Soutenance MCMC

Online EM Algorithm for Hidden Markov Model
Article d'Olivier Cappé

Dadi Charles et De Myttenaere Arnaud

ENS Cachan

January 6, 2013

L'algorithme EM classique :

- permet de maximiser une log-vraisemblance.
- comporte 2 étapes :
    - E : calcul de l'espérance de la vraisemblance des données complètes
    - M : maximisation de cette quantité et actualisation des paramètres
- de nombreuses applications : modèles HMM, biostatistiques, traitement du langage, ...

Cas de l'article : Comment adapter cet algorithme au cas où les données arrivent au fur et à mesure ?

L'algorithme EM online :

- l'étape E ne peut pas reposer sur l'ensemble des données
- principe : remplacer le calcul de l'espérance de la vraisemblance par une étape d'approximation stochastique, en utilisant la formule des probablités totales :
  - calculer
$$\hat{\phi}_{n+1}(x) = \frac{\sum_{x' \in \mathcal{X}} \hat{\phi}(x') q_{\hat{\theta}_n}(x', x) g_{\hat{\theta}_n}(x, Y_{n+1})}{\sum_{x', x'' \in \mathcal{X}^2} \hat{\phi}(x') q_{\hat{\theta}_n}(x', x'') g_{\hat{\theta}_n}(x'', Y_{n+1})}$$
  - et
$$\hat{\rho}_{n+1}(x) = \sum_{x' \in \mathcal{X}} \left(\gamma_{n+1} s(x', x, Y_{n+1}) + (1 - \gamma_{n+1}) \hat{\rho}_n(x') \right) \frac{\hat{\phi}_n(x') q_{\hat{\theta}_n}(x', x)}{\sum_{x'' \in \mathcal{X}} \hat{\phi}_n(x'') q_{\hat{\theta}_n}(x'', x)}$$

- approximer l'espérance conditionnelle par

$$\sum_{x \in \mathcal{X}} \hat{\phi}_{n+1}(x) \hat{\rho}_{n+1}(x)$$

- En pratique, on laisse un certain temps de chauffe à l'algorithme.
  - L'étape M n'est donc réalisée pour la première fois qu'après un nombre n d'itérations E, fixé au préalable.

Objectifs :

- comprendre l'algorithme EM online,
- reproduire les résultats présentés dans l'article,
- discuter des hypothèses nécessaire à la convergence de l'algorithme.

Nous avons implémenté deux applications de l'algorithme EM online:

- mélanges de gaussiennes
- HMM (présenté ici).

Dans le cadre des HMM, la convergence est liée au théorème suivant : Si

i) la vraisemblance des données complète est de la forme exponentielle,
ii) $\mathcal{X}$ est un ensemble de cardinal fini,
iii) Θ est un ensemble compact et $\theta^\star$ est dans l'intérieur de Θ,
iv) la matrice de transition vérifie $q_\theta(x, x') \geq \varepsilon > 0$ pour tout $\theta \in \Theta$,
v) $\sup_\theta \sup_y \bar{g}_\theta(y) < \infty$
vi) $\mathbb{E}_{\theta^\star}[|\log \inf_\theta \bar{g}_\theta(Y_0)] < \infty$, où $\bar{g}_\theta(y) = \sum_x g_\theta(x, y)$

Alors :

i)

$$\frac{1}{n}\mathbb{E}_{\nu,\theta}\left[\sum_{t=1}^{n} s(X_{t-1}, X_t, Y_t)|Y_{0:n}\right] \rightarrow \mathbb{E}_{\theta^\star}\left[\mathbb{E}_\theta[s(X_{t-1}, X_t, Y_t)|Y_{-\infty;+\infty}]\right]$$

ii) Les points fixes de l'algorithme sont les points stationnaires de la vraisemblance.

Une grande partie de notre travail a consisté à vérifier ces hypothèses, et à les aléger si nécessaire :

- Sont-elles nécessaires ?
- Sont-elles toutes vérifiées dans le cas du HMM présenté dans l'article ?
- Si non, que vérifie le modèle présenté dans l'article ?

- forme exponentielle : OK, on retrouve les équations de l'article
- cardinal fini : OK car $\chi = \{0, 1\}$
- $sup_\theta sup_y \bar{g}_\theta(y) < \infty$: vrai presque surement
- $\mathbb{E}_{\theta^\star}[| \log \inf_\theta \bar{g}_\theta(Y_0)] < \infty$: vrai sur tout compact inclut dans Θ
- compacité : fausse. Cependant il s'agit d'un ensemble convexe
- $q_\theta(x, x') \geq \varepsilon > 0$ pour tout $\theta \in \Theta$: faux. Cependant, $q_\theta(x, x') > 0$ sur tout compact inclut dans Θ.

Cas de l'algorithme EM classique

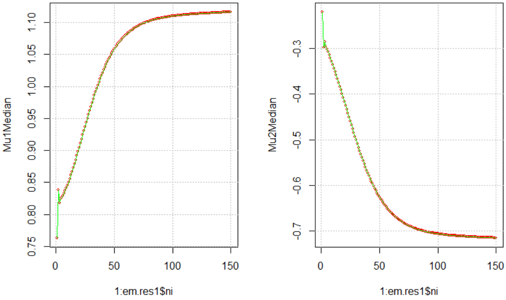
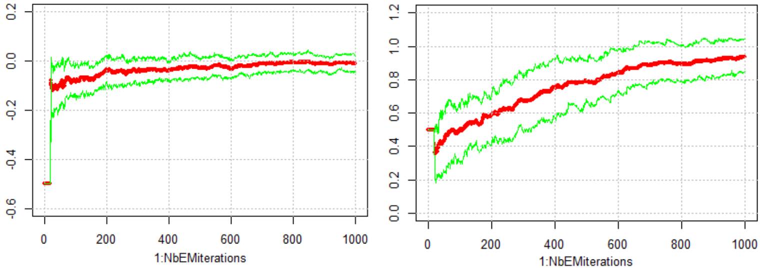


Figure : Estimation des moyennes

Cas de l'algorithme EM online



(a) $\hat{\mu}(1)$ (b) $\hat{\mu}(2)$

Figure : Estimations de la matrice de $\mu(1)$ et $\mu(2)$

Parmi les différences, on peut noter :

- La convergence s'établit plus lentement que dans le cas classique (5 itérations vs 30 à 40 itérations)

- Cette différence s'explique par le fait que le nombre de données utilisées à chaque itérations est considérablement moindre dans le cas en ligne
  - 20 observations à la 20^{eme} itération en ligne
  - 100 observations à chaque itération de l'algorithme EM classique.

- On peut envisager un algorithme qui combine EM en ligne et classique :
    - si les données mettent du temps à arriver, faire tourner le EM classique
    - prendre en compte la nouvelle données lorsqu'elle arrive
- Dans le cas où l'espérance de l'étape E n'est pas calculable, on peut penser à utiliser des méthodes de MCMC pour l'estimer...

- L'algorithme EM en ligne est une version adaptée de l'EM classique pour le cas où les données arrivent au fur et à mesure.
- Applications nombreuses : visions, langage, biostatistique, ...
- La convergence s'effectue sous certaine hypothèses suffisantes mais
  - ces hypothèses ne sont pas forcément nécessaires
  - hypothèses vraies sur tout compact dans Θ est suffisant ?
  - θ convexe suffisant ?