Master Masef - Mido 2017-2018

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

Exercice 1. [10]*pt*

1. In supervised learning which hypothesis is made 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

2. for 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

3. 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

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

5. if we find k points of $\mathbf{R}^d$ which can be classified in all possible ways by the family $\mathcal{C}_1$ of classifiers and not by the family $\mathcal{C}_2$, does it mean automatically that $VC(\mathcal{C}_1) > VC(\mathcal{C}_2)$?

1. Pierre Brugière Université Paris 9 Dauphine

Answers : no

6. what is the geometric configuration of $d+1$ points of a sphere of $\mathbf{R}^d$ which enables to classify them in all possible ways with the maximum margin
Answers : a simplex

7. in $\mathbf{R}^d$ under which condition(s) can we for sure 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)

8. 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)

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

10. 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)

11. 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)

12. 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)

13. in a C-SVM if x_i is a support vector which is not well classified what is the value of α_i
Answers : c

14. which of these inequalities is correct
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) $\max_{z \in \mathcal{Z}} \left[\min_{y \in \mathcal{Y}} g(y, z) \right] \geq \min_{y \in \mathcal{Y}} \left[\max_{z \in \mathcal{Z}} g(y, z) \right]$
Answers : a)

15. 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

16. 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

17. 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 strictly less than 90^o
Answers : 1, yes, yes (assuming $x \neq y$)

18. if $\{x_i\}_{i\in[\![1,n]\!]}$ is a family of orthormal vectors of $\mathbf{R}^d$ with $n^+ > 0$ vectors labelled $\{1\}$ and $n^- > 0$ vectors labelled $\{-1\}$ with what maximal margin can we separate the two classes
Answers : $\sqrt{\frac{1}{n^+} + \frac{1}{n^-}}$

19. gives the expression of $\frac{\partial L}{\partial w}$ when $L(w) = \frac{1}{2}\|w\|^2 - \alpha\langle w, x\rangle$
Answers : $w - \alpha x$

20. if $\{x_i\}_{i\in[\![1,n]\!]}$ is a family of orthormal vectors of the affine space $\mathbf{R}^d$ what is the minimum distance between an hyperplane which separates these points from the origin and the origin
Answers : $\frac{1}{\sqrt{d}}$

Exercise 1. [5pt]

Let $\langle . \rangle$ be defined on $l_2(\mathbf{R})$ by $\langle x, y\rangle = \sum\limits_{i=1}^{+\infty} x^i y^i$ where x^i and y^i are the components of x and y

Let $\|.\|_2$ be the corresponding norm defined by $\|x\|_2 = \big(\sum\limits_{i=1}^{+\infty}(x^i)^2\big)^{\frac{1}{2}}$

For $\omega \in l_2(\mathbf{R})$ we define $f_\omega(.)$ on $l_2(\mathbf{R})$ by

$$\begin{cases} f_\omega(x) = 1 \text{ if } \langle\omega, x\rangle \geq 1 \\ f_\omega(x) = -1 \text{ if } \langle\omega, x\rangle \leq -1 \\ f_\omega(x) = 0 \text{ if } -1 < \langle\omega, x\rangle < 1 \end{cases}$$

and we define a classification linked to $f_\omega(.)$ by :
x is classified in class $\{1\}$ iff $f_\omega(x) = 1$
x is classified in class $\{-1\}$ iff $f_\omega(x) = -1$
x is not classified by $f_\omega(.)$ iff $f_\omega(x) = 0$

For $\Delta > 0$ we defined the family $\mathcal{F}_\Delta$ of classifiers of $l_2(\mathbf{R})$ by :
$\mathcal{F}_\Delta = \{f_\omega(.), \|\omega\| \leq \frac{2}{\Delta}\}$

Let $\mathcal{B}_1 = \{x \in l_2(\mathbf{R}), \|x\| \leq 1\}$
Let $(x_i)_{i\in[\![1,d]\!]}$ be a family of d distinct vectors in $\mathcal{B}_1$
We make the assumption that the $(x_i)_{i\in[\![1,d]\!]}$ can be classified perfectly (no point unclassified and all points with their correct classification) in all possible ways by the classifiers of $\mathcal{F}_\Delta$.

1. **[0.5pt]** how many different ways are there to label the $(x_i)_{i\in[\![1,d]\!]}$ with the labels $\{-1, 1\}$?

Answer : 2^d

2. **[1pt]** if $y = (y^i)_{i\in[\![1,d]\!]} \in \{-1,1\}^d$ is a labelling of the $(x_i)_{i\in[\![1,d]\!]}$ and $f_{\omega^y}(.)$ is a classifier of $\mathcal{F}_\Delta$ which classifies all the $(x_i)_{i\in[\![1,d]\!]}$ perfectly, show that the classification is done with a margin of at least Δ (we assume there is at least one x_i in each class)

   Answer : $f_{\omega^y}(.)$ classifies all the x_i so for all x_i either $\langle \omega^y, x_i\rangle \geq 1$ or $\langle \omega^y, x_i\rangle \leq -1$ which means that the hyperplanes $\langle \omega^y, x_i\rangle = 1$ and $\langle \omega^y, x_i\rangle = -1$ separate the two classes of points defined by the labels y. The distance between the two hyperplanes is $\frac{2}{\|\omega^y\|} \geq \Delta$ Q.E.D

3. **[1.5pt]** if $y = (y^i)_{i\in[\![1,d]\!]} \in \{-1,1\}^d$ is a labelling of the $(x_i)_{i\in[\![1,d]\!]}$ show that $\|\sum_{i=1}^{i=d} y^i x_i\| \geq d\frac{\Delta}{2}$

   Answer : Let $f_{\omega^y}(.)$ classifies all the x_i with the labelling y correctly then $\forall i,\ y^i\langle \omega^y, x_i\rangle \geq 1 \implies \sum_{i=1}^{i=d} y^i\langle \omega^y, x_i\rangle \geq d \implies \langle \omega^y, \sum_{i=1}^{i=d} y^i x_i\rangle \geq d$

   $\implies \|\omega^y\|\|\sum_{i=1}^{i=d} y^i x_i\| \geq d$ (Cauchy-Schwarz)

   $\implies \frac{2}{\Delta}\|\sum_{i=1}^{i=d} y^i x_i\| \geq d \implies \|\sum_{i=1}^{i=d} y^i x_i\| \geq d\frac{\Delta}{2}$ Q.E.D

4. **[1.5pt]** Show that

   $\forall y = (y^i)_{i\in[\![1,d]\!]} \in \{-1,1\}^d,\ \|\sum_{i=1}^{i=d} y^i x_i\| \geq d\frac{\Delta}{2} \implies \sum_{i=1}^{i=d} \|x_i\|^2 \geq (\frac{d\Delta}{2})^2$

   Answer : Let Y^i be independent and be defined by $P(Y^i = 1) = 0.5$ and $P(Y^i = -1) = 0.5$. According to the hypothesis $\|\sum_{i=1}^{i=d} Y^i x_i\| \geq d\frac{\Delta}{2}$ so $\|\sum_{i=1}^{i=d} Y^i x_i\|^2 \geq d^2\frac{\Delta^2}{4}$ and so $\sum_{i=1}^{i=d}\sum_{i=1}^{i=d} Y^i Y^j\langle x_i, x_j\rangle \geq d^2\frac{\Delta^2}{4}$
   taking expectations we get
   $E[Y^i Y^j\langle x_i x_j\rangle] = 0$ if $i \neq j$ and $E[Y^i Y^i\langle x_i, x_i\rangle] = \|x_i\|^2$ leading to the result

5. **[0.5pt]** Deduct from what precedes a majorant for $VC(\mathcal{F}_\Delta)$

   Answer : $\sum_{i=1}^{i=d} \|x_i\|^2 \geq (\frac{d\Delta}{2})^2 \implies d \geq (\frac{d\Delta}{2})^2 \implies d \leq \frac{4}{\Delta^2}$
   so $VC(\mathcal{F}_\Delta) \leq \frac{4}{\Delta^2}$

Exercice 2. [5pt]
We consider a family of $d+1$ points $(M_i)_{i\in[\![1,d]\!]}$ of $\mathbf{R}^d$ which forms a simplex, i.e $\exists C > 0, \forall i \neq j\ d(M_i, M_j) = C$.
We note G the barycentre of the $(M_i)_{i\in[\![1,d]\!]}$ and $(x_i)_{i\in[\![1,d]\!]}$ the vectors of $\mathbf{R}^d$ defined by $x_i = \overrightarrow{GM_i}$.
Let $\phi_\sigma(.)$ be the transformation associated to the kernel $K_\sigma(x,y) = exp(-\frac{\|x-y\|^2}{2\sigma^2})$ and $\psi_\sigma(.)$ the transformation defined by $\psi_\sigma(M) = \phi_\sigma(\overrightarrow{GM})$

1. **[0.5pt]**show that $\forall i \in [\![1, d+1]\!]$, $\|\psi_\sigma(M)\| = 1$.
Answer : $\|\psi_\sigma(M)\|^2 = \|\phi_\sigma(x)\|^2 = \langle \phi_\sigma(x), \phi_\sigma(x)\rangle = exp(0) = 1$
2. **[0.5pt]**show that $\exists C_1 > 0, \forall i \neq j$, $\|\psi_\sigma(M_i) - \psi_\sigma(M_j)\| = C_1$
Answer : $\|\psi_\sigma(M_i) - \psi_\sigma(M_j)\|^2 = \|\phi_\sigma(x_i) - \phi_\sigma(x_j)\|^2 = \|\phi_\sigma(x_i)\|^2 + \|\phi_\sigma(x_j)\|^2 - 2\langle \phi_\sigma(x_i), \phi_\sigma(x_j)\rangle = 2 - 2exp(-\frac{d(M_i,M_j)^2}{2\sigma^2})$ Q.E.D
3. **[2pt]**show that if H is an hyperplane of $l_2(\mathbf{R})$ which separates the $\psi_\sigma(M_i)$ from the null vector 0 then $d(0,H) \leq \sqrt{g(d+1,C)}$ with
$g(d+1, C) = \frac{1}{d+1} + \frac{d}{d+1}exp(-\frac{C^2}{2\sigma^2})$
Answer : Let H be an hyperplane and $\langle w, z_i\rangle + b = 0$ its equation we can take w pointing out and thus
$\forall i \in [\![1, d]\!]$ $\langle w, \phi_\sigma(x_i)\rangle + b \geq 0$
$\implies \langle w, \frac{1}{d+1}\sum_{i=1}^{d+1} \phi_\sigma(x_i)\rangle + b \geq 0$ (by averaging the $d+1$ equations)
$\implies \|w\|\| \frac{1}{d+1}\sum_{i=1}^{d+1} \phi_\sigma(x_i)\| \geq |b|$ (by Cauchy Schwartz)
we use that $\| \frac{1}{d+1}\sum_{i=1}^{d+1} \phi_\sigma(x_i)\|^2 = \sum_{i=1}^{d+1} \|\phi_\sigma(x_i)\|^2 + \sum_{i\neq j}\langle \phi_\sigma(x_i)\phi_\sigma(x_j)\rangle$
$= \frac{1}{d+1} + \frac{d}{d+1}exp(-\frac{C^2}{2\sigma^2})$ which implies that $\|w\| \geq \frac{|b|}{\sqrt{g(d+1,C)}}$
as $d(0, H) = \frac{|b|}{\|w\|}$ we get the result
4. **[1pt]** find explicitly an hyperplane H^* which separates the $\phi_\sigma(x_i)$ from 0 with maximum margin
Answer : Let $w^* = \frac{1}{d+1}\sum_{i=1}^{d+1} \phi_\sigma(x_i)$. We have seen previously that for any hyperplane which separates the points from 0 we have $d(0, H) \geq \|w^*\|$
Now by symmetry and averaging the relations
$\forall j \in [\![1, d+1]\!]$, $\langle w^*, \phi_\sigma(x_j)\rangle = \langle w^*, \frac{1}{d+1}\sum_{j=1}^{d+1} \phi_\sigma(x_j)\rangle = \langle w^*, w^*\rangle > 0$
so the hyperplane H^* of equation $\langle w^*, \phi_\sigma(x_j)\rangle = \|w^*\|^2$ separates the points from 0 and as max margin as $d(0, H^*) = \frac{\|w^*\|^2}{\|w^*\|} = \|w^*\|$. [Q.E.D]
5. **[1pt]** Let $\mathcal{F}$ be the curve in $\mathbf{R}^d$ which corresponds to the hyperplane H^*. Show that if $\sigma < \frac{C}{2\sqrt{2ln(d+1)}}$ the curve $\mathcal{F}$ is made of several distinct components
Answer : $\mathcal{F} = \{x \in \mathbf{R}^d, \frac{1}{d+1}\sum_{i=1}^{d+1} exp(-\frac{\|x_i-x\|^2}{2\sigma^2}) = \frac{1}{d+1} + \frac{d}{d+1}exp(-\frac{C^2}{2\sigma^2})\}$
$= \{x \in \mathbf{R}^d, \sum_{i=1}^{d+1} exp(-\frac{\|x_i-x\|^2}{2\sigma^2}) = 1 + dexp(-\frac{C^2}{2\sigma^2})\}$
Let's show that $\sigma < \frac{C}{4\sqrt{ln(d+1)}} \implies \mathcal{F} \subset \cup\mathcal{B}(x_i, \frac{C}{2})$ (strict inclusion) and as $\cup\mathcal{B}(x_i, \frac{C}{2})$ is not connexe this will show that $\mathcal{F}$ is made of several distinct components
when $\sigma < \frac{C}{2\sqrt{2ln(d+1)}}$

$x \notin \cup\mathcal{B}(x_i, \frac{C}{2}) \Longrightarrow \sum_{i=1}^{d+1} exp(-\frac{\|x_i - x\|^2}{2\sigma^2}) < (1+d)exp(-\frac{C^2}{8\sigma^2}) < 1 \Longrightarrow x \notin \mathcal{F}$
so in conclusion $\sigma < \frac{C}{2\sqrt{2ln(d+1)}} \Longrightarrow \mathcal{F} \subset \cup\mathcal{B}(x_i, \frac{C}{2}) \Longrightarrow \mathcal{F}$ is made of several distinct components Q.E.D

ps by symmetry there is either one single or $d+1$ connexe components and when there is one single connexe component by symmetry the barycentre x_G must be inside so the limit case for one single component is when $(d+1)exp(-\frac{\|x_i - x_G\|^2}{2\sigma^2}) = 1 + dexp(-\frac{C^2}{2\sigma^2})$ (and it is not too difficult to express $\|x_i - x_G\|$ as a function of C) and thus get the value σ.