Examen du cours Apprentissage Statistique

Masters Math. UPMC – Spécialité Big Data

16/02/2016

Durée 2 h

Notes de cours autorisées

On dispose d'un ensemble de données non étiquetées, on s'intéresse à construire un système qui réponde « 1 » sur des données proches de cet ensemble d'apprentissage et «-1 » ailleurs. Pensez à un problème de classification où l'on ne dispose que d'exemples positifs et pas d'exemple négatifs.

On considère un ensemble de données $\{x^1, \dots, x^N\}$, avec $x^i \in R^n$, on va définir un classifieur à marge qui répond « 1 » dans un voisinage de ces données et « -1 » partout ailleurs.

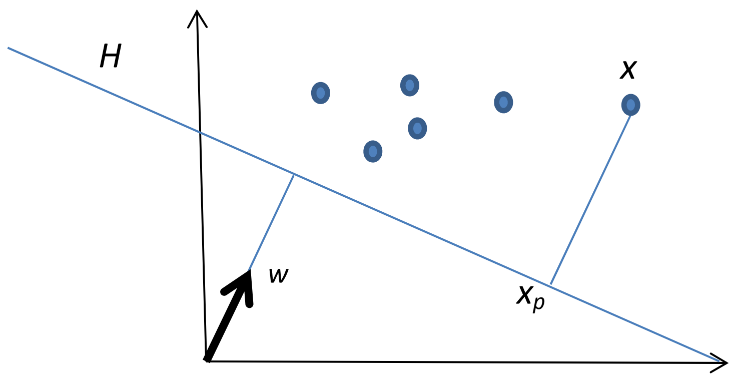
Pour cela, on va résoudre le problème suivant (pb 1):

$$\min_{w,\rho} \frac{1}{2} \|w\|^2 - \rho$$

sous les contraintes $w.x^i \geq \rho$

avec $\rho \in R$

1. Considérons la figure suivante



1.1. Soit H l'hyperplan d'équation $F(x) = w.x - \rho = 0$. Pour un point x , on note x_p sa projection sur l'hyperplan H. Montrer que la distance de x à H s'écrit $\frac{F(x)}{\|w\|}$. On pourra par exemple écrire x sous la forme $x = x_p + r\frac{w}{\|w\|}$ où r est la distance recherchée. En déduire que la distance de l'origine à H vaut $\frac{|\rho|}{\|w\|}$.

1.2. La fonction de décision sera $G(x) = sgn(F(x)$. Positionner sur la figure un hyperplan qui résout le problème (pb 1)

1.3. Cet hyperplan est-il unique (justifier) ?

2. Ecrire le Lagrangien $L(w, \rho, \alpha)$du problème (pb 1).

3. Calculer les dérivées $\frac{\partial L}{\partial w_j}$ et $\frac{\partial L}{\partial \rho}$.
4. En déduire $w = \sum_{i=1}^{N} \alpha_i x_i$ et $\sum_{i=1}^{N} \alpha_i = 1$
5. Montrer que la formulation duale du problème (pb 1) s'écrit (pb 2) :

$$\max_{\alpha} \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j x^i . x^j$$

sous les contraintes $\alpha_i \geq 0 \; et \; \sum_{i=1}^{N} \alpha_i = 1$

6. Montrer qu'à l'optimum, seuls les points pour lesquels $w. x^i = \rho$ ont un coefficient $\alpha_i$non nul.
7. Montrer $\rho = w. x^i$ pour x^i vecteur support quelconque.

On va maintenant développer un algorithme pour résoudre le problème (pb 2) .

Considérons que toutes les variables $\alpha_i$sont fixées pour $i = 3..N$, on va chercher à optimiser $\alpha_1$et α_2.

8. Montrer que le problème de minimisation pour α_1, α_2 peut se re-écrire sous la forme:

$$\max_{\alpha_1,\alpha_2} \frac{1}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j x^i . x^j + \sum_{i=1}^{2} \alpha_i \, C_i + C$$

sous les contraintes $\alpha_1, \alpha_2 \geq 0$ et $\sum_{i=1}^{2} \alpha_i = R$ avec $R = 1 - \sum_{j=3}^{N} \alpha_j$

avec $C_i = \sum_{j=3}^{N} \alpha_j x^i . x^j$ et $C = \sum_{i,j=3}^{N} \alpha_i \alpha_j x^i . x^j$

9. On note Q $= \frac{1}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j x^i . x^j + \sum_{i=1}^{2} \alpha_i \, C_i$

Montrer en utilisant les contraintes du problème :

$$Q = \frac{1}{2}(R - \alpha_2)^2 (x^1)^2 + (R - \alpha_2)\alpha_2 x^1 . x^2 + \frac{1}{2}\alpha_2^2 (x^2)^2 + (R - \alpha_2) C_1 + \alpha_2 C_2$$

10. Exprimer la solution $\alpha_2$qui minimise *Q*, et en déduire α_1 et montrer que ces solutions résolvent $\min_{\alpha_1,\alpha_2} \frac{1}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j x^i . x^j + \sum_{i=1}^{2} \alpha_i \, C_i + C$.
11. Proposer la forme générale d'un algorithme pour résoudre le problème (pb 2).
12. Lien avec la classification à deux classes. Supposons que (w^*, ρ^*) soit une solution du problème (pb 1), montrer que $(w^*, 0)$ paramétrise l'hyperplan séparateur optimal pour l'ensemble étiqueté $\{(x^1, 1), \dots, (x^N, 1)(-x^1, -1), \dots (-x^N, -1)\}$.