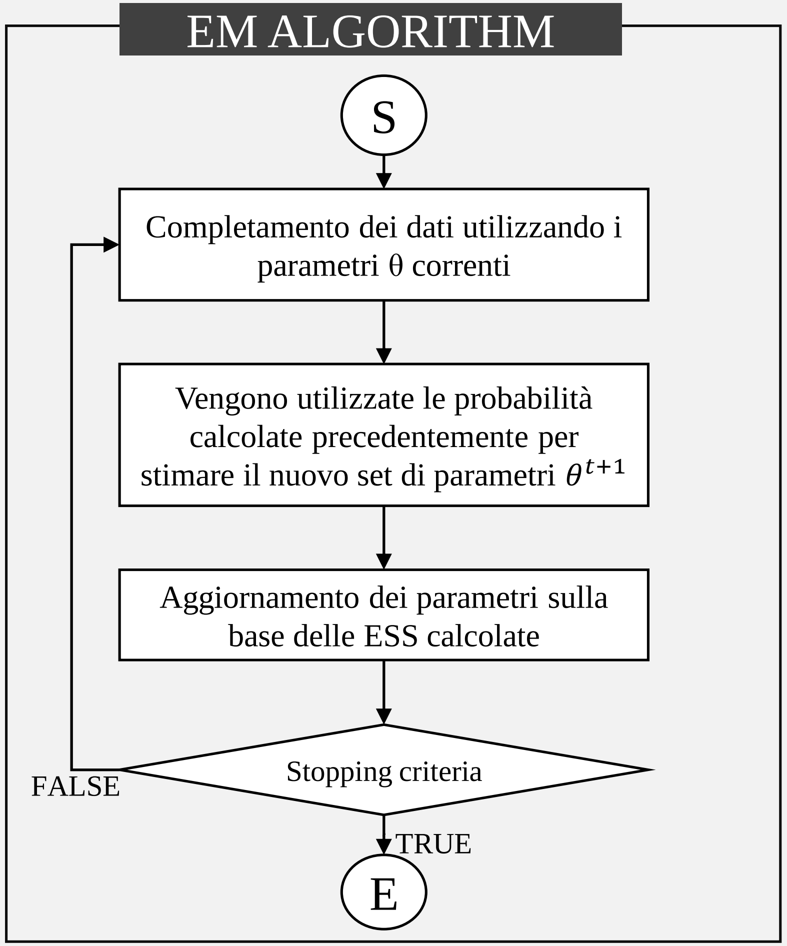
Lo step di maximisation è lineare. Tuttavia, lo step più difficile è quello di Expectation

ALGORITMO EM

Algorithm 19.2 Expectation-maximization algorithm for BN with table-CPDs

Procedure Compute-ESS (
$\mathcal{G}$, // Bayesian network structure over $X_1, \ldots, X_n$
$\boldsymbol{\theta}$, // Set of parameters for $\mathcal{G}$
$\mathcal{D}$ // Partially observed data set
)
// Initialize data structures
for each $i = 1, \ldots, n$
for each $x_i, \boldsymbol{u}_i \in Val(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})$
$\bar{M}[x_i, \boldsymbol{u}_i] \leftarrow 0$
// Collect probabilities from all instances
for each $m = 1 \ldots M$
Run inference on $\langle \mathcal{G}, \boldsymbol{\theta} \rangle$ using evidence $\boldsymbol{o}[m]$
for each $i = 1, \ldots, n$
for each $x_i, \boldsymbol{u}_i \in Val(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})$
$\bar{M}[x_i, \boldsymbol{u}_i] \leftarrow \bar{M}[x_i, \boldsymbol{u}_i] + P(x_i, \boldsymbol{u}_i \mid \boldsymbol{o}[m])$
return $\{\bar{M}[x_i, \boldsymbol{u}_i] : \forall i = 1, \ldots, n, \forall x_i, \boldsymbol{u}_i \in Val(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})\}$

Procedure Expectation-Maximization (
$\mathcal{G}$, // Bayesian network structure over $X_1, \ldots, X_n$
$\boldsymbol{\theta}^0$, // Initial set of parameters for $\mathcal{G}$
$\mathcal{D}$ // Partially observed data set
)
for each $t = 0, 1 \ldots,$ until convergence
// E-step
$\{\bar{M}_t[x_i, \boldsymbol{u}_i]\} \leftarrow$ Compute-ESS($\mathcal{G}, \boldsymbol{\theta}^t, \mathcal{D}$)
// M-step
for each $i = 1, \ldots, n$
for each $x_i, \boldsymbol{u}_i \in Val(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})$
$\theta^{t+1}_{x_i \mid \boldsymbol{u}_i} \leftarrow \frac{\bar{M}_t[x_i, \boldsymbol{u}_i]}{\bar{M}_t[\boldsymbol{u}_i]}$
return $\boldsymbol{\theta}^t$



ALGORITMO EM

L’idea appena presentata sta alla base dell’algoritmo che noi definiamo EM SOFT

Nella seconda parte della presentazione, si illustrerà un esempio dove si marcherà il concetto che la versione EM SOFT implementata e la versione EM STANDARD presentata da *Friedman* sono equivalenti

EM HARD VS EM SOFT

- Entrambi i metodi eseguono due step: Completare i dati utilizzando i parametri θ^t e usare questi dati per computare i nuovi parametri θ^{t+1}

- Tuttavia, diversamente da EM SOFT, EM HARD seleziona, per ogni istanza o[m] il singolo assegnamento h[m] che massimizza $P(h|o[m], \theta^t)$ (N.B. Friedman nel suo libro usa appunto il termine **hard-assignment EM**)

- EM HARD può essere visto come l'ottimizzazione di una diversa funzione obiettivo che coinvolge sia l’apprendimento di parametri sia il compito di apprendere il corretto assegnamento delle variabili mancanti. **L’obbiettivo è quello di massimizzare la likelihood dei dati completi dati i parametri**

$$\max_{\theta, H} l(\theta : H, D)$$

- EM SOFT al contrario, tenta di massimizzare $l(\theta : D)$, **considerando tutti i possibili assegnamenti dei dati mancanti**

EM HARD VS EM SOFT

EM HARD e EM SOFT tendono ad essere simili se $P(H|D,\theta)$ assegna una probabilità più alta a una delle possibili combinazioni dei dati mancanti durante l'**E-step**

In questo caso EM HARD si comporta in modo efficace e i dati rimpiazzati risultano essere uguali a EM SOFT

Al contrario, i due algoritmi potrebbero condurre a risultati veramente diversi

EM HARD VS EM SOFT

Seppur **in contesti diversi**, questa osservazione, era stata sollevata anche nel confronto tra EM SOFT e EM con BNLEARN che utilizza metodi di inferenza approssimata. Tale conclusione è stata raggiunta durante l'esposizione dei test effettuati

EM HARD VS EM SOFT

	EM SOFT	EM HARD
Inizializzazione parametri	Random, uniforme o qualsiasi altro modo	Random, uniforme o qualsiasi altro modo
Expectation	Si prendono in considerazione tutte le possibili combinazioni dei dati	Sceglie un singolo assegnamento che massimizza la joint distribution *(max)*
Expected sufficient statistics	$\bar{M}_{\theta}[u] = \sum_{m=1}^{M} \sum_{h[m] \epsilon Val(H[m])} Q(h[m]) \mathrm{I}\{\xi[m] < Y >= y\}$	$\bar{M}_{\theta^t}[x^1] = \sum_{m}^{M} \boldsymbol{I}\{\xi[m] < X >= x^1\}$
Maximisation step	Sulla base delle expected sufficient statistics si aggiornano i parametri	Sulla base delle expected sufficient statistics si aggiornano i parametri

EM SOFT E HARD
Esempio pratico

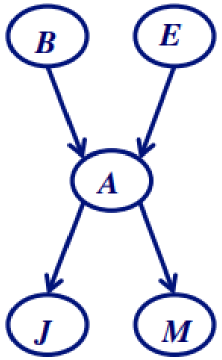
EM SOFT e EM HARD a confronto

ESEMPIO

Partiamo dall’esempio classico **Intrusione - Terremoto**

Si consideri la seguente situazione. Si è installata nella propria abitazione un sistema antifurto abbastanza affidabile ma occasionalmente risponde anche ai piccoli terremoti. Ci sono due vicini di casa John e Mary che hanno promesso di telefonare sul posto di lavoro dopo aver sentir suonare l’allarme del sistema antifurto. Inoltre si sa anche:

• John chiama sempre quando sente l’allarme suonare, ma in alcuni casi confonde lo squillo del telefono col suono dell’allarme;
• Mary ama ascoltare musica ad alto volume e occasionalmente non sente che l’allarme suona.



P(B)
0.1

P(E)
0.2

B	E	P(A)
t	t	0.9
t	f	0.6
f	t	0.3
f	f	0.2

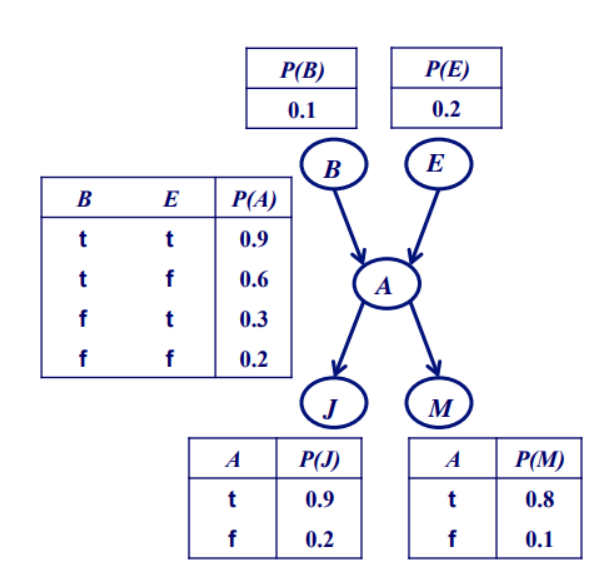
A	P(J)
t	0.9
f	0.2

A	P(M)
t	0.8
f	0.1

ESEMPIO

Sia fornito il seguente dataset

	D1	D2	D3
B	?	F	F
E	F	?	T
A	?	?	T
J	F	T	?
M	F	F	F



ESEMPIO

EM SOFT (libro Friedman) - EXPECTATION

$Q^1(< B^T, A^T >) = \alpha(0{,}1 \ * 0{,}8 * 0{,}6 * 0{,}1 * 0{,}2) = \alpha(9{,}6 \cdot 10^{-4}) = 2{,}17 \cdot 10^{-3}$

$Q^1(< B^T, A^F >) = \alpha(0{,}1 \ * 0{,}8 * 0{,}4 * 0{,}8 * 0{,}9) = \alpha(0{,}02304) \ = 0{,}05217$

$Q^1(< B^F, A^T >) = \alpha(0{,}9 \ * 0{,}8 * 0{,}2 * 0{,}1 * 0{,}2) = \alpha(2{,}88 \cdot 10^{-3}) = \ \ 6{,}52 * 10^{-3}$

$Q^1(< B^F, A^F >) = \alpha(0{,}9 \ * 0{,}8 * 0{,}8 * 0{,}8 * 0{,}9) = \alpha(0{,}4147) = 0{,}939$

Quattro possibili casi di completamento dei dati
$\alpha = 0{,}4416$

$Q^2(< E^T, A^T >) = \alpha(0{,}9 * 0{,}2 \ * 0{,}3 \ * 0{,}9 * 0{,}2) = \ \alpha(9{,}72 \cdot 10^{-3}\ ) = 0{,}006$

$Q^2(< E^T, A^F >) = \alpha(0{,}9 \ * 0{,}2 * 0{,}7 * 0{,}2 * 0{,}9) = \alpha(0{,}02268) = 0{,}14$

$Q^2(< E^F, A^T >) = \alpha(0{,}9 \ * 0{,}8 * 0{,}2 * 0{,}9 * 0{,}2) = \alpha(0{,}02592) = 0{,}16$

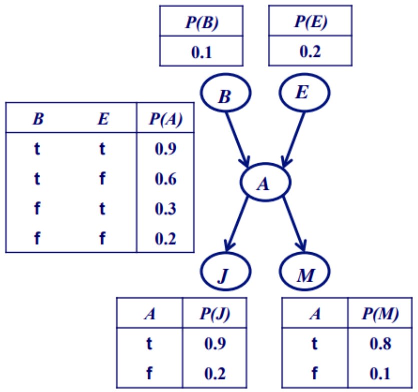
$Q^2(< E^F, A^F >) = \alpha(0{,}9 \ * 0{,}8 * 0{,}8 * 0{,}2 * 0{,}9) = \alpha(0{,}10368) = 0{,}64$

Quattro possibili casi di completamento dei dati
$\alpha = 0{,}162$

$Q^3(< J^T >) = \alpha(0{,}9 * 0{,}2 \ * 0{,}3 \ * 0{,}9 * 0{,}2) = \alpha(9{,}72 \cdot 10^{-3}) = 0{,}9$

$Q^3(< J^F >) = \alpha(0{,}9 \ * 0{,}2 * 0{,}3 * 0{,}1 * 0{,}2) = \alpha(1{,}08 \cdot 10^{-3}) = 0{,}1$

Due possibili casi di completamento dei dati
$\alpha = 0{,}0108$



	D1	D2	D3
B	?	F	F
E	F	?	T
A	?	?	T
J	F	T	?
M	F	F	F

ESEMPIO

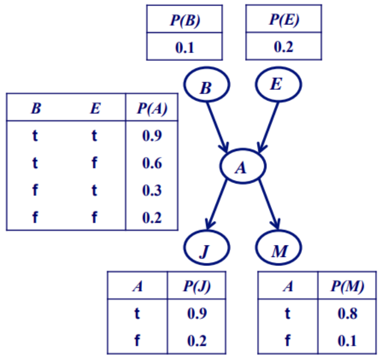
A questo punto possiamo stimare le **expected sufficient statistics** come:

$$\bar{M}_{\theta}[y] = \sum_{m=1}^{M} \sum_{h[m] \epsilon Val(H[m])} Q(h[m]) \mathrm{I}\{\xi[m] < Y >= y\}$$

Nel nostro caso usando le distribuzioni $Q^1, Q^2 \ e \ Q^3$ calcolate precedentemente abbiamo che:

$\bar{M}_{\theta}[J^F, A^F] = \ Q^1(< B^T, A^F >) + Q^1(< B^F, A^F >)$ + 0 + 0 = 0,05217 +0,939= 0,99117

$\bar{M}_{\theta}[A^F] = \ Q^1(< B^T, A^F >) + Q^1(< B^F, A^F >)+$
$Q^2(< E^T, A^F >) + Q^2(< E^F, A^F >)$ +
0 = 0,05217 +0,939+ 0,14+ 0,64= 1,77



	D1	D2	D3
B	?	F	F
E	F	?	T
A	?	?	T
J	F	T	?
M	F	F	F

ESEMPIO

Riprendiamo ora lo script **EM SOFT** e calcoliamo il passo di **Expectation**:

$$P(X|e) = \alpha P(X,e) = \alpha \sum_{y} P(x,e,y)$$

Dato 1

$P^1(A) = P(A|E = F, J = F, M = F) = \alpha \sum_B P(A,E,J,M,B) = < 0.0087; 0.991>$

$P^1(B) = P(B|E = F, J = F, M = F) = \alpha \sum_A P(B,E,J,M,A) = < 0.0544; 0.946>$

Dato 2

$P^2(A) = P(A|B = F, J = T, M = F) = \alpha \sum_E P(A,M,J,B,E) = < 0.22; 0.78>$

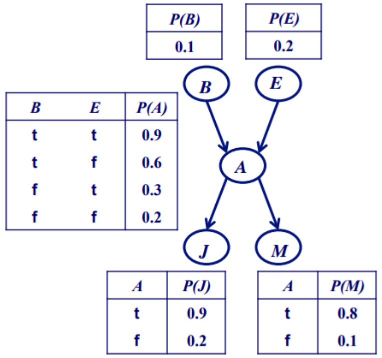
$P^2(E) = P(E|B = F, J = T, M = F) = \alpha \sum_A P(B,E,J,M,A) = < 0.2; 0.8>$

Dato 3

$P^3(J) = P(J|B = F, E = T, A = T, M = F) = < 0.9; 0.1>$

	Node	Value	Data_1	Data_2	Data_3
1	A	0	0.00869565217391304	0.22	0
2	A	1	0.991304347826087	0.78	0
3	B	0	0.0543478260869565	0	0
4	B	1	0.945652173913043	0	0
5	E	0	0	0.2	0
6	E	1	0	0.8	0
7	J	0	0	0	0.9
8	J	1	0	0	0.1
9	M	0	0	0	0
10	M	1	0	0	0

Convenzione: 0=TRUE, 1=FALSE



	D1	D2	D3
B	?	F	F
E	F	?	T
A	?	?	T
J	F	T	?
M	F	F	F

Dove $P^1(A)$ è soltanto una notazione che indica la distribuzione di probabilità della variabile A per il dato 1

ESEMPIO

Prendiamo in considerazione i due approcci E-step presentati precedentemente e consideriamo il dato 2:

$Q^2(< E^T, A^T >) = \alpha(0{,}9 * 0{,}2 * 0{,}3 * 0{,}9 * 0{,}2) = \alpha(9{,}72 \cdot 10^{-3}) = 0{,}06$

$Q^2(< E^T, A^F >) = \alpha(0{,}9 * 0{,}2 * 0{,}7 * 0{,}2 * 0{,}9) = \alpha(0{,}02268) = 0{,}14$

$Q^2(< E^F, A^T >) = \alpha(0{,}9 * 0{,}8 * 0{,}2 * 0{,}9 * 0{,}2) = \alpha(0{,}02592) = 0{,}16$

$Q^2(< E^F, A^F >) = \alpha(0{,}9 * 0{,}8 * 0{,}8 * 0{,}2 * 0{,}9) = \alpha(0{,}10368) = 0{,}64$

$P^2(E) = P(E \mid B{=}F, J{=}T, M{=}F) = \alpha \sum_A P(B,E,J,M,A) =$ <0.2;0.8>

	D2
B	F
E	?
A	?
J	T
M	F

Nota bene:

$$P(E|B = F, J = T, M = F) = \alpha \sum_A P(B,E,J,M,A) = \; < Q^2(< E^T, A^T >) + Q^2(< E^T, A^F >); \; Q^2(< E^F, A^T >) + Q^2(< E^F, A^F >) >$$

=< 0.2;0.8>

Allora:

$$\boldsymbol{Q(E^T) = P(E^T | o[m], \theta) = P(E^T | o[2], \theta) = Q^2(< E^T, A^T >) + Q^2(< E^T, A^F >)}$$

ESEMPIO

Considerazione:

Si consideri la formula presentata nel libro di Friedman e nelle slide precedenti per computare le **ESS**:

$$\bar{M}_{\theta}[y] = \sum_{m=1}^{M} \sum_{h[m] \epsilon Val(H[m])} Q(h[m]) \mathrm{I}\{\xi[m] < Y > = y\}$$

Collegando questa formula con il procedimento illustrato nella slide 24, è possibile notare che la formula sopra esposta può essere semplificata ulteriormente e risulta essere uguale a:

$$\boldsymbol{\bar{M}_{\theta^t}[x, u] = \sum_{m} P(x, u | o[m], \theta^t)}$$

ESEMPIO

Considerazione:

Si consideri la formula presentata nel libro di Friedman e nelle slide precedenti per computare le **ESS**:

$$\bar{M}_{\theta}[y] = \sum_{m=1}^{M} \sum_{h[m] \epsilon Val(H[m])} Q(h[m]) \mathrm{I}\{\xi[m] < Y >= y\}$$

Collegando questa formula con il procedimento illustrato nella slide 24, è possibile notare che la formula sopra esposta può essere semplificata ulteriormente e risulta essere uguale a:

$$\boldsymbol{\bar{M}_{\theta^t}[x, u]} = \sum_{\boldsymbol{m}} \boldsymbol{P(x, u | o[m], \theta^t)}$$

Questa rappresentazione facilita la comprensione dell'argomento e permette di ridurre i costi computazionali. Tuttavia, la formula $\boldsymbol{\bar{M}_{\theta^t}[x, u] = \sum_m P(x, u | o[m], \theta^t)}$ risulta ancora essere NP-HARD perché per ogni dato bisogna considerare sempre tutte le combinazioni di dati mancanti

ESEMPIO

Consideriamo ora **EM HARD,** lo scopo è quello di massimizzare la likelihood dei dati completi

$$\max_{\boldsymbol{\theta},\boldsymbol{H}} \boldsymbol{l}(\boldsymbol{\theta}: \boldsymbol{H}, \boldsymbol{D})$$

Abbiamo visto precedentemente che esistono:

- 4 combinazioni di assegnamento di valori alle variabili nascoste per completare il dato 1
- 4 combinazioni di assegnamento di valori alle variabili nascoste per completare il dato 2
- 2 combinazioni di assegnamento di valori alle variabili nascoste per completare il dato 3

Tuttavia, il nostro scopo è di **selezionare il singolo assegnamento h[m]** che massimizza $\boldsymbol{P}(\boldsymbol{h}|\boldsymbol{o}[\boldsymbol{m}], \boldsymbol{\theta}^{\boldsymbol{t}})$

Intuitivamente:

$Dato1$: $Q^1(< B^F, A^F >) = 0,9\ * 0,8 * 0,8 * 0,8 * 0,9 = 0,4147$

$Dato2$: $Q^2(< E^F, A^T >) = 0,9\ * 0,8 * 0,2 * 0,9 * 0,2 = 0,02592$

$Dato3$: $Q^3(< J^T >) = 0,9 * 0,2\ * 0,3\ * 0,9 * 0,2 = 9,72 \cdot 10^{-3}$

ESEMPIO

Una volta ottenuto l'assegnamento che massimizza la probabilità allora, è possibile calcolare le expected sufficient statistics prendendo in considerazione la seguente tabella

	D1	**D2**	**D3**
B	**F**	F	F
E	F	**F**	T
A	**F**	**T**	T
J	F	T	**T**
M	F	F	F

Possiamo quindi calcolare le expected sufficient statistics facilmente attraverso la seguente funzione:

$$\bar{M}_{\theta^t}[x^1] = \sum_{m}^{M} \boldsymbol{I}\{\xi[m] < X > = x^1\}$$

ESEMPIO

$$\bar{M}_{\theta^t}[x^1] = \sum_{m}^{M} \boldsymbol{I}\{\xi[m] < X >= x^1\}$$

$\bar{M}_\theta[J^F, A^F] = 1 + 0 + 0 = 1$
$\bar{M}_\theta[J^T, A^F] = 0 + 0 + 0 = 0$
$\bar{M}_\theta[A^F] = 1 + 0 + 0 = 1$

$\bar{M}_\theta[J^F, A^T] = 0 + 0 + 0 = 0$
$\bar{M}_\theta[J^T, A^T] = 0 + 1 + 1 = 2$
$\bar{M}_\theta[A^T] = 0 + 1 + 1 = 2$

Di conseguenza:

$$\tilde{\boldsymbol{\theta}}_{J^F|A^F} = \frac{\bar{M}_\theta[J^F, A^F]}{\bar{M}_\theta[A^F]} = \frac{1}{1} = 1$$

$$\tilde{\boldsymbol{\theta}}_{J^T|A^F} = \frac{\bar{M}_\theta[J^T, A^F]}{\bar{M}_\theta[A^F]} = \frac{0}{1} = 0$$

$$\tilde{\boldsymbol{\theta}}_{J^F|A^T} = \frac{\bar{M}_\theta[J^F, A^F]}{\bar{M}_\theta[A^F]} = \frac{0}{2} = 0$$

$$\tilde{\boldsymbol{\theta}}_{J^T|A^T} = \frac{\bar{M}_\theta[J^F, A^F]}{\bar{M}_\theta[A^F]} = \frac{2}{2} = 1$$

Risultati ottenuti a causa dei pochi dati a disposizione

	D1	D2	D3
B	F	F	F
E	F	F	T
A	F	T	T
J	F	T	T
M	F	F	F

ESEMPIO

Considerazioni:

EM HARD, esattamente come EM SOFT fa ricorso al concetto di **probabilità a posteriori** per computare il passo di expectation. Tuttavia, in EM HARD, esiste una differenza sostanziale. Rispetto a EM SOFT, seleziona per ogni istanza o[m] il singolo assegnamento alle variabili nascoste h[m] che massimizza $P(h|o[m],\theta)$