Estimation distribuée d'une espérance conditionnelle

Igor Colin

30 juin 2014

Rappels

Estimation de fonction

Objectif et formulation

- Objectif : regrouper les utilisateurs par centres d'intérêts communs
- Notations :
  - $(X_i)_{1 \leq i \leq n}$: caractéristiques des utilisateurs (musiques, historique des conversations, etc.)
  - $D : (X, Y) \mapsto D(X, Y)$: fonction de dissimilarité entre deux vecteurs de caractéristiques
  - P : partition des utilisateurs
  - Φ_P : fonction d'appartenance au même *cluster*

Problème

- Nouvel objectif : trouver la solution du problème

$$\min_P w(P) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} D(X_i, X_j) \Phi_P(X_i, X_j) \right)$$

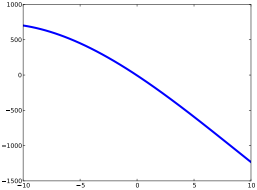
- Idée : estimer $f : x \mapsto \mathbb{E}[D(x, X)\Phi_P(x, X)]$
- Contrainte : les $(X_i)_{1 \leq i \leq n}$ ne sont pas simultanément accessibles

Rappels

Estimation de fonction

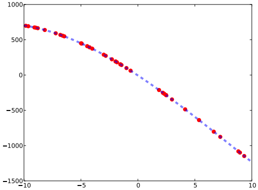
Méthode générale de regression

- Notations :
  - f : fonction à estimer



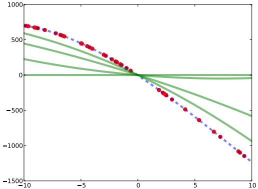
Méthode générale de regression

- Notations :
  - f : fonction à estimer
  - $\{(x_i, f(x_i))\}_{1 \leq i \leq n}$: observations



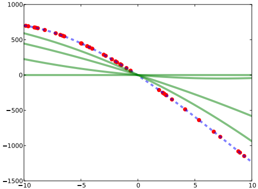
Méthode générale de regression

- Notations :
  - f : fonction à estimer
  - $\{(x_i, f(x_i))\}_{1\leq i\leq n}$: observations
  - $\hat{f} : (x;\theta) \mapsto \hat{f}(x;\theta)$: estimateur



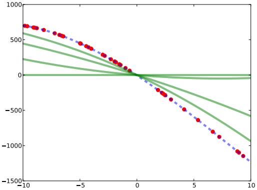
Méthode générale de regression

- Notations :
  - f : fonction à estimer
  - $\{(x_i, f(x_i))\}_{1 \leq i \leq n}$: observations
  - $\hat{f} : (x; \theta) \mapsto \hat{f}(x; \theta)$: estimateur
  - $\hat{R} : \theta \mapsto \hat{R}(\theta)$: risque empirique



Méthode générale de regression

- Notations :
  - f : fonction à estimer
  - $\{(x_i, f(x_i))\}_{1 \leq i \leq n}$: observations
  - $\hat{f} : (x;\theta) \mapsto \hat{f}(x;\theta)$: estimateur
  - $\hat{R} : \theta \mapsto \hat{R}(\theta)$: risque empirique



- Objectif : trouver θ^* solution de

$$\min_{\theta \in \Theta} \hat{R}(\theta)$$

Exemple

- Exemple : estimation polynomiale
  - $\hat{f} : (x;\theta) \mapsto \theta_0 + \theta_1 x + \theta_2 x^2$,
  - $\hat{R}(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left(\hat{f}(x_i) - f(x_i)\right)^2$
- Qualité dépendante du choix de $\hat{f}$

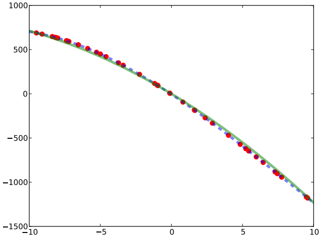


FIGURE: $\hat{f}$ adaptée.

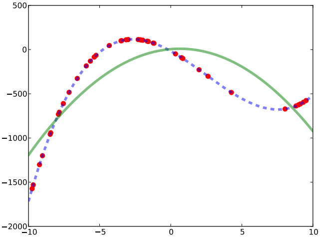


FIGURE: $\hat{f}$ non adaptée.

Application au problème initial

- Fonction à estimer : $f : x \mapsto \mathbb{E}\left[D(x,X)\Phi_P(x,X)\right]$
- Risque empirique : moindres carrés
- Estimateur à noyaux :

$$\hat{f}(x;\theta,\mathbf{w},a) = a + \sum_{j=1}^{m} w_j K(x-\theta_j)$$

où K est un noyau gaussien.

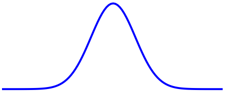


FIGURE: Noyau gaussien 1D

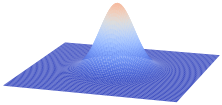


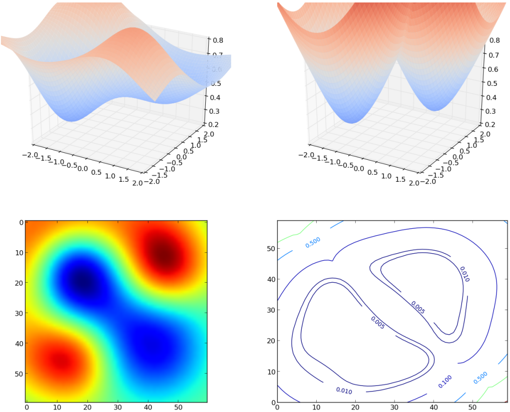
FIGURE: Noyau gaussien 2D

Résultats

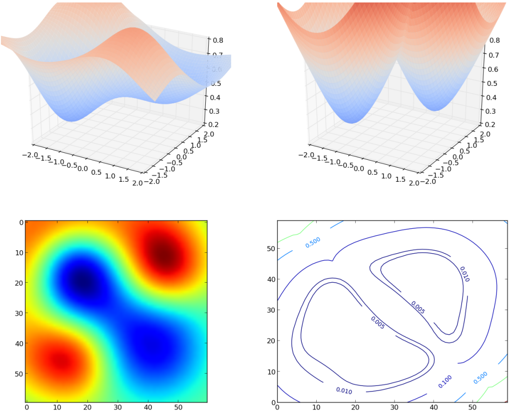
- Application :
  - $D : (x, y) \mapsto \|x - y\|_2$
  - $(X_i)_{1 \leq i \leq n}$: n tirages d'un mélange de gaussiennes
  - Observations empiriques $\tilde{f}(x_i) = \frac{1}{n} \sum_{j=1}^{n} D(x_i, x_j) \Phi_p(x_i, x_j)$
- Objectif :

$$\min_{\theta, w, a} \hat{R}(\theta, w, a) = \frac{1}{2n} \sum_{i=1}^{n} \left(\hat{f}(x; \theta, w, a) - \tilde{f}(x_i) \right)^2$$

Résultats optimisation globale



Résultats optimisation distribuée



Contrainte supplémentaire

- $\tilde{f}(X_i)$ non accessible
- Idée : si l'agent i a accès aux caractéristiques $(X_j)_{j \in \mathcal{A}_i}$ $(i \in \mathcal{A}_i)$, il peut :
  - estimer la distribution des X à partir des observations accessibles
  - utiliser le risque R_i :

$$R_i : (\theta, w, a) \mapsto \mathbb{E}_{\mu_i}\left[\left(\hat{f}(x; \theta, w, a) - \mathbb{E}_{\mu_i}[D(x, X)\Phi_P(x, X)]\right)^2\right]$$

- Le problème devient alors

$$\min_{\theta, w, a} \frac{1}{n} \sum_{i=1}^{n} R_i(\theta, w, a)$$

Résultats