Master M280 - Mido 9th March 2018

Exam : Machine Learning in Finance [1] : Time 1h30

Exercice 1. [14]*pt*

Q1 : In supervised learning what hypothesis do we make on the learning sample $(X^i, Y^i)_{i \in [\![1,n]\!]}$

a) the (X^i, Y^i) have all the same laws

b) the (X^i, Y^i) are independent

c) the X^i have all a normal distribution

Answers : a,b

Q2 : when we have a learning sample $(X^i, Y^i)_{i \in [\![1,n]\!]}$ and try to estimate f such that $Y = f(X)$

a) we are in the framework of supervised learning

b) we are in the framework of unsupervised learning

c) we are either solving a classification or regression problem

Answers : a,c

Q3 : when we have a learning sample $(X^i, Y^i)_{i \in [\![1,n]\!]}$ how is the calibration error the most likely to be defined in a classification problem

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

b) $E[\frac{1}{n} \sum\limits_{i=1}^{i=n} 1_{f(X_i) \neq Y_i}]$

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

Answers : c

Q4 : which one of these relationships is true for the classification error $R_n(f_n)$ and the prediction error $R(f_n)$

a) $R(f_n) < R_n(f_n)$

b) $E(R(f_n)) \geq R_n(f_n)$

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

d) $R(f_n) \geq E[R_n(f_n)]$

Answers : d

Q5 : which one of these expressions is the correct Vapnik Chervonenkis formula

a) $P\left(R(f_n) > R_n(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \leq \eta$

b) $P\left(R_n(f_n) > R(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \leq \eta$

c) $P\left(R_n(f_n) > R(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \leq \frac{\eta}{n}$

a) $P\left(R(f_n) > R_n(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \leq \frac{\eta}{n}$

1. Pierre Brugière Université Paris 9 Dauphine

Answers : a

Q6 : following the principles of Structural Risk Minimisation with nested ensemble of classifiers $\mathcal{F}_1 \subset \mathcal{F}_2 \cdots \subset \mathcal{F}_k \cdots$ what would be the basis to pick the classifier $f_{n,k}$?
Answers : that it minimizes over k, $R_n(f_{n,k}) + \phi_{n,\eta}(\frac{VC(\mathcal{F}_k)}{n})$

Q7) in $\mathbf{R}^d$ under which condition(s) can we be sure that we can separate with an hyperplane the null point 0 from the points $x_1, x_2, \cdots x_n$
a) if $x_1, x_2, \cdots x_n$ are independant
b) if $x_2 - x_1, x_3 - x_1, \cdots x_n - x_1$ are independent
c) if $x_2 - x_1, x_3 - x_1, \cdots x_n - x_1$ are dependent
Answers : a)

Q8) which one of these assertions is (are) true for two sets of vectors $(x_i)_{i \in I}$ and $(x_i)_{i \in J}$ of $\mathbf{R}^d$?
a) the two sets of vectors can be separated by an hyperplane if and only if the two convexe envelopes of these two sets can be separated by an hyperplane
b) it may be possible that the two sets can be separated by an hyperplane but not their convexe envelope
c) if may be possible that the two convexe envelopes have a nul intersection but cannot be separated by an hyperplane
d) if the two sets of points can be separated by an hyperplane there is in fact an infinity of hyperplanes that can separate them
Answers : a, d)

Q9) What is the Vapnik dimension of the family of hyperplane classifiers of $\mathbf{R}^d$?
a) d
b) $d + 1$
c) $+\infty$
Answers : b)

Q10) If a family of classifiers $\mathcal{F}$ is defined by d parameters $d \geq 1$ which proposition(s) are true
a) $VC(\mathcal{F}) < d + 2$
b) $VC(\mathcal{F}) = d + 1$
c) there may be some cases for which $VC(\mathcal{F}) = +\infty$
Answers : c)

Q11) what is the distance between the two hyperplanes of equations :
$\langle w, x \rangle + b = 0$ and $\langle -w, x \rangle + c = 0$
a) $\frac{|b-c|}{\|w\|^2}$
b) $\frac{|b+c|}{\|w\|}$

c) $\frac{|b-c|}{\|w\|}$
Answers : b)

Q12) which assertion(s) is\are true in $\mathbf{R}^2$
a) the VC dimension of GAP tolerant classifiers of radius 1 and margin 1.8 is 3
b) the VC dimension of GAP tolerant classifiers of radius 1 and margin 1 is 3
c) the VC dimension of GAP tolerant classifiers of radius 1 and margin 1.8 is 2
d) the VC dimension of GAP tolerant classifiers of radius 1 and margin 1 is 2
Answers : b, c)

Q13) which formula is true for a family $\mathcal{F}$ of GAP tolerant classifiers of radius R
a) $VC(\mathcal{F}_{\Delta,D}) \leq 1 + Min(\frac{D}{\Delta}, d)$
b) $VC(\mathcal{F}_{\Delta,D}) \leq 1 + Min(\frac{\Delta}{D}, d)$
c) $VC(\mathcal{F}_{\Delta,D}) \leq 1 + Min(\frac{D^2}{\Delta^2}, d)$
d) $VC(\mathcal{F}_{\Delta,D}) \leq 1 + Min(\frac{\Delta^2}{D^2}, d)$
Answers : c)

Q14) write the expression of the optimisation problem of classification for a C-SCM

Answers : $(P_C) \begin{cases} \min\limits_{w,b,\{\xi_i\}_{i\in[\![1,n]\!]}} \|w\|^2 + C\sum\limits_{i=1}^{i=n}\xi_i \\ \forall (x_i, y_i) \in \mathcal{S},\, y_i[\langle w, x_i\rangle + b] \geq 1 - \xi_i \\ \forall i \in [\![1,n]\!], \xi_i \geq 0 \end{cases}$

Q15) which of these inequalities is correct
a) $\max\limits_{z\in\mathcal{Z}}\left[\min\limits_{y\in\mathcal{Y}} g(y,z)\right] \leq \min\limits_{y\in\mathcal{Y}}\left[\max\limits_{z\in\mathcal{Z}} g(y,z)\right]$
b) $\max\limits_{z\in\mathcal{Z}}\left[\min\limits_{y\in\mathcal{Y}} g(y,z)\right] \geq \min\limits_{y\in\mathcal{Y}}\left[\max\limits_{z\in\mathcal{Z}} g(y,z)\right]$
Answers : a)

Q16) which of the following assertions are true :
a) if the KKT conditions are satisfied the primal and dual problems have the same value
b) if the primal and dual problems have the same value the KKT conditions are satisfied
c) in the SVMs problems we studied the KKT conditions may not be satisfied
Answers : a,b

Q17) which of the following assertions are true :
a) $\langle x, y\rangle^4 + \langle x, y\rangle^2$ is a Kernel
b) $\langle x, y\rangle^4 - \langle x, y\rangle^2$ is a Kernel
c) $exp(-\|x-y\|_d^2)$ is a Kernel
Answers : a,c

Q18) if $\phi_\sigma(.)$ is the transformation linked to the Kernel $K(x,y) = exp(-\frac{\|x-y\|_d^2}{2\sigma^2})$
a) what is the value of $\|\phi_\sigma(x)\|$?
b) is is true that $\forall x, y$ $\|\phi_\sigma(x) - \phi_\sigma(y)\| \leq \sqrt{2}$
c) is is true that $\forall x, y$ the angle between $\phi_\sigma(x)$ and $\phi_\sigma(y)$ is less than 90^o
Answers : 1, yes, yes

[3pt]Exercise
Let $\{(x_i, y_i)\}_{i\epsilon\{1,2,..,l\}}$ be a learning sample with $x_i \in \mathbf{R}^d$ and $y_i\epsilon\{-1,1\}$. We consider for $\mu > 0$ the optimisation problem (P_μ) :

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

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

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

[0.5pt]2) explain why for the solution of (P_μ) we have $\rho \geq 0$
Answers : if $\rho < 0$ was a solution the value of the problem would be strictly positive which is not the case as we can reach the value 0 by taking w, b and ρ equal to zero

[1pt]3) calculate : $\frac{\partial L}{\partial w}$, $\frac{\partial L}{\partial b}$, $\frac{\partial L}{\partial \rho}$, $\frac{\partial L}{\partial \zeta_i}$

Answers :
$$\frac{\partial L}{\partial w} = \left[\frac{\partial L}{\partial w^1}, \frac{\partial L}{\partial w^2} \cdots, \frac{\partial L}{\partial w^l}\right] = w' - \sum_{i=1}^{i=l} \alpha_i y_i x_i'$$
$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{i=l} \alpha_i y_i,$$
$$\frac{\partial L}{\partial \rho} = -2 + \sum_{i=1}^{i=l} \alpha_i,$$
$$\frac{\partial L}{\partial \zeta_i} = \mu - \alpha_i - \beta_i$$

[1pt]4) write the dual problem (D_μ) of (P_μ)
Answers :
$$\max_{\alpha_i,\beta_i} -\frac{1}{2}\|\sum_{i=1}^{i=l} \alpha_i y_i x_i\|^2,$$

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

[3pt]Exercise

Let $\{x_i\}_{i\in[\![1,n^+]\!]}$ be n^+ vectors of $S^1_{\mathbf{N}}$ labelled 1 and $\{z_j\}_{j\in[\![1,n^-]\!]}$ be n^- vectors of $S^1_{\mathbf{N}}$ labelled -1. We assume here that the $\{x_i, z_j\}$ form a family of orthogonal vectors.

[1.5pt]1) show that any hyperplane of margin Δ which separates the $\{x_i\}$ from the $\{z_j\}$ satisfies $\Delta \leq \sqrt{\frac{1}{n^-}+\frac{1}{n^+}}$

Answers : $w^+ = \frac{1}{n^+}\sum_{i=1}^{i=n^+} x_i$ belongs to the convex envelope of the $\{x_i\}$ and $w^- = \frac{1}{n^-}\sum_{j=1}^{j=n^-} z_j$ belongs to the convex envelope of the $\{z_j\}$.

As the maximum margin is the distance between the two convex envelopes we have : $\Delta \leq MaxMargin = d(\mathcal{C}_x, \mathcal{C}_z) \leq d(w^+, w^-) = \sqrt{\frac{1}{n^-}+\frac{1}{n^+}}$ which proves the result

[1.5pt] 2) find an hyperplane of margin Δ which separates the $\{x_i\}$ from the $\{z_j\}$ which satisfies $\Delta = \sqrt{\frac{1}{n^-}+\frac{1}{n^+}}$

Answers :

Let $w = \frac{1}{n^+}\sum_{i=1}^{i=n^+} x_i - \frac{1}{n^-}\sum_{j=1}^{j=n^-} z_j$

As the vectors are orthogonal we have :

$\forall x_i,\ \langle w, x_i\rangle = \frac{1}{n^+}$ and $\forall z_j,\ \langle w, z_j\rangle = -\frac{1}{n^-}$

so, $H_{w,-\frac{1}{n^+}}$ and $H_{w,\frac{1}{n^-}}$ separate the points. and

$d(H_{w,-\frac{1}{n^+}}, H_{w,\frac{1}{n^-}}) = \frac{|-\frac{1}{n^+}-\frac{1}{n^-}|}{\|w\|} = \sqrt{\frac{1}{n^-}+\frac{1}{n^+}}$