

Mido - Master MASEF : 8 mars 2017
Examen : Machine Learning in Finance¹ : Durée 2 heures

Quizz [10 points]

Q1 : Que signifie « SVM » ?

- a) Special Vector Matrix
- b) Support Vector Machine
- c) Symmetric Vector Module

Q2 : Si $(X_i, Y_i)_{i \in \llbracket 1, n \rrbracket}$ est un échantillon d'apprentissage comment est définie l'erreur de calibration pour un problème de classification où f est le classificateur ?

- a) $\frac{1}{n} \sum_{i=1}^{i=n} 1_{f(X_i) \neq Y_i}$
- b) $E[|f(X) - Y|]$
- c) aucun des choix proposés ci-dessus

Q3 : Si $(X_i, Y_i)_{i \in \llbracket 1, n \rrbracket}$ est un échantillon d'apprentissage, parmi les hypothèses suivantes lesquelles font parties des hypothèses toujours faites dans le cours ?

- a) les X_i sont gaussiennes
- b) les (X_i, Y_i) sont toutes de même loi
- c) les (X_i, Y_i) sont toutes indépendantes
- d) aucun des choix proposés ci-dessus

Q4 : Si $R_n(f_n)$ est l'erreur de calibration pour le classificateur f_n associé à l'échantillon $(X_i, Y_i)_{i \in \llbracket 1, n \rrbracket}$ et si l'on note $R(f_n) = E[1_{f_n(X_{n+1}) \neq Y_{n+1}}]$ quelles propriétés sont vraies ?

- a) $R_n(f_n) = R(f_n)$
- b) $R_n(f_n) > R(f_n)$
- c) $E[R_n(f_n)] \leq R(f_n)$
- d) aucun des choix proposés ci-dessus

Q5 : Si il existe k points de $\mathbf{R}^d$ qui peuvent être classifiés de toutes les façons possible par une famille de classificateurs $\mathcal{F}_1$ et pas par $\mathcal{F}_2$ alors parmi les propositions suivantes lesquelles sont forcément vraies ?

- a) $VC(\mathcal{F}_1) = k$
- b) $VC(\mathcal{F}_1) \geq k$
- c) $VC(\mathcal{F}_1) > VC(\mathcal{F}_2)$
- d) aucun des choix proposés ci-dessus

Q6 : Si une famille de classificateurs est paramétrisée par p paramètres, parmi les assertions suivantes lesquelles sont vraies ?

- a) nécessairement $VC(\mathcal{F}) < +\infty$

1. Pierre Brugière Université Paris 9 Dauphine

- b) nécessairement $VC(\mathcal{F}) = p + 1$
- c) aucun des choix proposés ci-dessus

Q7 : L'inégalité de Vapnik Chervonenkis permet connaissant la complexité du modèle utilisé et l'erreur de calibration réalisée sur l'échantillon :

- a) de calculer la valeur exacte de $R(f_n)$
- b) de définir un intervalle de confiance pour $R(f_n)$
- c) de calculer $P(R(f_n) > R_n(f_n))$
- d) aucun des choix proposés ci-dessus

Q8 : Quelles propositions sont vraies ?

- a) une VC infinie pour une famille de classificateurs entraîne obligatoirement une bonne qualité prédictive
- b) une VC infinie pour une famille de classificateurs entraîne obligatoirement une calibration parfaite
- c) la VC d'une famille de classificateurs de $\mathbf{R}^d$ est nécessairement inférieure ou égale à $d + 1$
- d) le nombre de façon de $\{0, 1\}$ -classifier k points est de 2^k

Q9 : Dans $\mathbf{R}^d$ pour $w \neq 0$ on note $H_{w,b}$ l'hyperplan défini par $H_{w,b} = \{x \in \mathbf{R}^d, \langle w, x \rangle + b = 0\}$. Quelles propositions sont vraies ?

- a) $\forall x \in \mathbf{R}^d \ d(x, H_{w,b}) = \frac{|\langle w, x \rangle + b|^2}{\|w\|^2}$
- b) $\forall x \in \mathbf{R}^d \ d(x, H_{w,b}) = \frac{|\langle w, x \rangle + b|}{\|w\|}$
- c) $\forall x \in \mathbf{R}^d \ d(x, H_{w,b}) = \frac{|\langle w, x \rangle - b|}{\|w\|}$
- d) aucun des choix proposés ci-dessus

Q10 : Dans $\mathbf{R}^d$ pour $w \neq 0$ on note $H_{w,b}$ l'hyperplan défini par $H_{w,b} = \{x \in \mathbf{R}^d, \langle w, x \rangle + b = 0\}$. Quelles propositions sont vraies ?

- a) $d(H_{w,b_1}, H_{w,b_2}) = \frac{|b_2 + b_1|}{\|w\|}$
- b) $d(H_{w,b_1}, H_{w,b_2}) = \frac{|b_2 - b_1|}{\|w\|}$
- c) $d(H_{-w,b_1}, H_{w,b_2}) = \frac{|b_2 + b_1|}{\|w\|}$
- d) aucun des choix proposés ci-dessus

Q11 : Si vous pouvez classifier parfaitement un échantillon à l'aide des hyperplans H_{w_1,b_1} avec une marge Δ_1 et H_{w_2,b_2} avec une marge Δ_2 vous avez une bonne raison de choisir l'hyperplan H_{w_1,b_1} plutôt que H_{w_2,b_2} si :

- a) $\Delta_1 < \Delta_2$
- b) $\Delta_1 > \Delta_2$
- c) $b_1 > b_2$
- d) $b_2 > b_1$

Q12 : Si on peut classifier parfaitement un échantillon formé de deux classes à l'aide d'un hyperplan $H_{w,b}$ de marge Δ que dire des enveloppes convexes des

deux classes ?

- a) elles sont séparables
- b) Δ est plus grand que la distance entre les deux enveloppes convexes
- c) certains points des enveloppes convexes sont sur les bords de l'hyperplan si il est de marge maximale
- d) aucun des choix proposés ci-dessus

Q13 : Quelle majoration peut on utiliser pour la dimension de Vapnik de la famille $\mathcal{F}$ d'hyperplans de marges minimum Δ classifiant des points à l'intérieur de la boule de centre 0 et diamètre D de $\mathbf{R}^d$?

- a) $VC(\mathcal{F}) \leq \min(d, \frac{\Delta}{D}) + 1$
- b) $VC(\mathcal{F}) \leq \min(d, (\frac{D}{\Delta})^2) + 1$
- c) $VC(\mathcal{F}) \leq \min(d, (\frac{\Delta}{D})^2) + 1$
- d) $VC(\mathcal{F}) \leq \min(d, D\Delta) + 1$

Q14 : Le théorème du minimax assure que :

- a) $\max_{z \in \mathcal{Z}} \left[\min_{y \in \mathcal{Y}} g(y, z) \right] \leq \min_{y \in \mathcal{Y}} \left[\max_{z \in \mathcal{Z}} g(y, z) \right]$
- b) $\min_{y \in \mathcal{Y}} \left[\max_{z \in \mathcal{Z}} g(y, z) \right] \leq \max_{z \in \mathcal{Z}} \left[\min_{y \in \mathcal{Y}} g(y, z) \right]$
- c) $\max_{z \in \mathcal{Z}} \left[\min_{y \in \mathcal{Y}} g(y, z) \right] = \min_{y \in \mathcal{Y}} \left[\max_{z \in \mathcal{Z}} g(y, z) \right]$
- d) aucun des choix proposés ci-dessus

Q15 : Quand on résoud par C-SVM une classification $\{-1, 1\}$ quelles assertions sont vraies ?

- a) d'après le théorème du minimax les solutions d^* du problème Dual et p^* du problème Primal vérifient nécessairement $d^* \leq p^*$
- b) le problème Primal et Dual ont la même solution uniquement si on rajoute les conditions de KKT dans l'écriture du problème Dual
- c) le problème Primal et Dual ont la même solution vue la nature particulière du problème
- d) les conditions de KKT sont automatiquement satisfaites vue la nature particulière du problème

Q16 : Quand on résoud avec un C-SVM une classification $\{-1, 1\}$ quelles assertions sont vraies ?

- a) tous les vecteurs supports sont toujours classifiés correctement
- b) tous les vecteurs supports sont toujours sur les bords de l'hyperplan solution
- c) l'hyperplan solution est orthogonal à une certaine combinaison linéaire des vecteurs x_i de l'échantillon

Q17 : Parmi les fonctions suivantes de $\mathbf{R}^d \times \mathbf{R}^d$ dans $\mathbf{R}$ lesquelles sont des noyaux ?

- a) $\exp(-\frac{\|x-y\|^2}{2}) + \exp(-\frac{\|x-y\|^2}{4})$

- b) $\langle x, y \rangle \exp(-\frac{\|x-y\|^2}{2})$
c) $\exp(-\frac{\|x-y\|^2}{2}) - \exp(-\frac{\|x-y\|^2}{4})$

On considère dans toute le suite du QCM le noyau $K_\sigma(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2})$ de $\mathbf{R}^n$ et ϕ_σ telle que : $\langle \phi_\sigma(x), \phi_\sigma(y) \rangle = \exp(-\frac{\|x-y\|^2}{2\sigma^2})$

Q18 : Quelles assertions sont vraies ?

- a) $\forall x \in \mathbf{R}^n, \|\phi_\sigma(x)\| = 1$
b) $\forall x, y \in \mathbf{R}^n \|\phi_\sigma(x) - \phi_\sigma(y)\| > \sqrt{2}$
c) $\forall (x_i)_{i \in \llbracket 1, l \rrbracket} \in \mathbf{R}^n \exists w \in Vect\{\phi_\sigma(\mathbf{R}^n)\}, \|w\| = 1$ et $\forall i \in \llbracket 1, l \rrbracket, \langle \phi_\sigma(x_i), w \rangle \geq \frac{1}{n}$

Q19 : Quelles assertions sont vraies ?

- a) $\forall (x_i)_{i \in \llbracket 1, l \rrbracket}$ la matrice $[K_\sigma(x_i, x_j)]_{i, j \in \llbracket 1, l \rrbracket}$ est symétrique
b) $\forall (x_i)_{i \in \llbracket 1, l \rrbracket}$ la matrice $[K_\sigma(x_i, x_j)]_{i, j \in \llbracket 1, l \rrbracket}$ définit une forme quadratique positive
c) $\forall x, y \in \mathbf{R}^n, \langle \phi_\sigma(x), \phi_\sigma(y) \rangle \geq \frac{1}{\sqrt{2}}$

Q20 : Les $(x_i)_{i \in \llbracket 1, l \rrbracket}$ étant des points distincts de $\mathbf{R}^n$ que dire de la fonction

$$f(x) = \lim_{\sigma \rightarrow 0} \sum_{i=1}^l K_\sigma(x_i, x) ?$$

- a) elle vaut zéro partout
b) elle vaut 1 partout
c) elle vaut 1 en chaque point x_i et zéro ailleurs

Problème :

On note $\mathbf{N}$ l'ensemble des entiers naturels et pour tout élément z de $\mathbf{R}^{\mathbf{N}}$ on note $(z^j)_{j \in \mathbf{N}}$ ses composantes et $H = \{z \in \mathbf{R}^{\mathbf{N}}, \sum_{j \in \mathbf{N}} (z^j)^2 < +\infty\}$

On note $\langle \cdot, \cdot \rangle$ le produit scalaire sur H défini par $\langle u, v \rangle = \sum_{j \in \mathbf{N}} u^j v^j$ et on note

$\|\cdot\|$ la norme associée, définie par $\|z\| = \sqrt{\langle z, z \rangle}$.

On note ξ le vecteur de $\mathbf{R}^n$ de composantes $(\xi^i)_{i \in \llbracket 1, n \rrbracket}$

Pour une famille finie de points $(z_i)_{i \in \llbracket 1, n \rrbracket}$ de H on considère le problème :

$$(P_\nu) \begin{cases} \inf_{w \in \mathbf{R}^n, \rho \in \mathbf{R}, \xi \in \mathbf{R}^n} \frac{1}{2} \|w\|^2 - 2\rho + \frac{\nu}{n} \sum_{i=1}^n \xi^i \\ \forall i \in \llbracket 1, n \rrbracket, \langle w, z_i \rangle \geq \rho - \xi^i & (C1) \\ \forall i \in \llbracket 1, n \rrbracket, \xi^i \geq 0 & (C2) \end{cases}$$

[0.5pt]1) Ecrire le Lagrangien $L(w, \rho, \xi, \alpha, \beta)$ du système où $\alpha = (\alpha^i)_{i \in \llbracket 1, n \rrbracket}$ est associé aux n contraintes (C1) et $\beta = (\beta^i)_{i \in \llbracket 1, n \rrbracket}$ est associé aux n contraintes (C2)

[1pt] 2) Calculer les dérivées $\frac{\partial L}{\partial w}$, $\frac{\partial L}{\partial \rho}$, $\frac{\partial L}{\partial \xi^i}$

[0.5pt] 3) On note $w(\alpha, \beta)$, $\rho(\alpha, \beta)$, $\xi(\alpha, \beta)$ les solutions (dependant de α et β) de $\frac{\partial L}{\partial w} = 0$, $\frac{\partial L}{\partial \rho} = 0$, $\frac{\partial L}{\partial \xi^i} = 0$.
Calculer l'expression de $L(w(\alpha, \beta), \rho(\alpha, \beta), \xi(\alpha, \beta), \alpha, \beta)$ en fonction de α et β et des $(z_i)_{i \in \llbracket 1, n \rrbracket}$.

[1pt] 4) Montrez que le problème dual du problème (P_ν) est le problème (D_ν) défini par :

$$(D_\nu) \begin{cases} \sup_{\alpha \in \mathbf{R}^n} -\frac{1}{2} \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \alpha^i \alpha^j \langle z_i, z_j \rangle \\ \sum_{i=1}^{i=n} \alpha^i = 2 \\ \forall i \in \llbracket 1, n \rrbracket, 0 \leq \alpha^i \leq \frac{\nu}{n} \end{cases}$$

On note $\alpha(\nu)$ la solution en α de (D_ν) quand ce problème a une solution (et dans ce cas on admet que le sup est atteint et en un point unique).

5)[1pt] a) Que se passe-t-il pour les solutions de (P_ν) et (D_ν) quand $\nu \rightarrow 0$ et en dessous de quelle valeur ν_L de ν observe-t-on un phénomène "critique" ?

[1pt] b) Que se passe-t-il pour les solutions de (P_ν) et (D_ν) quand $\nu \rightarrow +\infty$ et au dessus de quelle valeur ν_H de ν observe-t-on un phénomène "critique" ?
Que valent les solutions $\xi(\nu)$ de (P_ν) si $\nu > \nu_H$.

[1pt] c) On suppose ici $\nu_L < \nu < \nu_H$. Comment fait-on une fois les $\alpha(\nu)$ trouvés pour (D_ν) pour calculer l'argument $\rho(\nu)$ de (P_ν) (on suppose que l'inf est atteint et en un point unique) ?

[0.5pt] d) Pour quelle valeur de $\alpha_i(\nu)$ un vecteur z_i peut-il être un vecteur support mal classé pour le problème (P_ν) ?

On suppose maintenant que $\forall i \in \llbracket 1, n \rrbracket, \|z_i\| = 1$ et on considère le problème :

$$(B_\nu) \begin{cases} \inf_{c \in \mathbf{R}^n, R \in \mathbf{R}, \xi \in \mathbf{R}^n} 2R^2 + \frac{\nu}{n} \sum_{i=1}^{i=n} \xi^i \\ \forall i \in \llbracket 1, n \rrbracket, \|z_i - c\|^2 \leq R^2 + \xi^i \\ \forall i \in \llbracket 1, n \rrbracket, \xi^i \geq 0 \end{cases}$$

[0.5pt] 6) Ecrire le Lagrangien $G(c, R, \xi, \alpha, \beta)$ du système (B_ν)

[1pt] 7) Calculer les dérivées $\frac{\partial G}{\partial c}$, $\frac{\partial G}{\partial R}$, $\frac{\partial G}{\partial \xi^i}$

[1.5pt] 8) Calculez le problème dual noté (C_ν) du problème (B_ν)

[0.5pt] 9) Que remarquez-vous ?