

Estimation distribuée d'une espérance conditionnelle

Igor Colin

29 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

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

- ▶ 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
- ▶ $\{(x_i, f(x_i))\}_{1 \leq i \leq n}$: observations
- ▶ Estimateur $\hat{f} : (x; \theta) \mapsto \hat{f}(x; \theta)$
- ▶ $\hat{R} : \theta \mapsto \hat{R}(\theta)$: risque empirique associé à θ

- ▶ Objectif : trouver θ^* solution de

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

- ▶ Exemple : estimation polynomiale

- ▶ $\hat{f} : (x; \theta) \mapsto \theta_0 + \theta_1 x + \theta_2 x^2,$
- ▶ $\hat{R}(\theta) = \sum_{i=1}^n \left(\hat{f}(x_i) - f(x_i) \right)^2$