1 Pistes de réflexion Thèse - Janvier 2015

1.1 Introduction

Soient X_1,...,X_n, $X_i \in \mathbb{R}^p$ un échantillon de n variables aléatoires i.i.d de loi P^*. On suppose que P^* est un mélange de K composantes Gaussiennes $P_j, \quad j \in [K]$ et que K est connu:

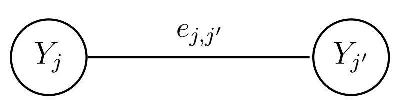
$$P^* = \sum_{j=1}^{K} \pi_j P_j \qquad avec \qquad \forall i, \quad \pi_i > 0 \qquad et \qquad \sum_{j=1}^{K} \pi_j = 1$$

Où P_j a pour densité la loi normale $\mathcal{N}_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$.

Une approche classique dans le cadre de mélange de gaussien est d'utiliser l'algorithme EM pour estimer les paramètres $\boldsymbol{\Pi} = (\pi_1, \ldots, \pi_n), \boldsymbol{\mu}, \boldsymbol{\Sigma}$. Sans hypothèse supplémentaire, la complexité de cette approche rend le calcul impossible quand la dimension p devient grande. Une approche est de considérer $\boldsymbol{\Sigma}$ diagonale mais ceci revient à considérer que les features sont indépendantes ce qui est un inconvénient.

1.2 Approche par analyse structurelle de Σ

Soient Y_1,...,Y_n un échantillon de n variables aléatoires i.i.d de loi $\mathcal{N}_p(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, $Y_i \in \mathbb{R}^p$. Supposons p grand. Nous cherchons a estimer μ^* et Σ^*.
Nous pouvons estimer la moyenne empirique $\widehat{\mu} = \bar{Y_n}$ et donc supposer que $\mu = 0$. Pour Σ^* ceci est plus délicat et nous chercherons des hypothèses structurelles sur $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega}$ la matrice de précision. Notamment en considérant que $\boldsymbol{\Omega}$ est creuse. Interprétation: si $\omega_{j,j'} = 0$ alors $Y^j \perp\!\!\!\perp Y^{j'} | \{Y^l \backslash (Y^j, Y^{j'})\}$.
Nous pouvons considérer le graphe $G = (V, E)$ avec card(V)=p alors,



et $e_{j,j'} \in E \iff \omega_{j,j'} \neq 0$

1.2.1 Graphical Lasso

On veut l'estimateur du maximum de vraisemblance pénalisé par la norme L_1.

$$\widehat{\mathbf{\Omega}} \in \arg\min_{\mathbf{\Omega} \geq 0} \Big\{ \log(\det \Omega) + \frac{1}{2} tr(\Omega S_n) + \lambda \parallel \Omega \parallel_1 \Big\}$$

avec $S_n = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^T$ la matrice de covariance et $\|\mathbf{\Omega}\|_1 = \sum_{j,j':j\neq j'} |\omega_{j,j'}|$ Ceci est un problème d'optimisation convexe, (problème d'optimisation SDP)

1.2.2 Column-wise Lasso

(Théorème) Si $Y \sim \mathcal{N}_p(0, (\Omega^*)^{-1})$, alors

$$Y_j = \sum_{j' \in [p] \backslash j} \big(- \frac{\omega^*_{jj'}}{\omega^*_{jj}} \big) Y^{j'} + \frac{1}{\sqrt{\omega^*_{jj}}} \xi^j$$

où $\xi^j \sim \mathcal{N}_p(0,1) \perp\!\!\!\perp \{Y^{j'} : j' \in [p] \backslash j\}$

1.2.3 Méthode d'estimation

$\forall j \in \{1, .., p\}$ on estime le vecteur $\frac{1}{\omega_{jj}} \begin{bmatrix} \omega_{j1} \\ \ddots \\ \omega_{jj} \end{bmatrix}$ On estime β par la méthode du Lasso:

$$\widehat{\beta} \in \arg\min_{\widehat{\beta} \in \mathbb{R}^{p-1}} \big\{ \parallel Y^j - Y^{-j}\beta_{-j} \parallel_2^2 + \lambda \parallel \beta_{-j} \parallel_1 \big\}$$

ou Y^{-j} matrice des $Y_{j'} \backslash Y_j$.
On pose

$$\widehat{\omega}_{jj} = \left(\frac{\parallel Y^j - Y{-}j\beta_{-j} \parallel_2^2}{n} \right)^{-1} \qquad \text{et} \qquad \widehat{\omega}_{jj'} =$$

On a un lasso en dimension p, la theorie recommande:

$$\lambda = \sqrt{\frac{2\log(p)}{n}} \frac{1}{\sqrt{\omega^*_{jj}}}$$

1.2.4 Square-root Lasso

$$\widehat{\beta} \in \arg \min_{\widehat{\beta} \in \mathbb{R}^{p-1}} \left\{ \| Y^j - Y^{-j}\beta_{-j} \|_2 + \lambda \| \beta_{-j} \|_1 \right\}$$

la theorie recommande:

$$\lambda = \sqrt{\frac{2 \log(p)}{n}}$$

(Theoreme) $\quad \forall \lambda > 0, \exists \lambda_1 > 0 \quad | \quad \widehat{\beta}^{\sqrt{Lasso}}(\lambda) = \widehat{\beta}^{Lasso}(\lambda_1)$

1.3 Objectifs

Injecter cette méthode du column-wise lasso dans EM sous l'hypothese $\Omega_j = \Sigma_j^{-1}$ est creuse. Nous pourrons commencer par un cas simple, $\Omega_j = \Omega, \forall j$

1.4 Notation pratique

Nous pourrons utiliser la notation suivante:

$$\widehat{\beta}^{Lasso} \in \arg \min_{\substack{\widehat{B} \in \mathbb{R}^p \\ B_{jj}=1}} \left\{ \| Y.B \|_2^2 + \lambda \| B \|_1 \right\}$$

avec $Y = [Y_1, ..., Y_n]^T \in \mathbb{R}^{n,p}$ et $\| Y\widehat{B} \|_2^2 = \sum_{j=1}^{p} \| Y\widehat{B}_j \|_2^2$