Master Masef - Mido 2015-2016

Examen : Machine Learning in Finance [1] : Durée 1h30 - Correction

Exercice 1. [9]pt

Q1 : Support Vector Machine

Q2 : Nombre Maximum de points que l'on puisse trouver que l'on peut classifier de toutes les façons possible

Q3 : Oui voir l'exemple du cours avec la fonction sinus

Q4 : $d+1$

Q5 : Pas forcément car il est possible qu'il existe une autre famille de k points de $\mathbf{R}^d$ qui puissent être classifiés de toutes les façons possible par la famille de classificateurs $\mathcal{C}_2$

Q6 : La probabilité d'erreur en prédiction est controlée pas la la qualité de la calibration et la VC de cette famille

Q7 : Non mais on ne peut plus utiliser l'inégalité de VC

Q8 :

a) Pour avoir plus de robustesse et controler mieux la VC

b) Un seul.

c) Il est orthogonal au segment reliant les deux points où la distance entre les deux enveloppes convexes des points à séparer est minimale

Q9 : Un simplexe

Q10 : Oui car la VC des hyperplans de marge 0.1 est de l'ordre de $(\frac{1}{0.1})^2$

Q11 : De potentiellement mieux séparer les données en dimension supérieure (et induire dans l'espace de départ des séparations plus complexes).

Q12 : A trouver des classificateurs dans l'espace des observations qui correspondent à des SVMs dans un autre espace

Q13 : Non parmi les points mal classifiés peuvent figurer des vecteurs supports (cas $\alpha_i = C$)

Q14 :

a) Ils sont tous sur une sphère de centre 0 et de rayon 1

b) Un pourcentage fixe des points, d'après les théorème de convergence des $\nu-$ SVM

c) $\sum \alpha_i y_i K(x_i, x)$

Exercice 2. [11pt]

[0.5pt] 1) $L(w,b,\rho,\zeta_i,\alpha_i,\beta_i) = \frac{1}{2}\|w\|^2 - 2\rho + \mu \sum\limits_{i=1}^{i=l} \zeta_i - \sum\limits_{i=1}^{i=l} \alpha_i \left[y_i(w.x_i+b) - \rho + \zeta_i\right] - \sum\limits_{i=1}^{i=l} \beta i \zeta_i$

[0.5pt] 2) $\max\limits_{\alpha_i \geq 0, \beta_i \geq 0} L(w,b,\rho,\zeta,\alpha_i,\beta_i) = +\infty$ si (1) ou (2) non satisfaites

$\max\limits_{\alpha_i \geq 0, \beta_i \geq 0} L(w,b,\rho,\zeta,\alpha_i,\beta_i) = \frac{1}{2}\|w\|^2 - 2\rho + \mu \sum\limits_{i=1}^{i=l} \zeta_i$ si (1) et (2) satisfaites

d'où le résultat

1. Pierre Brugière Université Paris 9 Dauphine

3)

[0.5pt] $\frac{\partial L}{\partial w} = \left[\frac{\partial L}{\partial w^1}, \frac{\partial L}{\partial w^2} \cdots, \frac{\partial L}{\partial w^l}\right] = w' - \sum\limits_{i=1}^{i=l} \alpha_i y_i x_i'$

[0.5pt] $\frac{\partial L}{\partial b} = -\sum\limits_{i=1}^{i=l} \alpha_i y_i,$

[0.5pt] $\frac{\partial L}{\partial \rho} = -2 + \sum\limits_{i=1}^{i=l} \alpha_i,$

[0.5pt] $\frac{\partial L}{\partial \zeta_i} = \mu - \alpha_i - \beta_i$

4) (D_μ) est le problème de minimisation sous contraintes de $\max\limits_{\alpha_i \geq 0, \beta_i \geq 0} \min\limits_{w,b,\rho,\zeta_i} L(w, b, \rho, \zeta, \alpha_i, \beta_i)$,

$\min\limits_{w,b,\rho,\zeta_i} L(w, b, \rho, \zeta, \alpha_i, \beta_i)$ est obtenue en utilisant les valeurs de w, b, ρ et β_i obtenues en annulant les dérivées partielles $\frac{\partial L}{\partial w}$, $\frac{\partial L}{\partial b}$, $\frac{\partial L}{\partial \rho}$, $\frac{\partial L}{\partial \zeta_i}$, calculées précédemment

[1.50pt] (D_μ) est le problème

$$\max_{\alpha_i \geq 0, \beta_i \geq 0} -\frac{1}{2}\| \sum_{i=1}^{i=l} \alpha_i y_i x_i \|^2,$$

$$\begin{cases} \sum\limits_{i=1}^{i=l} \alpha_i y_i = 0 \\ \sum\limits_{i=1}^{i=l} \alpha_i = 2 \\ \mu - \alpha_i = \beta_i \end{cases}$$

que l'on peut aussi écrire sous la forme

$$-\min_{\alpha_i} \frac{1}{2}\| \sum_{i=1}^{i=l} \alpha_i y_i x_i \|^2,$$

$$\begin{cases} \sum\limits_{i=1}^{i=l} \alpha_i y_i = 0 \\ \sum\limits_{i=1}^{i=l} \alpha_i = 2 \\ 0 \leq \alpha_i \leq \mu \end{cases}$$

5)

[0.5pt] a) Lorsque μ devient trop petit (D_μ) n'a plus de solution car les conditions

$$\begin{cases} \sum\limits_{i=1}^{i=l} \alpha_i = 2 \\ 0 \leq \alpha_i \leq \mu \end{cases}$$

ne peuvent être satisfaites

[0.5pt] b) Lorsque $\mu \to +\infty$ l'erreur de classement à un coût infini, donc soit les points peuvent peuvent être tous classifiés correctement et la solution limite est la solution du système à classiication parfaite :

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

$y_i(w.x_i + b) \geq \rho$.

soit il n'est pas possible de classifier tous les points correctement et la solution du système va être $+\infty$.

6) **[0.25pt]** a) Trivial

[0.25pt] b) pour $\mu \geq 1$ $H_\mu = H_1$ et $H_{-\mu} = H_{-1}$

en dessous d'une certaine valeur de μ les contraintes ne peuvent être satisfaites et $H_{+\mu} = H_{-\mu} = \emptyset$

[0.25pt] c) pour $\mu \geq 1$ H_μ et $H_{-\mu}$ sont les enveloppes convexes au sens usuel

pour $\mu \leq 1$ H_μ et $H_{-\mu}$ sont les enveloppes convexes réduites

[0.25pt] 7) Le carré de la distance entre l'enveloppe convexe des points classifiés $+1$ et l'enveloppe convexe des points classifiés -1

[1pt] 8) Il suffit de remarquer que $\| \sum_{i:y_i=1} \alpha_i x_i - \sum_{i:y_i=-1} \alpha_i x_i \|^2 = \| \sum_{i=1}^{i=l} \alpha_i y_i x_i \|^2$ et que

$$\begin{cases} \sum_{i:y_i=1} \alpha_i = 1 \\ \sum_{i:y_i=-1} \alpha_i = 1 \\ 0 \leq \alpha_i \leq \mu \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^{i=l} \alpha_i = 2 \\ \sum_{i=1}^{i=l} \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq \mu \end{cases}$$

les deux problèmes vont donc avoir les même solutions.

9) **[0.5pt]** Les problèmes Primal et Dual ont les mêmes solutions donc les contraintes de KKT sont satisfaites pour les solutions et on a

$$\begin{cases} \alpha_i(\mu)\left[y_i(w(\mu).x_i + b(\mu)) - \rho(\mu) + \zeta_i(\mu)\right] = 0 \\ (\mu - \alpha_i(\mu))\zeta_i(\mu) = 0 \end{cases}$$

[0.5pt] Si $0 < \alpha_i(\mu) < \mu$ alors nécessairement d'après les équations de KKT on a $\zeta_i(\mu) = 0$ et donc $[y_i(w(\mu).x_i + b(\mu)) - \rho(\mu)] = 0$

[0.5pt] 10) Mal classé signifie que $\zeta_i(\mu) > 0$ donc d'après KKT nécéssairement $\alpha_i(\mu) = \mu$

11) on peut réécrire $p(\mu) = \frac{1}{2} \sum_{i=1}^{i=l} \alpha_i(\mu) x_i$ et on remarque que $p(\mu)$ est le milieu du segment qui définit la distance minimum entre les deux convexes $H_{+\mu}$ et $H_{-\mu}$ et $w(\mu)$ est le vecteur qui relie ces deux points

[1pt] a) on multiplie par y_i chaque équation $\alpha_i(\mu)\left[y_i(w(\mu).x_i + b(\mu)) - \rho(\mu) + \zeta_i(\mu)\right] = 0$ ce qui donne : $\alpha_i(\mu)\left[w(\mu).x_i + b(\mu) - y_i\rho(\mu) + y_i\zeta_i(\mu)\right] = 0$

et en sommant toutes ces équations on obtient

$$\sum_{i=1}^{i=l} \alpha_i(\mu) x_i.w(\mu) + b(\mu) \sum_{i=1}^{i=l} \alpha_i(\mu) - \rho(\mu) \sum_{i=1}^{i=l} \alpha_i(\mu) y_i + \sum_{i=1}^{i=l} \alpha_i(\mu)\zeta_i(\mu) y_i = 0$$

en utilisant

$\sum\limits_{i=1}^{i=l} \alpha_i(\mu)x_i = 2p(\mu)$,

$\sum\limits_{i=1}^{i=l} \alpha_i(\mu) = 2$,

$\sum\limits_{i=1}^{i=l} \alpha_i(\mu)y_i = 0$ et

$(\mu - \alpha_i(\mu))\zeta_i(\mu) = 0$ on obtient :

$2p(\mu).w(\mu) + 2b(\mu) - \rho(\mu) \times 0 + \mu \sum\limits_{i=1}^{i=l} \zeta_i(\mu)y_i = 0$

soit

$b(\mu) = -w(\mu).p(\mu) - \frac{\mu}{2} \sum\limits_{i=1}^{i=l} y_i\zeta_i(\mu)$

[1pt] b) si on somme les équations :

$\alpha_i(\mu) \left[y_i(w(\mu).x_i + b(\mu)) - \rho(\mu) + \zeta_i(\mu)\right] = 0$ on obtient

$\sum\limits_{i=1}^{i=l} \alpha_i(\mu)y_ix_i.w(\mu) + b(\mu) \sum\limits_{i=1}^{i=l} \alpha_i(\mu)y_i - \rho(\mu) \sum\limits_{i=1}^{i=l} \alpha_i(\mu) + \sum\limits_{i=1}^{i=l} \alpha_i(\mu)\zeta_i(\mu) = 0$

ce qui entraine cette fois

$w(\mu).w(\mu) + b(\mu) \times 0 - \rho(\mu) \times 2 + \sum\limits_{i=1}^{i=l} \mu\zeta_i(\mu) = 0$

soit

$\rho(\mu) = \frac{1}{2}w(\mu).w(\mu) + \frac{\mu}{2} \sum\limits_{i=1}^{i=l} \zeta_i(\mu)$

[1pt] c) l'hyperplan est orthogonal au segment joignant les points réalisant la distance minimum entre les deux enveloppes convexes réduites $H_{+\mu}$ et $H_{-\mu}$. Il passe par la milieu de ce segment ssi $\rho(\mu) = 0$ et si il y a des erreurs de classification dans la solutions de D_μ ($\sum\limits_{i=1}^{i=l} \zeta_i(\mu) \neq 0$) il est translaté.