Master Masef - Mido 2015-2016

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

Exercice 1. [9]*pt*

Q1 : Que signifie « SVM » ?

Q2 : Quelle est la définition de la « VC »d'une famille de classificateurs ?

Q3 : Est-il possible de trouver un exemple de famille de classificateurs dont la dimension est supérieure au nombre de paramètres ?

Q4 : Quand on fait de la classification dans $\mathbf{R}^d$ par les hyperplans de $\mathbf{R}^d$ quelle est la VC que l'on obtient pour cette famille de classificateurs ?

Q5 : Si il existe k points de $\mathbf{R}^d$ qui peuvent être classifiés de toutes les façons possible par une famille de classificateurs $\mathcal{C}_1$ et pas par la famille de classificateurs $\mathcal{C}_2$ cela signifie-t'il obligatoirement que $VC(\mathcal{C}_1) > VC(\mathcal{C}_2)$?

Q6 : Quand on choisit un classificateur dans une famille, comment la VC de cette famille relie-t'elle la qualité de la calibration au contrôle de la qualité de la prédiction ?

Q7 : Une VC élevée voir infinie pour une famille de classificateurs entraine t'elle obligatoirement une mauvaise qualité prédictive ?

Q8 : Quand des observations $\mathbf{R}^d$ peuvent être parfaitement classifiées à l'aide d'hyperplans de $\mathbf{R}^d$

a) pourquoi cherche t-on à les séparer par un hyperplan de marge maximale plutôt que par un hyperplan quelconque ?

b) combien y a t'il d'hyperplans de marge maximale qui séparent ces points ?

c) que peut-on dire géométriquement d'un hyperplan de marge maximale par rapport aux deux classes de points qu'il sépare ?

Q9 : Quelle est la configuration géométrique des points sur un sphère de $\mathbf{R}^d$ qui va permettre d'en séparer le plus grand nombre possible de toutes les facons possible par des hyperplans de marge donnée ?

Q10 : Y a t'il une grande différence entre i) la VC de la famille des hyperplans de $\mathbf{R}^{10^9}$ appliqués à la boule de centre 0 et de rayon 1 et ii) la VC de la famille des hyperplans de $\mathbf{R}^{10^9}$ et de marge 0.1 appliqués à cette même boule ?

Q11 : Quel est l'intérêt en classification de transformer des observations x_i de $\mathbf{R}^d$ en de nouvelles données $\phi(x_i)$ dans un nouvel espace de dimension supérieure voir infinie ?

Q12 : A quoi sert le théorème de Mercer en classification ?

Q13 : Les vecteurs supports ($\alpha_i \neq 0$) sont-ils toujours correctement classifiés dans un problème de classification par hyperplans de marge maximale avec erreurs pénalisées ?

Q14 : Quand on utilise en classification le noyau $exp(-\frac{\|x-y\|^2}{2})$ et la transformation $\phi()$ associée de $\mathbf{R}^d$ dans l'espace de Hilbert $(H, <>)$

a) citer une propriété géométrique importante des points transformés dans $(H, <>)$?

1. Pierre Brugière Université Paris 9 Dauphine

b) que pensez-vous obtenir en général comme nombre de vecteurs supports dans $(H, <>)$?

c) quelle grandeur regarde-t'on dans $\mathbf{R}^d$ pour classifier les points ?

Exercice 2. [11pt]

Pour une famille $\{(x_i, y_i)\}_{i\epsilon\{1,2,..,l\}}$ de points de $\mathbf{R}^d$ de classification $y_i\epsilon\{0,1\}$ on considère pour $\mu > 0$ le problème d'optimisation (P_μ) suivant :

$$\min_{w,b,\rho,\zeta_i} \frac{1}{2}\|w\|^2 - 2\rho + \mu \sum_{i=1}^{i=l} \zeta_i,$$

$$\begin{cases} y_i(w.x_i + b) \geq \rho - \zeta_i \\ \zeta_i \geq 0 \end{cases}$$

1) Ecrire le Lagrangien du système $L\,(w, b, \rho, \zeta_i, \alpha_i, \beta_i)$

2) Justifier pourquoi résoudre le problème d'optimisation (P_μ) est équivalent à résoudre $\min\limits_{w,b,\rho,\zeta_i} \max\limits_{\alpha_i\geq 0,\beta_i\geq 0} L\,(w, b, \rho, \zeta, \alpha_i, \beta_i)$,

3) Calculer les dérivés directionnelles : $\frac{\partial L}{\partial w}$, $\frac{\partial L}{\partial b}$, $\frac{\partial L}{\partial \rho}$, $\frac{\partial L}{\partial \zeta_i}$

4) Ecrire le problème dual du problème (P_μ) que l'on appelle (D_μ)

5) Que se passe t-il pour les solutions de (D_μ) quand :

a) $\mu \to 0$

b) $\mu \to +\infty$

On considère maintenant les ensembles

$$H_{+\mu} = \{ \sum_{i:y_i=1} \alpha_i x_i / \sum_{i:y_i=1} \alpha_i = 1 \text{ et } 0 \leq \alpha_i \leq \mu \}$$

et

$$H_{-\mu} = \{ \sum_{i:y_i=-1} \alpha_i x_i / \sum_{i:y_i=-1} \alpha_i = 1 \text{ et } 0 \leq \alpha_i \leq \mu \}$$

6) a) montrer que $H_{+\mu}$ et $H_{-\mu}$ sont convexes

b) comment $H_{+\mu}$ et $H_{-\mu}$ varient-ils lorsque μ varie ?

c) que représentent les ensembles $H_{+\mu}$ et $H_{-\mu}$?

7) Que représente géométriquement la solution du problème (Q_μ) suivant

$$\min_{0\leq\alpha_i\leq\mu} \| \sum_{i:y_i=1} \alpha_i x_i - \sum_{i:y_i=-1} \alpha_i x_i \|^2$$

$$\begin{cases} \sum\limits_{i:y_i=1} \alpha_i = 1 \\ \sum\limits_{i:y_i=-1} \alpha_i = 1 \end{cases}$$

8) Montrer que les problèmes (Q_μ) et (D_μ) ont les mêmes solutions en $\alpha_i(\mu)$.

On suppose maintenant que les $\{(x_i, y_i)\}_{i\epsilon\{1,2,..,l\}}$ et μ sont tels que (D_μ) a une solution non nulle et qui est la même que celle du problème (P_μ).

9) Trouver une condition suffisante sur $\alpha_i(\mu)$ pour que le vecteur x_i vérifie pour le problème (P_μ) : $w(\mu).x_i + b(\mu) = \rho(\mu)$

10) Si x_i est un vecteur support mal classé dans le problème (P_μ) que peut on dire de $\alpha_i(\mu)$?

On note : $p(\mu) = \frac{1}{2}(\sum\limits_{i:y_i=1} \alpha_i(\mu)x_i + \sum\limits_{i:y_i=-1} \alpha_i(\mu)x_i)$ et $w(\mu) = \sum\limits_{i=1}^{i=l} \alpha_i(\mu)y_ix_i$

11) Montrer que les solutions $b(\mu)$ et $\rho(\mu)$ du problème (P_μ) vérifient :

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

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

c) interpréter la position géométrique de l'hyperplan d'équation :
$w(\mu).x + b(\mu) = \rho(\mu)$