Sélection de modèle

15 septembre 2010

litis
Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes

Aurélie Boisbunon, Stéphane Canu, Dominique Fourdrinier
Université de Rouen et INSA de Rouen, LITIS EA 4108

Etat de l'art

- Perspective décisionnelle $\rightarrow$ approche estimation de coût.
- Cadre distributionnel sphérique.
- 1$^{\text{ère}}$ mise en oeuvre $\rightarrow$ sélection de variables / estimateur MC

Contexte

Modèle linéaire

$$Y = X\beta + \varepsilon$$

où $Y \in \mathbb{R}^n$, $X \in \mathcal{M}_{n,p}$, $\beta \in \mathbb{R}^p$ inconnu, et $\varepsilon \in \mathbb{R}^n$ le bruit.
On suppose que $p < n$.

- Solution usuelle : $\hat{\beta}^{MC} = (X^T X)^{-1} X^T Y$
  - Problèmes d'interprétabilité, d'instabilité, d'inadmissibilité pour plus de 3 dimensions, ..., de cette solution.
- Hypothèse : $Y = X_I \beta_I + \varepsilon$
- **Problème : Comment trouver I ?**

Estimateur de β

LASSO (Tibshirani 1996)

$$\hat{\beta}^{LASSO} = \arg\min \left\{ \| Y - X\beta \|^2 + \lambda \| \beta \|_1 \right\}, \quad \lambda \geq 0$$

- Propose un chemin de régularisation quand λ varie
- **Estimation de $I \equiv$ estimation de λ**

Problème : comment choisir l'hyperparamètre λ optimal ?

- CV, GCV $\rightarrow$ très coûteux, pas adapté pour grandes dimensions,
- AIC, BIC $\rightarrow$ basé sur la connaissance a priori de la distribution des erreurs,
- **Notre proposition : estimation de coût.**

Estimation du coût

Transformation orthogonale du modèle sous forme canonique

$$Y \xrightarrow{G} \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + \eta$$

Etape 1 : définition d'une fonction de coût pour $\hat{\beta}^{LASSO}$

$$L(\hat{\beta}^{LASSO}, \beta) = \| \hat{\beta}^{LASSO} - \beta \|^2$$

Etape 2 : Définition de l'estimateur du coût de $\hat{\beta}^{LASSO}$

- Estimateur sans biais : $\mathbb{E}[\delta_0] = \mathbb{E}\left[L(\hat{\beta}^{LASSO}, \beta)\right] = R(\hat{\beta}^{LASSO})$
- Estimateur amélioré : $\delta_\gamma(\lambda) = \delta_0(\lambda) - \| U \|^4 \gamma$

Etape 3 : Recherche du minimum de l'estimateur de coût

$$\hat{\lambda} = \arg\min_{\lambda \in \mathbb{R}_+} \delta(\lambda)$$

Résultats préliminaires

Protocole : X fixé, $\beta = (0, \ldots, 0, 10, \ldots, 10, 0, \ldots, 0, 10, \ldots, 10, 0, \ldots, 0)^T$ fixé, r répliques de ε générées.

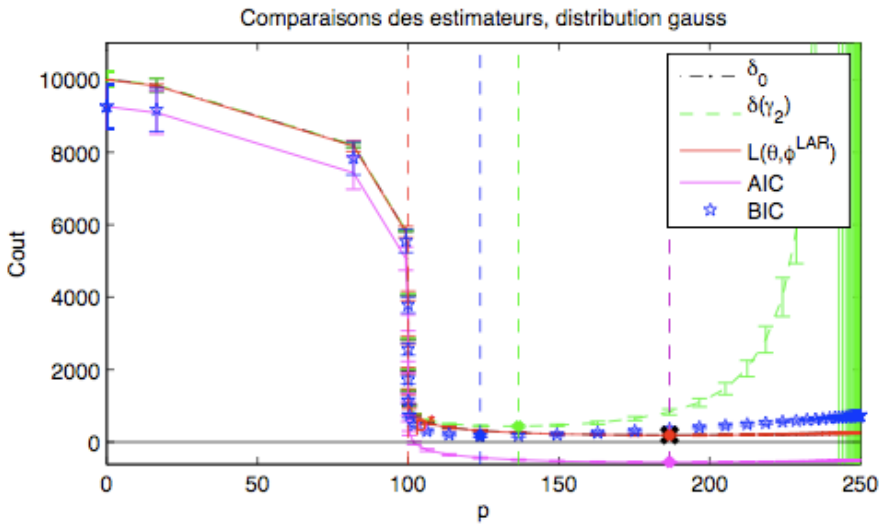

Figure: Comparaison des estimateurs

- δ_0 estime mieux le coût que δ_γ
- Le coût réel estime mal I (biais du LASSO).
- δ_γ meilleur que δ_0 pour estimer I.
- δ_γ meilleur que AIC mais moins bon que BIC et δ_0 équivalent à AIC.

Avantages de notre proposition et perspectives

- Robustesse distributionnelle
  - Lois à symétrie sphérique
- Applicable à d'autres fonctions de coût et d'autres estimateurs :
  - Amélioration pour l'estimation du LASSO
  - Adaptation pour la classification
- Généralisation au cas $p \geq n$ à développer.