Modeles Probabilistes Graphiques Intelligents

Principe général

Les modèles probabiliste graphiques sont l'évolution naturelle des systèmes experts des années 1980 ou l'on a cherché a donner une base solide et complètement mathématique c'est-à-dire probabiliste aux mécanisme d'inférence et aux coefficients de certitude qui y étaient attachés.

Le problème est que les mécanismes d'inférence reposent maintenant sur des optimisations de vraisemblance qui sont très consommateur de temps de calcul quand ils ne sont pas infaisables.

On est donc forcé de réaliser ces optimisations sur un petit nombre de données et avec des petits réseaux bayésiens.

Il apparait alors naturel d'essayer de concevoir des systèmes qui partent de solutions existantes et tentent d'améliorer ces solutions sans revenir a zéro à chaque fois qu'un nouveau réseau est essayé.

L'idée est même de rechercher de manière automatique des modifications incrémentale de la structure afin d'améliorer l'efficience du système sans remettre en question l'acquis.

Parmi ces évolutions de structure figure en bonne place l'enrichissement des informations prises en compte par le système.

Algorithme EM

La partie E de l'algorithme EM est censée s'occuper de l'apprentissage de la chaine de variables cachée par calcul d'Esperance des transitions. Ceci provient de ce que si on maximise le loglikelihood en supposant connu les probabilités d'émissions (donc les paramètres associés), on est réduit a maximiser des quantités comme $\sum_{i=1}^{n} a_i \, Log(\delta_i)$ ceci va en effet nous donner que les paramètres optimaux de la chaines sont obtenu par calcul d'espérance de transition. D'où le nom donné a cette partie E.

Le mechanisme de base

Théorème 1

On veut maximiser $\varphi\left(\delta_i|_{i=1,n}\right) = \sum_{i=1}^{n} a_i \, Log(\delta_i)$ sous la contrainte :

$$\sum_{i=1}^{n} \delta_i = E$$

Alors le résultat est :

On introduit les multiplicateurs de Lagrange : soit a maximiser $\sum_{i=1}^{n} a_i \, Log(\delta_i) - \lambda(\sum_{i=1}^{n} \delta_i - E)$

On dérive : $\frac{a_i}{\delta_i} - \lambda = 0$ pour tous les i

En réinjectant dans la contrainte on trouve : $\delta_i = \frac{a_i \, E}{\sum_{i=1}^{n} a_i}$

Qui est bien E fois une probabilité normalisée

Rapprochement avec une transformée de Legendre

1) Nous allons partir de l'equilibre

La contrainte est équivalente a :

Soit $E = \sum_{i=1}^{n} \delta_i$

On transforme de Legendre φ par rapport a la variable E :

On doit donc minimiser $E\lambda - \varphi\left(\delta_i|_{i=1,n}\right)$

Le minimum est donné par $\lambda = \frac{\partial\varphi(\delta_i|_{i=1,n})}{\partial E}$ ce qui a l'optimum définit la transformation inverse

$$\varphi(E) = \sum_{i=1}^{n} a_i \, Log(a_i \, E) - \sum_{i=1}^{n} a_i \, Log\left(\sum_{i=1}^{n} a_i\right)$$

Soit

$$\varphi(E) = \left(\sum_{i=1}^{n} a_i\right) Log(\, E) + \sum_{i=1}^{n} a_i \, Log(a_i) - \left(\sum_{i=1}^{n} a_i\right) Log\left(\sum_{i=1}^{n} a_i\right)$$

Donc $\lambda = \frac{\sum_{i=1}^{n} a_i}{E}$

Donc l'énergie libre s'écrit :

$$F(\lambda) = \sum_{i=1}^{n} a_i - \varphi\left(\frac{\sum_{i=1}^{n} a_i}{\lambda}\right)$$

Qui se simplifie en $F(\lambda) = (\sum_{i=1}^{n} a_i) Log\lambda + \sum_{i=1}^{n} a_i \left(1 - Log(a_i)\right)$

Qui vaut bien $E\lambda - \varphi = \sum_{i=1}^{n} a_i \left(1 - Log(a_i)\right)$ quand $\lambda = 1$

2) Nous considerons un cas plus general,

Si on ecrit les variables $\delta_i\ par$:

$\delta_i = {}^{E}/_{n} + \omega_i$ pour $i = 2, n$ et $\delta_1 = {}^{E}/_{n} - \sum_{i=2}^{n} \omega_i$

On est alors capable d'ecrire

$$\varphi(E, \omega_i|_{i=2,n}) = a_1 Log\left({}^{E}/_{n} - \sum_{i=2}^{n} \omega_i\right) + \sum_{i=2}^{n} a_i\, Log({}^{E}/_{n} + \omega_i)$$

On doit donc minimiser $E\lambda - \ \varphi(E, \omega_i|_{i=2,n})$

Le minimum est donné par $\lambda(E, \omega_i|_{i=2,n}) = \ \frac{\partial \varphi(\delta_i|_{i=1,n})}{\partial E} = \frac{a_1}{E - n\sum_{i=2}^{n} \omega_i} + \sum_{i=2}^{n} \frac{a_i}{E + n\ \omega_i}$

Ce qui définit implicitement $E(\lambda, \omega_i|_{i=2,n})$

L'énergie libre s'écrit

$$F(\lambda, \omega_i|_{i=2,n}) = E(\lambda, \omega_i|_{i=2,n})\lambda - \ \varphi(E(\lambda, \omega_i|_{i=2,n}), \omega_i|_{i=2,n})$$

Quand on minimise cet énergie libre, par rapport a $\omega_i|_{i=2,n}$ on obtient :

$$\frac{\partial F(\lambda)}{\partial \omega_i} = \frac{\partial E(\lambda, \omega_i|_{i=2,n})}{\partial \omega_i}\left(\lambda - \ \frac{\partial \varphi(E(\lambda, \omega_i|_{i=2,n}), \omega_i|_{i=2,n})}{\partial E}\right) - \frac{\partial \varphi(E(\lambda, \omega_i|_{i=2,n}), \omega_i|_{i=2,n})}{\partial \omega_i}$$

Que l'on peut ré exprimer en :

$$\frac{\partial F(\lambda)}{\partial \omega_i} = \frac{\partial E(\lambda, \omega_i|_{i=2,n})}{\partial \omega_i}\left(\lambda - \ \frac{a_1}{E - n\sum_{j=2}^{n} \omega_j} - \sum_{j=2}^{n} \frac{a_j}{E + n\ \omega_j}\right) + \frac{a_1}{{}^{E}/_{n} - \sum_{j=2}^{n} \omega_j} - \frac{a_i}{{}^{E}/_{n} + \omega_i}$$

A l'équilibre :

$$\frac{\partial F(\lambda)}{\partial \omega_i} = 0$$

On en déduit une valeur pour la dérivée :

$$\frac{\partial E(\lambda, \omega_i|_{i=2,n})}{\partial \omega_i} = \frac{\frac{-a_1}{{}^{E}/_{n} - \sum_{j=2}^{n} \omega_j} + \frac{a_i}{{}^{E}/_{n} + \omega_i}}{\lambda - \ \frac{a_1}{E - n\sum_{j=2}^{n} \omega_j} - \sum_{j=2}^{n} \frac{a_j}{E + n\ \omega_j}}$$

Ces équations alliées avec $\lambda\left(E, \omega_i|_{i=2,n}\right) = \frac{\partial\varphi(\delta_i|_{i=1,n})}{\partial E} = \frac{a_1}{E - n\sum_{i=2}^{n}\omega_i} + \sum_{i=2}^{n}\frac{a_i}{E + n\,\omega_i}$

Déterminent $\omega_i(\lambda)$ a l'équilibre.

A l'équilibre le dénominateur est nul lorsque l'on substitue la valeur de $\lambda\left(E, \omega_i|_{i=2,n}\right)$,

Donc le numérateur doit aussi estre nul.

cela implique que $\frac{-a_1}{E/n - \sum_{j=2}^{n}\omega_j} + \frac{a_i}{E/n + \omega_i} = 0$

soit $\frac{a_1}{E/n - \sum_{j=2}^{n}\omega_j} = \frac{a_i}{E/n + \omega_i}$

Soit encore $\left(\omega_i + E/n\right) = \frac{a_1}{a_i}\left(E/n - \sum_{j=2}^{n}\omega_j\right)$

Soit encore $(\delta_i) = \frac{a_1}{a_i}(\delta_1)$

Soit encore $\delta_i = E/n$

Ce qui est effectivement le cas a l'équilibre,

Théorème 2

On veut maximiser $\sum_{i,j,t} v_{i,j,t} Log\left(\gamma_{i,j}\right)$ sous la contrainte
$\sum_j \gamma_{i,j} = 1$ pour tout i

On écrit le lagrangien $\sum_{i,j,t} v_{i,j,t} Log\left(\gamma_{i,j}\right) - \sum_i \lambda_i\left(\sum_j \gamma_{i,j} - 1\right)$

La stationnarité par rapport a $\gamma_{i,j}$ implique que :

$\sum_t \frac{v_{i,j,t}}{\gamma_{i,j}} - \lambda_i$ pour tout i

Donc on en tire la valeur de $\gamma_{i,j} = \frac{\sum_t v_{i,j,t}}{\lambda_i}$

Remplaçant dans la contrainte , on en tire la valeur de λ_i :

$$\lambda_i = \sum_{j,t} v_{i,j,t}$$

Et donc la valeur finale de $\gamma_{i,j}\ est$:

$$\gamma_{i,j} = \frac{\sum_t v_{i,j,t}}{\sum_{j,t} v_{i,j,t}}$$

Théorème 3

On veut maximiser $\sum_{i,j,t} v_{i,j,t}\alpha(i,j)Log(\gamma_{i,j})$

Ou $\alpha(i,j) = 0\ ou\ 1$

Dans ce cas la même formule conduit a : $\gamma_{i,j} = \frac{\sum_t v_{i,j,t}\alpha(i,j)}{\sum_{j,t} v_{i,j,t}\alpha(i,j)}$

Expression d'un HHMM de correlation dans le cadre d'un HMM :

Dans le cas d'un HMM , Le CDLL est : $Log(\delta_{c_1}\prod_t \gamma_{c_{t-1},c_t}\prod_t p_{c_t}(x_t))$

Qui se réexprime en $Log(\delta_{c_1}) + \sum_t Log(\gamma_{c_{t-1},c_t}) + \sum_t Log\left(p_{c_t}(x_t)\right)$

Et quand on introduit

$u_j(t) = 1$ quand $c_t = j$

Et

$v_{j,k}(t) = 1$ quand $c_{t-1} = j$ et $c_t = k$

Cela donne CDLL=

$$\sum_{j=1}^{m} u_j(1)\, Log(\delta_j) + \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{t=2}^{T} v_{j,k}(t)\, Log(\gamma_{j,k}) + \sum_{j=1}^{m}\sum_{t=1}^{T} u_j(t)\, Log\left(p_j(x_t)\right)$$

Le cas d'un double HMM bi indépendant est bien sûr :

CDLL=

$$Log(\delta_{c_1}) + Log(\delta_{d_1}) + \sum_t \left(Log(\gamma_{c_{t-1},c_t}) + Log(\gamma_{d_{t-1},d_t})\right) + \sum_t \left(Log\left(p_{c_t}(x_t)\right) + Log\left(p_{d_t}(y_t)\right)\right)$$

Si on rajoute une troisième variable cachée représentant la corrélation sans observation :

CDLL=

$$CDLL = \sum_{c_t,d_t,e_t} Log(\delta_{c_1}) + Log(\delta_{d_1}) + Log(\delta_{e_1}) + \sum_t \left(Log(\gamma_{c_{t-1},c_t}) + Log(\gamma_{d_{t-1},d_t}) + Log(\gamma_{e_{t-1},e_t})\right) + \sum_t \left(Log\left(p_{c_t}(x_t)\right) + Log\left(p_{d_t}(y_t)\right)\right)$$

L'ajout de la dépendance de corrélation change le graphe en :

$$CDLL = \sum_{c_t,d_t,e_t} Log(\delta_{c_1}) + Log(\delta_{d_1}) + Log(\delta_{e_1})$$
$$+ \sum_t \left(Log(\gamma_{c_{t-1},e_t,c_t}) + Log(\gamma_{d_{t-1},e_t,d_t}) + Log(\gamma_{e_{t-1},e_t})\right)$$
$$+ \sum_t \left(Log\left(p_{c_t}(x_t)\right) + Log\left(p_{d_t}(y_t)\right)\right)$$

C'est équivalent a considérer des produit du type : $Log\left(\prod_t \gamma_{c_{t-1},e_t,c_t}\gamma_{d_{t-1},e_t,d_t}\gamma_{e_{t-1},e_t}\, p_{c_t}(x_t)p_{d_t}(y_t)\right)$

Qui se mets sous la forme récursive :

$$CDLL = Log\left(\prod_{c_t,d_t,e_t} p_{c_t}(x_t)p_{d_t}(y_t) \prod_{c_{t-1},d_{t-1},e_{t-1}} \dots .\gamma_{c_{t-1},e_t,c_t}\gamma_{d_{t-1},e_t,d_t}\gamma_{e_{t-1},e_t}p_{c_{t-1}}(x_{t-1})p_{d_{t-1}}(y_{t-1}) \prod_{c_{t-2},d_{t-2},e_{t-2}} \dots .\right)$$

Donc Faire un produit sur des triplets c_t, d_t, e_t est équivalent à faire un produit simple sur un espace tensoriel.

Cela est donc un produit matriciel sur les espaces tensoriels $c_{t-1}\otimes d_{t-1}\otimes e_{t-1}$, $c_{t-2}\otimes d_{t-2}\otimes e_{t-2}$ etc.. Suivit par une contraction sur les derniers indices : c_t, d_t, e_t

Ce qui se ré-exprime en :

$$CDLL = \sum_{j=1}^{m} u_j^c(1)\, Log(\delta_j^c) + \sum_{j=1}^{m} u_j^d(1)\, Log(\delta_j^d) + \sum_{j=1}^{m} u_j^e(1)\, Log(\delta_j^e)$$
$$+ \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{t=2}^{T} v_{j,k,l}^c(t)\, Log(\gamma_{j,k,l}^c)$$
$$+ \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{t=2}^{T} v_{j,k,l}^d(t)\, Log(\gamma_{j,k,l}^d) + \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{t=2}^{T} v_{j,k}^e(t)\, Log(\gamma_{j,k}^e)$$
$$+ \sum_{j=1}^{m}\sum_{t=1}^{T} u_j^c(t)\, Log\left(p_j^c(x_t)\right) + \sum_{j=1}^{m}\sum_{t=1}^{T} u_j^d(t)\, Log\left(p_j^d(y_t)\right)$$

Ou les sufficient statistics sont données par :

$$v_{j,k,l}^{x1} = prob[x1(t) = j \; \vdots \; x1(t-1) = k, e(t) = l]$$

$$v_{j,k,l}^{x2} = prob[x2(t) = j \; \vdots \; x2(t-1) = k, e(t) = l]$$

$v^{e}_{j,k} = prob[e(t) = j \vdots e(t-1) = k]$

$u^{x1}_{j} = prob[x1(t) = j\,]$

$u^{x2}_{j} = prob[x2(t) = j\,]$

$u^{e}_{j} = prob[e(t) = j\,]$

On calcule les quantités suivantes :

$v^{T}_{j,k,l,m,n,p} = prob[e(t) = j \& x_1(t) = k \& x_2(t) = l \vdots e(t-1) = m \& x_1(t-1) = n \& x_2(t-1) = p]$

donc $v^{e}_{j,k} = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{n=1}^{m} \sum_{p=2}^{m} v^{T}_{j,k,l,m,n,p}$ quantité que l'on obtient par quadruple marginalisation sur la statistique double slice marginal sur le triple produit tensoriel.

de même pour les autres statistiques

Quand on varie de manière indépendante les $\gamma^{c}_{j,k,l}$, $\gamma^{c}_{j,k,l}$ et $\gamma^{e}_{j,k}$, on obtient de manière évidente les problèmes stationnaires suivants :

On maximise $\sum_{j=1}^{m} \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{t=2}^{T} v^{x1}_{j,k,l}(t)\, Log\left(\gamma^{x1}_{j,k,l}\right)$

Sous la contrainte $\sum_{j=1}^{n} \gamma^{x1}_{j,k,l} = 1$ qui provient de ce que conditionnellement, la matrice de transition est une probabilité. Ce qui nous donne comme solution stationnaire :

$\gamma^{x1}_{j,k,l} = \frac{\sum_t v^{x1}_{j,k,l}(t)}{\sum_{j,t} v^{x1}_{j,k,l}(t)}$ Et de manière analogue pour les autres quantités :

$\gamma^{x2}_{j,k,l} = \frac{\sum_t v^{x2}_{j,k,l}(t)}{\sum_{j,t} v^{x2}_{j,k,l}(t)}$ Et $\gamma^{e}_{j,k} = \frac{\sum_t v^{e}_{j,k}(t)}{\sum_{j,t} v^{e}_{j,k}(t)}$

Généralisation

CDLL=

$$\sum_{i,[c_{1,i}]_i} \left(Log(\delta_{c_{1,i}})\right) + \sum_{t,[c_{t-1,i}]_i [c_{t,j}]_j} \left(Log\left(\gamma_{c_{t-1,i},c_{t,j}}\right)\right) + \sum_{t,i,[c_{t,i}]_i} \left(Log\left(p_{c_{t,i}}(x_{t,i})\right)\right)$$

Ou $[c_{1,i}]_i$ Signifie $c_{1,1}, c_{1,2}, c_{1,3}, \ldots.$

$\gamma_{c_{t-1,i},c_{t,j}}$ Signifie donc $\gamma_{c_{t-1,1},c_{t-1,2},c_{t-1,3},\ldots,c_{t,1},c_{t,2},c_{t,3},\ldots}$

Mais ceux-ci sont identiquement nul si il n'existe pas de nœud avec $c_{t-1,1}, c_{t-1,2}, c_{t-1,3}, \ldots, c_{t,1}$ en entrée ;

De même, seul un seul des j est actif dans $c_{t,j}$

Le nombre de i peut être quelconque, et en particulier $p_{c_{t,i}}(x_{t,i})$ peut être identiquement nul, ce qui signifie qu'il s'agit d'une chaine de markov de degre superieure (abstraite)

C'est équivalent à considérer des produits du type : $Log\left(\prod_{t,k,l,p,q>p} \gamma_{c_{t-1,k},c_{t,l}} \gamma_{d_{t,p},d_{t,q}} \prod_i p_{d_{t,i}}(x_{t,i})\right)$

Qui se mets sous la forme récursive :

$$Log\left(\prod_{c_t,d_t,e_t} p_{c_t}(x_t) p_{d_t}(y_t) \prod_{c_{t-1,1},\ldots,c_{t-1,n} d_{t,1},\ldots,d_{t,n}} \gamma_{c_{t-1,1},\ldots,c_{t,n}} \gamma_{d_{t,1},\ldots,d_{t,n}} p_{c_{t-1,i}}(x_{t-1}) \prod_{c_{t-2},d_{t-2},e_{t-2}} \ldots\right)$$

C'est à dire que l'on calcule tranche de date par tranche de date

HHMM avec 3 niveaux de nœuds cachés

$$CDLL = \sum_{c_t,d_t,e_t} Log(\delta_{c_1}) + Log(\delta_{d_1}) + Log(\delta_{e_1}) + \sum_t \left(Log(\gamma_{c_{t-1},c_t}) + Log(\gamma_{d_{t-1},c_t,d_t}) + Log(\gamma_{e_{t-1},d_t,e_t})\right) + \sum_t Log\left(p_{e_t}(x_t)\right)$$

$$CDLL = Log\left(\prod_{c_t,d_t,e_t} p_{e_t}(x_t) \prod_{c_{t-1},d_{t-1},e_{t-1}} \ldots \gamma_{c_{t-1},c_t} \gamma_{d_{t-1},c_t,d_t} \gamma_{e_{t-1},d_t,e_t} p_{e_{t-1}}(x_{t-1}) \prod_{c_{t-2},d_{t-2},e_{t-2}} \ldots\right)$$

Donc Faire un produit sur des triplets c_t, d_t, e_t est équivalent à faire un produit simple sur un espace tensoriel.

Cela est donc un produit matriciel sur les espaces tensoriels $c_{t-1} \otimes d_{t-1} \otimes e_{t-1}$, $c_{t-2} \otimes d_{t-2} \otimes e_{t-2}$ etc.. Suivit par une contraction sur les derniers indices : c_t, d_t, e_t

Ce qui se ré-exprime en :

$$LL = \sum_{j=1}^{m} u_j^c(1)\, Log\big(\delta_j^c\big) + \sum_{j=1}^{m} u_j^d(1)\, Log\big(\delta_j^d\big) + \sum_{j=1}^{m} u_j^e(1)\, Log\big(\delta_j^e\big) + \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{t=2}^{T} v_{j,k}^c(t)\, Log\big(\gamma_{j,k}^c\big)$$
$$+ \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{t=2}^{T} v_{j,k,l}^d(t)\, Log\big(\gamma_{j,k,l}^d\big) + \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{t=2}^{T} v_{j,k,l}^e(t)\, Log\big(\gamma_{j,k,l}^e\big)$$
$$+ \sum_{j=1}^{m}\sum_{t=1}^{T} u_j^c(t)\, Log\left(p_j^c(x_t)\right)$$

Ou les sufficient statistics sont données par :

$v_{j,k,l}^e = prob[e(t) = j \ \vdots\ e(t-1) = k, d(t) = l]$

$v_{j,k,l}^d = prob[d(t) = j \ \vdots\ d(t-1) = k, c(t) = l]$

$v_{j,k}^c = prob[c(t) = j \ \vdots\ c(t-1) = k]$

$u_j^e = prob[e(t) = j\,]$

$u_j^d = prob[d(t) = j\,]$

$u_j^c = prob[c(t) = j\,]$

On calcule les quantités suivantes :

$$v_{j,k,l,m,n,p}^T = prob[e(t) = j \& d(t) = k \& c(t) = l \ \vdots\ e(t-1) = m \& d(t-1) = n \& c(t-1) = p]$$

donc

$$v_{j,m,k}^e = \sum_{l=1}^{m}\sum_{n=1}^{m}\sum_{p=2}^{m} v_{j,k,l,m,n,p}^T$$

$$v_{j,m,k}^d = \sum_{l=1}^{m}\sum_{n=1}^{m}\sum_{p=2}^{m} v_{l,j,n,m,k,p}^T$$

$$v_{j,k}^c = \sum_{l=1}^{m}\sum_{n=1}^{m}\sum_{m=1}^{m}\sum_{p=2}^{m} v_{l,n,j,m,p,k}^T$$

quantité que l'on obtient par quadruple marginalisation sur la statistique double slice marginal sur le triple produit tensoriel .

Quand on varie de manière indépendante les $\gamma_{j,k,l}^e$, $\gamma_{j,k,l}^d$ et $\gamma_{j,k}^c$, on obtient de manière évidente les problèmes stationnaires suivants :

On maximise $\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{t=2}^{T} v_{j,k,l}^e(t)\, Log\big(\gamma_{j,k,l}^e\big)$

Sous la contrainte $\sum_{j=1}^{n} \gamma_{j,k,l}^{e} = 1$ qui provient de ce que conditionnellement, la matrice de transition est une probabilité. Ce qui nous donne comme solution stationnaire :

$\gamma_{j,k,l}^{e} = \frac{\sum_t v_{j,k,l}^{e}(t)}{\sum_{j,t} v_{j,k,l}^{e}(t)}$ Et de manière analogue pour les autres quantités :

$\gamma_{j,k,l}^{d} = \frac{\sum_t v_{j,k,l}^{d}(t)}{\sum_{j,t} v_{j,k,l}^{d}(t)}$ Et $\gamma_{j,k}^{c} = \frac{\sum_t v_{j,k}^{c}(t)}{\sum_{j,t} v_{j,k}^{c}(t)}$

HHMM avec une infinité de niveaux de nœuds cachés et Activation

Variable n_t

Représente le nombre de niveau réinitialisé. Valeur entre 0 et l'infini. Distribution :

$$\sum_{k=0}^{\infty} Prob[n_t = k] = 1$$

Et $i \to Prob[n_t = i]$ décroissante afin d'implémenter la notion d'abstraction croissante du HHMM

A priori $n_t = n_t(j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \dots, \infty)$

Donc $Prob[n_t = n \,|c_{0,t-1} = j_{0,t-1}, c_{1,t-1} = j_{1,t-1}, \dots, \infty] = b_n(j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \dots, \infty)$

Avec

$$\sum_{n=0}^{\infty} b_n(j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \dots, \infty) = 1$$

Initialisation des niveaux :

$$a_{n,j_{n,t}}(j_{n+1,t}) = Prob[c_{n,t} = j_{n,t}, c_{n+1,t} = j_{n+1,t}, n_t > n\,]$$

$$\sum_{i=1}^{Nmax[n]} a_{n,i}(j_{n+1,t}) = 1$$

Transition régulière:

$$a_{0,i}(j_{1,t}, j_{0,t-1}) = Prob[c_{0,t} = i, c_{0,t-1} = j_{0,t-1}, c_{1,t} = j_{1,t}, n_t = 0\,]$$

$$\sum_{i=1}^{Nmax[n]} a_{0,i}(j_{1,t}, j_{0,t-1}) = 1$$

Reconstruction de la flat matrice de transition (associée au flat HMM équivalent) les conditions en rouge sont celle synchrones , et les conditions en bleu sont celle de la date précédente

$Prob[c_{0,t} = j_{0,t}, c_{1,t} = j_{1,t}, \dots, c_{n,t} = j_{n,t}, c_{0,t-1} = j_{0,t-1}, c_{1t-1} = j_{1t-1}, \dots, c_{n,t-1} = j_{n,t-1}\,]$=

$$Prob[n_t = 0]Prob[c_{0,t} = j_{0,t}, c_{0,t-1} = j_{0,t-1}, c_{1,t} = j_{1,t} | n_t = 0] + \sum_{n=1}^{\infty} Prob[n_t = n]Prob[c_{n,t} = j_{n,t}, c_{n+1,t} = j_{n+1,t} \; | n_t > n]$$

Donc si n est le cutoff ou valeur maximale pour les niveaux

$$Prob[c_{0,t} = j_{0,t}, c_{1,t} = j_{1,t}, \ldots, c_{n,t} = j_{n,t}, c_{0,t-1} = j_{0,t-1}, c_{1t-1} = j_{1t-1}, \ldots, c_{n,t-1} = j_{n,t-1}] =$$

$$A_{j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \ldots, j_{n,t-1}, j_{0,t}, j_{1,t}, j_{2,t}, \ldots, j_{n,t}} =$$

$$b_0(j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \ldots, j_{n,t}) a_{0,j_{0,t}}(j_{1,t}, j_{0,t-1}) + \sum_{n=1}^{\infty} b_n(j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \ldots, \infty) a_{n,j_{n,t}}(j_{n+1,t})$$

Ayant identifié les etats et la matrice de transition ,

On a bien sur la probabilité initiale du reseau caché

$$Prob[c_{0,0} = j_{0,0}, c_{1,0} = j_{1,0}, \ldots, c_{n,0} = j_{n,0}] = \delta_{j_{0,0}, j_{1,0}, \ldots j_{n,0}}$$

$$\sum_{j_{0,0}, j_{1,0}, \ldots j_{n,0}} \delta_{j_{0,0}, j_{1,0}, \ldots j_{n,0}} = 1$$

La distribution de la variable observée est representée pala distribution

$$p_{t,j_{0,t}}(x, \theta)$$

Nous pouvons alors écrire la vraisemblance :

CDLL=

$$\sum_{j_{0,0}, j_{1,0}, \ldots j_{n,0}} \left(Log(\delta_{j_{0,0}, j_{1,0}, \ldots j_{n,0}}) \right) + \sum_{t, j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \ldots, j_{n,t-1}, j_{0,t}, j_{1,t}, j_{2,t}, \ldots, j_{n,t}} \left(Log(A_{j_{0,t-1}, j_{1,t-1}, j_{2,t-1}, \ldots, j_{n,t-1}, j_{0,t}, j_{1,t}, j_{2,t}, \ldots, j_{n,t}}) \right) + \sum_{t, j_{0,t}} \left(Log\left(p_{t,j_{0,t}}(x_{t,j_{0,t}}, \theta) \right) \right)$$

On effectue le remplacement dans cette formule de la matrice de transition pour examiner les simplifications et refactorings :

$$\sum_{j_{0,0},j_{1,0},..j_{n,0}} \left(Log\left(\delta_{j_{0,0},j_{1,0},..j_{n,0}}\right)\right)$$

$$+ \sum_{t,j_{0,t-1},j_{1,t-1},j_{2,t-1},\dots,j_{n,t-1},j_{0,t},j_{1,t},j_{2,t},\dots,j_{n,t}} \left(Log\left(b_0\left(j_{0,t-1},j_{1,t-1},j_{2,t-1},\dots,j_{n,t}\right)a_{0,j_{0,t}}\left(j_{1,t},j_{0,t-1}\right)\right.\right.$$

$$\left.\left.+\sum_{n=1}^{\infty} b_n\left(j_{0,t-1},j_{1,t-1},j_{2,t-1},\dots,\infty\right)a_{n,j_{n,t}}\left(j_{n+1,t}\right)\right)\right) + \sum_{t,j_{0,t}}\left(Log\left(p_{t,j_{0,t}}\left(x_{t,j_{0,t}},\theta\right)\right)\right)$$

Modularité du likelihood - Inference constante et modulaire