Master ISF Apprentissage - Mido 10 novembre 2016

Quizz : Machine Learning in Finance [1] : Durée 1 heure

Quizz [15points]

Q1 : Que signifie « SVM » ?

Support Vector Machines

Q2 : Si $(X_i, Y_i)_{i\in[\![1,n]\!]}$ est un échantillon d'apprentissage comment est définie l'erreur de calibration pour un problème de classification ?

a) $\frac{1}{n}\sum\limits_{i=1}^{i=n} 1_{f(X_i)\neq Y_i}$

b) $\frac{1}{n}\sum\limits_{i=1}^{i=n} |f(X_i) - Y_i|$

c) $E[|f(X) - Y|]$

d) aucun des choix proposés ci-dessus

Q3 : Si $(X_i, Y_i)_{i\in[\![1,n]\!]}$ est un échantillon d'apprentissage quelles sont les hypothèses 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 retenu pour l'échantillon $(X_i, Y_i)_{i\in[\![1,n]\!]}$ et si l'on note $R(f_n) = E[1_{f(X_{n+1})\neq Y_{n+1}}]$ quelles propriétés suivantes sont vraies ?

a) $R_n(f_n) = R(f_n)$

b) $R_n(f_n) > R(f_n)$

c) $R_n(f_n) \leq R(f_n)$

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 la famille de classificateurs $\mathcal{F}_2$ alors quelles propositions suivantes sont forcémment 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 deux familles de classificateurs $\mathcal{F}_1$ et $\mathcal{F}_2$ ont la même erreur de calibration sur un échantillon d'apprentissage donné quel classificateur l'inégalité de Vapnik Chernovenkis encourage t'elle à retenir pour la prédiction :

a) le classificateur de $\mathcal{F}_1$ si $VC(\mathcal{F}_1) > VC(\mathcal{F}_2)$

[1] Pierre Brugière Université Paris 9 Dauphine

b) le classificateur de $\mathcal{F}_2$ si $VC(\mathcal{F}_1) > VC(\mathcal{F}_2)$
c) aucun des choix proposés ci-dessus

Q7 : 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 ?
a)2^d
b)$d-1$
c)$d+1$
d)aucun des choix proposés ci-dessus

Q8 : L'inégalité de Vapnik Chervonenkis permet connaissant la complexité du modèle utilisé et l'erreur de calibration réalisée sur l'échantillon :
a) de définir un intervalle de confiance pour $R(f_n)$
b) de calculer la valeur exacte de $R(f_n)$
c) de déterminer les outliers dans les observations
d) aucun des choix proposés ci-dessus

Q9 : Dans $\mathbf{R}$ on considère la famille de $\{0,1\}$-classifiers $\mathcal{F} = \{1_{x<\alpha}, 1_{x\geq\alpha}\}_{\alpha\in\mathbf{R}}$ quelle est la valeur de $VC(\mathcal{F})$:
a) 1
b) 2
c) 3
d) aucun des choix proposés ci-dessus

Q10 : Quelles propositions sont vraies
a) une VC élevée voir infinie pour une famille de classificateurs entraine obligatoirement une mauvaise qualité prédictive
b) une VC élevée voir infinie pour une famille de classificateurs entraine obligatoirement une calibration parfaite
c) la VC d'une famille de classificateurs est toujours proche du nombre de paramètres
d) le nombre de façon de $\{0,1\}$-classifier k points est de 2^k

Q11 : Quelles propositions sont vraies
a) il y a peu de chances qu'une machine qui calibre mal prédise bien
b) le théorème de Vapnik Chervonenkis ne donne pas de garantie qu'une machine très complexe predise aussi bien qu'elle calibre
c) SRM signifie Structural Risk Maximization
d) une machine très complexe au sens de VC prédit toujours très bien

Q12 : 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) $\forall x \in \mathbf{R}^d \ d(x, H_{w,b}) = \frac{|\langle w,x\rangle+b|}{\|w\|}$
b) $\forall x \in \mathbf{R}^d \ d(x, H_{w,b}) = \frac{|\langle w,x\rangle+b|}{\|w\|^2}$
c) $\forall x \in \mathbf{R}^d \ d(x, H_{w,b}) = \frac{|\langle w,x\rangle-b|}{\|w\|^2}$

d) aucun des choix proposés ci-dessus

Q13 : 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

Q14 : Si vous pouvez classifier parfaitement un échantillon à l'aide des hyperplans H_{w_1,b_1} de marge Δ_1 et H_{w_2,b_2} de marge Δ_2 vous avez une bonne raison de choisir l'hyperplan H_{w_1,b_1} si :
a) $\Delta_1 < \Delta_2$
b) $\Delta_1 > \Delta_2$
c) $b_1 > b_2$
d) $b_2 > b_1$

Q15 : 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 separables
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

Q16 : Pour la famille $\mathcal{F}$ de classificateurs de diamètre 1 de $\mathbf{R}^{1000}$ formé d'hyperplans de marge 0.1
a) $VC(\mathcal{F}) = 1001$
b) $VC(\mathcal{F}) = 1000$
c) $VC(\mathcal{F}) \approx 100$
d) $VC(\mathcal{F}) \approx 10$

Q17 : Le théorème du minimax assure que
a) $\min_{y\in\mathcal{Y}} \left[\max_{z\in\mathcal{Z}} g(y,z)\right] \leq \max_{z\in\mathcal{Z}} \left[\min_{y\in\mathcal{Y}} g(y,z)\right]$
b) $\max_{z\in\mathcal{Z}} \left[\min_{y\in\mathcal{Y}} g(y,z)\right] \leq \min_{y\in\mathcal{Y}} \left[\max_{z\in\mathcal{Z}} g(y,z)\right]$
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

Q18 : Quand on résoud un SVM pour une classification $\{-1, 1\}$ quelles assertions sont vraies :

a) le problème Primal et Dual ont la même solution uniquement si on prend en compte les conditions de KKT dans l'écriture du problème dual
b) le problème Primal et Dual ont la même solution vu la nature particulière du problème
c) les conditions de KKT sont automatiquement satisfaites vu la nature particulière du problème
d) il y a trois conditions de KKT

Q19 : Quand on résoud un C-SVM pour une classification $\{-1, 1\}$ quelles assertions sont vraies :
a) tous les vecteurs supports sont forcémment classifiés correctement
b) il est possible que certains vecteurs supports soient classifiés incorrectement
c) tous les vecteurs supports sont forcémment sur les bords de l'hyperplan séparateur à marge solution
d) l'hyperplan séparateur est orthogonal à une certaine combinaison linéaire des x_i de l'échantillon

Q20 : A quoi sert le théorème de Mercer en classification ?
à faire un changement de coordonnées sur les variables afin de pouvoir mieux les séparer et de savoir quelles fonctions $K(x, y)$ correspondent à un produit scalaire $< \phi(x), \phi(y) >$

Q21 : Parmi les fonctions suivantes de $\mathbf{R}^d \times \mathbf{R}^d$ dans $\mathbf{R}$ lesquels sont des noyaux d'après vous
a) $exp(-\frac{\|x-y\|^2}{2}) + exp(-\frac{\|x-y\|^2}{4})$ (car somme de noyaux donc positive aussi)
b) $< x, y > exp(-\frac{\|x-y\|^2}{2})$
c) $< x, y >$ (car $\iint < x, y > f(x)f(y)dxdy = \iint < xf(x), yf(y) > dxdy$ $=< \int xf(x)dx, \int yf(y)dy >\geq 0$)

Q22 : Citer deux propriétés géométriques importantes des points $\phi(x_i)$ d'un échantillon, transformés par l'immersion ϕ associé au noyau $exp(-\frac{\|x-y\|^2}{2})$ par la relation $< \phi(x), \phi(y) >= exp(-\frac{\|x-y\|^2}{2})$
i) sur une sphère de rayon 1 ii) de corrélations positives donc dans un orthant

Exercice 1 [3 points]
Soit $(x_i, y_i)_{i\in[\![1,n]\!]}$ un échantillon de $\mathbf{R}^d$ et soit ϕ_σ l'immersion associée au noyau $K_\sigma(x, y) = exp(-\frac{\|x-y\|^2}{2\sigma^2})$ càd $< \phi_\sigma(x), \phi_\sigma(y) >= exp(-\frac{\|x-y\|^2}{2\sigma^2})$
a) exprimer en fonction des $K_\sigma(x_i, x_j)$ et des α_i la quantité $\|\sum_{i=1}^{i=n} \alpha_i \phi_\sigma(x_i)\|$
réponse : $(\alpha'[K(x_i, x_j)]\alpha)^{\frac{1}{2}}$
b) que vaut la fonction $f(x) = \lim_{\sigma\to 0} \sum_{i=1}^{i=n} K_\sigma(x_i, x)$
réponse : en supposant les x_i distints : 1 si $x = x_i$ zéro sinon
c) que vaut $\lim_{\sigma\to 0} d(\phi_\sigma(x_i), \phi_\sigma(x_j))$

réponse : à la limite les vecteurs sont orthogonaux et de norme 1 donc la distance limite est de $\sqrt{2}$

Exercice 2 [2 points]
Montrer que pour un estimateur f_n qui vérifie $R_n(f_n) = \min_{f \in \mathcal{F}} R_n(f)$ on a :
$E[R_n(f_n)] \leq R(f_n)$
réponse : voir cours v3 page 7