M2 Probabilités et Statistiques, Université Paris-Sud
Apprentissage statistique (Sylvain Arlot et Francis Bach)
Convexification du risque

Francis Bach

Cours 3, 15 février 2010

1 Rappels d'optimisation

Bonnes références : [3] (disponible gratuitement en ligne), [2]

- Problème primal : $f^* = \inf_{x \in \mathcal{X}} f(x)$ tel que $\forall i \in \{1, \dots, m\}, h_i(x) = 0, \forall j \in \{1, \dots, r\}, g_j(x) = 0$
- Lagrangien : $\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x)$
- Fonction duale : $q(\lambda, \mu) = \inf_{x \in \mathcal{X}} \mathcal{L}(x, \lambda, \mu)$
- Problème dual (toujours concave) : $d^* = \sup_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r_+} q(\lambda, \mu)$
- Dualité faible (sans hypothèses) : $d^* \leqslant f^*$
- Dualité forte (avec hypothèses de convexité) : $d^* \leqslant f^*$
- Inversion min/max
- Conditions de Slater
- Conditions d'optimalité de KKT

2 Convexification du risque en classification binaire

- Etant donné une fonction f de $\mathcal{X}$ and $\mathbb{R}$, on considère le classifieur g de $\mathcal{X}$ and $\mathbb{R}$, défini par $g(x) = \text{sign}(f(x))$.
- Risque : $R(f) = \mathbb{E}\phi_{0-1}(Yf(X))$
- ϕ-Risque : $R_\phi(f) = \mathbb{E}\phi_{0-1}(Yf(X))$ où ϕ est une fonction de contraste
- Régression logistique : $\phi(u) = \log(1 + e^{-u})$ (interprétation en modèle probabiliste, bien spécifié ou non)
- Moindres carrés : $\phi(u) = (u - 1)^2$
- SVM : $\phi(u) = \max(0, 1 - u)$ (interprétation géométrique)

3 Liens entre les risques [1]

- Hypothèse : ϕ convexe
- Définition d'un contraste bien calibré, équivalent à ϕ dérivable en 0 et $\phi'(0) < 0$.
- Théorème :

$$\psi(R(f) - R^*) \leqslant R_\phi(f) - R_\phi^*$$

  où $\psi(\theta) = \phi(0) - \inf_\alpha \frac{1+\theta}{2}\phi(\alpha) + \frac{1-\theta}{2}\phi(-\alpha)$

Références

[1] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473) :138–156, 2006. (Was Department of Statistics, U.C. Berkeley Technical Report number 638, 2003).

[2] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization*. Number 3 in CMS Books in Mathematics. Springer-Verlag, 2000.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2003.