Master Masef - Mido 2018-2019

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

Exercise 1 (QCM) : [10pts]

1. If $(X_i, Y_i)_{i\in[\![1,n]\!]}$ is a learning sample, how is the calibration error generally defined for a classification problem ?
a) $\frac{1}{n}\sum_{i=1}^{i=n} 1_{f(X_i)\neq Y_i}$
b) $\frac{1}{n}\sum_{i=1}^{i=n} |f(X_i) - Y_i|$
c) $E[|f(X_{n+1}) - Y_{n+1}|]$
d) none of the answers above

2. If $(X_i, Y_i)_{i\in[\![1,n]\!]}$ is a learning sample, which hypothesis is always made on the variables ?
a) the X_i are all Gaussian
b) the (X_i, Y_i) have all the same laws
c) the X_i have all the same laws but not necessarily the Y_i
d) the (X_i, Y_i) are all independent
e) none of the answers above

3. If $R_n(f_n)$ is the calibration error for the optimal classifier f_n obtained for the learning sample $(X_i, Y_i)_{i\in[\![1,n]\!]}$ and if we note $R(f_n) = E[1_{f(X_{n+1})\neq Y_{n+1}}]$ the prediction error, which properties are always verified ?
a) $R_n(f_n) = R(f_n)$
b) $R_n(f_n) \leq R(f_n)$
c) $E(R_n(f_n)) \leq R(f_n)$
d) none of the answers above

4. If there are k particular points in $\mathbb{R}^d$ which can be classified in all possible ways by the family of classifiers $\mathcal{F}_1$ and if it is not possible to classify them in all possible ways by the family of classifiers $\mathcal{F}_2$ then which assertions is/are necessarily true :
a)$VC(\mathcal{F}_1) = k$
b)$VC(\mathcal{F}_2) < k$
c)$VC(\mathcal{F}_1) > VC(\mathcal{F}_2)$
d)none of the answers above

5. If for two families of classifiers/machines $\mathcal{F}_1$ and $\mathcal{F}_2$ the error of calibration is the same on a learning sample, which machine does the Vapnik Chernovenkis inequality and SRM principle encourage to use to predict :

1. Pierre Brugière University Paris 9 Dauphine

a) $\mathcal{F}_1$ if $VC(\mathcal{F}_1) > VC(\mathcal{F}_2)$
b) not $\mathcal{F}_2$ if $VC(\mathcal{F}_1) > VC(\mathcal{F}_2)$
c) none of the answers above

6. If $x_1, x_2 \cdots x_n$ form a family of independent vectors, and if 0 is the null vector, in how many different ways is it possible to classify : $0, x_1, x_2 \cdots x_n$?
a) 2^n
b) $n-1$
c) 2^{n+1}
d) none of the answers above

7. The inequality of Vapnik Chervonenkis enables, knowing the complexity of the family of classifiers/machine used and the error on calibration to :
a) define a confidence interval for $R(f_n)$
b) calculate the exact value of $R(f_n)$
c) define a boundary $\epsilon < 1$ for $R(f_n)$
d) none of the answers above

8. Which assertions are true :
a) a high VC for a family of classifiers implies necessarily a bad quality of prediction
b) an infinite VC for a family of classifiers implies always a perfect calibration
c) the VC for a parametric family of classifiers is always close to the number of parameters of the family of classifiers
d) none of the answers above

9. Which assertions is/are true ?
a) It is unlikely that a machine which calibrates badly will predict accurately
b) VC gives no guarantee that a complex machine which calibrates well will predict accurately
c) SRM means Structural Risk Maximization
d) a machine which is very complex in a VC sense predicts always very accurately

10. In $\mathbb{R}^d$ for $w \neq 0$ let $H_{w,b} = \{x \in \mathbb{R}^d, \langle w, x\rangle + b = 0\}$. Which assertions are true :
a) $\forall x \in \mathbb{R}^d \ d(x, H_{w,b}) = \frac{|\langle w,x\rangle + b|}{\|w\|^2}$
b) $\forall x \in \mathbb{R}^d \ d(x, H_{w,b}) = \frac{|\langle w,x\rangle + b|}{\|w\|}$
c) $\forall x \in \mathbb{R}^d \ d(x, H_{w,b}) = \frac{|\langle w,x\rangle - b|}{\|w\|}$
d) none of the answers above

11. Let $w \in \mathbb{R}^d \setminus \{0\}$. Let $H_{w,b} = \{(x,y) \in \mathbb{R}^{d+1}, \langle w, x\rangle + b - y = 0\}$. Which assertions are true :
a) $d(H_{w,b_1}, H_{w,b_2}) = \frac{|b_2+b_1|}{\|w\|_d}$
b) $d(H_{w,b_1}, H_{w,b_2}) = \frac{|b_2-b_1|}{\sqrt{1+\|w\|_d^2}}$
c) $d(H_{w,b_1}, H_{w,b_2}) = \frac{|b_2-b_1|}{\|w\|_d}$
d) none of the answers above

12. If you can classify perfectly a learning sample with the hyperplanes $H_{w_1,b_1} = \{x \in \mathbb{R}^d, \langle w_1, x\rangle + b_1 = 0\}$ of margin Δ_1 and $H_{w_2,b_2} = \{x \in \mathbb{R}^d, \langle w_2, x\rangle + b_2 = 0\}$ of margin Δ_2 you have a good reason to choose H_{w_1,b_1} if :
a) $\Delta_1 < \Delta_2$
b) $\Delta_1 > \Delta_2$
c) $b_1 > b_2$
d) $b_2 > b_1$

13. If you can classify perfectly a learning sample with two classes with an hyperplane H of margin Δ and if we call $\mathcal{C}_1$ and $\mathcal{C}_2$ the convex envelopes of the two classes of points then :
a) for sure $d(\mathcal{C}_1, \mathcal{C}_2) \geq \Delta$
b) for sure $d(\mathcal{C}_1, \mathcal{C}_2) > \Delta$
c) H of maximum margin $\Longleftrightarrow \big(H \cap \mathcal{C}_1 \neq \emptyset$ and $H \cap \mathcal{C}_2 \neq \emptyset\big)$
d) none of the answers above

14. If $\mathcal{F}$ is a family of classifiers of $\mathbb{R}^{10^6}$ of diameters 1 of hyperplanes of margin 0.1 then :
a) $VC(\mathcal{F}) = 10^6$
b) $VC(\mathcal{F}) = 10^6 + 1$
c) $VC(\mathcal{F}) \approx 100$
d) $VC(\mathcal{F}) \approx 10$

15. According to the minimax theorem for any function $g : \mathcal{Y} \times \mathcal{Z} \longrightarrow \mathbb{R}$
a) $\inf_{y\in\mathcal{Y}} \left[\sup_{z\in\mathcal{Z}} g(y,z)\right] \leq \sup_{z\in\mathcal{Z}} \left[\inf_{y\in\mathcal{Y}} g(y,z)\right]$
b) $\sup_{z\in\mathcal{Z}} \left[\inf_{y\in\mathcal{Y}} g(y,z)\right] \leq \inf_{y\in\mathcal{Y}} \left[\sup_{z\in\mathcal{Z}} g(y,z)\right]$
c) $\sup_{z\in\mathcal{Z}} \left[\inf_{y\in\mathcal{Y}} g(y,z)\right] = \inf_{y\in\mathcal{Y}} \left[\sup_{z\in\mathcal{Z}} g(y,z)\right]$
d) none of the answers above

16. When solving a SVM for a classification $\{-1, 1\}$ which assertions are true :
a) The Primal and Dual problems have the same solution if and only if

the KKT conditions are added to the constraints of the dual problem
b) The Primal and Dual problems have the same solution because of the particular nature of the problem
c) The KKT conditions are automatically satisfied because of the particular nature of the problem
d) none of the answers above

17. When solving a C-SVM for a classification $\{-1, 1\}$ which assertions are true :
a) all the support vectors are necessarily classified correctly
b) some support vectors may be classified incorrectly
c) the support vectors are necessarily on the border of the maximum margin hyperplane
d) if $0 < \alpha_i < C$, x_i is a support vector on the margin of the maximum margin hyperplane

18. Among the following functions from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}$ which ones are Kernel
a) $exp(-\frac{\|x-y\|^2}{2}) + exp(-\frac{\|x-y\|^2}{4})$
b) $exp(-\|x\| - \|y\|)$
c) $< x, y >$

19. If $\phi(.)$ is the transformation associated to the kernel $K(x, y) = exp(-\frac{\|x-y\|^2}{2})$ by the relationship $< \phi(x), \phi(y) >= K(x, y)$ among these properties which ones are true for the image points $\phi(x_i)$ from a learning sample.
a) the $\phi(x_i)$ are necessarily on a sphere of radius 1
b) the $\phi(x_i)$ for sure can be separated from 0 by an hyperplane
c) the $\phi(x_i)$ for sure are independent if the x_i are distinct

20. If $K(., .)$ is a kernel with values in $\mathbb{R}$ and $(H, \langle ., . \rangle_{RK})$ is a Reproducing Hilbert Space for K then which of the following assertions are true :
a) for all $f \in H$, $\langle K(x, .), f \rangle_{RK} = f(x)$
b) $K(x_1, x_2)^2 \leq K(x_1, x_1)K(x_1, x_2)$
c) for all $f \in H, f(x)^2 \leq K(x, x)\langle f, f \rangle_{RK}$

Exercise : [4pts]

We consider the following Ridge Regression problem :

$$(P)\ \ \arg\min_{\beta\in\mathbb{R}^d} \|Y - X\beta\|_n^2 + \lambda\|\beta\|_d^2$$

where Y is a vector of $\mathbb{R}^n$, X is a $n \times d$ matrix and $\|.\|$ is the euclidean norm. We remark that (P) can be written in the form :

$$(Q)\begin{cases} \arg\min_{z\in\mathbb{R}^n,\beta\in\mathbb{R}^d} \|z\|_n^2 + \lambda\|\beta\|_d^2 \\ z = Y - X\beta \end{cases}$$

1. **[0.5pt]** Determine the expression of the Lagrangian $L(z, \beta, \gamma)$ of (Q) with $\gamma \in \mathbb{R}^n$.
2. **[0.5pt]** Calculate $\frac{\partial L}{\partial z}$ and $\frac{\partial L}{\partial \beta}$
3. **[1pt]** Write the dual formulation (D) of (Q)
4. **[1pt]** Solve (D) by determining its solution γ^* and calculate as well the corresponding β^* it produces.
5. **[1pt]** Is β^* the expression you expected knowing that the solution of (P) is known to be $(X'X + \lambda Id_d)^{-1}X'Y$? Comment.

Exercise : [8.5pts]

Let $\{(x_i, y_i)\}_{i\in[\![1,n]\!]}$ be a learning sample with $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$. Let $\epsilon \in \mathbb{R}^{+*}$. We consider the problem :

$$(P_\epsilon)\begin{cases} \arg\min_{w\in\mathbb{R}^d,b\in\mathbb{R},\xi\in\mathbb{R}^n,\tilde{\xi}\in\mathbb{R}^n} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \tilde{\xi}_i) \\ y_i - \langle w, x_i\rangle - b \le \epsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \le \epsilon + \tilde{\xi}_i \\ \xi_i \ge 0 \\ \tilde{\xi}_i \ge 0 \end{cases}$$

1. **[1pt]** Write the Lagrangian $L(w, b, \xi, \tilde{\xi}, \alpha, \tilde{\alpha}, \mu, \tilde{\mu})$ of the problem (P_ϵ)
2. **[1pt]** Calculate :
   (a) $\frac{\partial L}{\partial w}$
   (b) $\frac{\partial L}{\partial b}$
   (c) $\frac{\partial L}{\partial \xi_i}$
   (d) $\frac{\partial L}{\partial \tilde{\xi}_i}$
3. **[1pt]** Show that the dual (D_ϵ) of (P_ϵ) is of the form

$$(D_\epsilon)\begin{cases} \arg\max_{\alpha\in\mathbb{R}^n,\tilde{\alpha}\in\mathbb{R}^n} -\frac{1}{2}(\alpha - \tilde{\alpha})'K(\alpha - \tilde{\alpha}) + \sum_{i=1}^{i=n}[\alpha_i(y_i - \epsilon) - \tilde{\alpha}_i(y_i + \epsilon)] \\ 0 \le \alpha_i \le C \\ 0 \le \tilde{\alpha}_i \le C \end{cases}$$

(a) **[0.25pt]** what is the expression of K ?

(b) **[0.25pt]** say, without demonstration, under what conditions on the $\{(x_i)\}_{i\in[\![1,n]\!]}$ K is invertible.

(c) **[0.50pt]** express the solution w^* of (P_ϵ) as a function of the solutions α^* and $\tilde{\alpha}^*$ of (D_ϵ) and of the x_i.

4. Due to the form of the problem, we assume in all what follows that the KKT conditions are satisfied.

(a) **[0.50pt]** write the KKT conditions

(b) **[0.50pt]** show that $\alpha_i\tilde{\alpha}_i = 0$

(c) **[0.50pt]** show that $\xi_i\tilde{\xi}_i = 0$

5. We call "support vectors" the x_i which appears with a non zero coefficient in the expression of w^*.

(a) **[0.5pt]** show that, x_i is a support vector if an only if :
$\alpha_i \neq 0$ or $\tilde{\alpha}_i \neq 0$

(b) **[0.5pt]** show that, $0 < \alpha_i < C$ or $0 < \tilde{\alpha}_i < C$ implies :
$|\langle w, x_i\rangle + b - y_i| = \epsilon$

(c) **[0.5pt]** show that, $\alpha_i = C$ or $\tilde{\alpha}_i = C$ implies $|\langle w, x_i\rangle + b - y_i| \geq \epsilon$

(d) **[0.5pt]** conclude on the location of the points (x_i, y_i) for the support vectors x_i.

6. **[1pt]** What is the aim of the problem (P_ϵ) ? is there a geometric interpretation in $\mathbb{R}^{d+1}$? and what is the interest to have $\|w^*\|$ small ?