ISF App 28th January 2021

Exam : Introduction to Machine Learning [1] : Time 1h30

Exercise 1. [10pts]

Select the correct answers

1. By assumption, the $(X^i, Y^i)_{i\in[\![1,n]\!]}$ in the learning sample, in supervised learning are such :
   a) the (X^i, Y^i) have all the same laws and are independent
   b) the (X^i, Y^i) have the same laws but may be dependent
   c) the X^i have all a normal distribution
   Answers : a

2. if $(X^i)_{i\in[\![1,n]\!]}$ is a learning sample without any labelling.
   a) we are in the framework of supervised learning
   b) we are in the framework of unsupervised learning
   c) generally we try to derive a structure for the X^i
   Answers : b,c

3. if $R_n(f_n)$ is the classification error and $R(f_n)$ is the prediction error then
   a) $R(f_n) < R_n(f_n)$
   b) $E(R(f_n)) \geq R_n(f_n)$
   c) $R(f_n) \geq E[R_n(f_n)]$
   d) $E(R(f_n)) \leq R_n(f_n)$
   Answers : c

4. in Vapnik Chervonenkis's formula
   a) $P\left(R(f_n) \leq R_n(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \geq 1-\eta$
   b) $P\left(R_n(f_n) \leq R(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \geq 1-\eta$
   c) $P\left(R_n(f_n) \leq R(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \geq 1-\frac{\eta}{n}$
   a) $P\left(R(f_n) \leq R_n(f_n) + \phi_{n,\eta}\left(\frac{VC(\mathcal{F})}{n}\right)\right) \geq 1-\frac{\eta}{n}$
   Answers : a

5. in the Vapnik Chervonenkis's formula
   a) $\phi_{n,\eta}(t)$ is a function which is decreasing with t
   b) $\phi_{n,\eta}(t)$ is a function which is increasing with t
   c) $\phi_{n,\eta}(\frac{1}{n})$ tends to zero when n tends to ∞
   Answers : b,c

1. Pierre Brugière University Paris Dauphine PSL

6. in Structural Risk Minimisation with $\mathcal{F}_1 \subset \mathcal{F}_2 \cdots \subset \mathcal{F}_k \cdots$ the classifier $f_{n,k}$ from class $\mathcal{F}_k$ is chosen if
a) it minimizes over k, $R_n(f_{n,k}) + \phi_{n,\eta}(\frac{VC(\mathcal{F}_k)}{n})$
b) it minimizes over k, $R_n(f_{n,k})$
c) it minimizes over k, $R_n(f_{n,k}) - \phi_{n,\eta}(\frac{VC(\mathcal{F}_k)}{n})$
Answers : a

7. if $x_1, x_2, \cdots x_n$ is a family of independent vectors then
a) $x_1, x_2, \cdots x_n$ can be classified in all possible ways with hyperplanes
b) 0 can be separated from $x_1, x_2, \cdots x_n$ with a hyperplane
c) 0 can be separated from the convex envelope of $x_1, x_2, \cdots x_n$ with a hyperplane
Answers : a,b,c

8. if a family of classifiers $\mathcal{F}$ is defined by d parameters $d \geq 1$ then
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

9. what is the distance between the point represented by the vector z and the hyperplane of equation $\langle w, x \rangle = b$
a) $\frac{|b+\langle w,z \rangle|}{\|w\|^2}$
b) $\frac{|\langle w,z \rangle - b|}{\|w\|}$
c) $\frac{|\langle w,z \rangle - b|}{\|w\|^2}$
Answers : b

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

11. in $\mathbf{R}^d$
a) the VC dimension of hyperplane classifiers of margin 1 is $d + 1$
b) the VC dimension of hyperplane classifiers within a ball of radius 1 is $d + 1$
c) the VC dimension of hyperplane classifiers of radius 1 and margin 0.1 is less than $d + 1$ when d is large
Answers : a,b,c

12. in a C-SVM
a) all the support vectors are always on the borders of the separating hyperplanes
b) it is possible that some support vectors are not correctly classified
c) it is impossible that some support vectors lay between the two separating hyperplanes
Answers : b

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

14. 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) the KKT conditions enable to calculate some of the parameters of the separating hyperplanes
Answers : a,b,c

15. which of the following assertions are true for the Gaussian Kernel K_σ and associated transformation ϕ_σ :
a) all the transformed points $\phi_\sigma(x_i)$ are on sphere or radius 1
b) for any sample $\{(x_i, y_i)\}_{i\in[\![1,n]\!]}$ it is possible to find σ such that the classification with the Gaussian Kernel K_σ is perfect
c) a notion of margin can be associated to Gaussian Kernel classifiers
Answers : a,b,c

16. single class SVMs can be used
a) to estimate the support of a distribution
b) to define different clusters amongst a sample
c) to solve a regression problem
Answers : a,b

17. if $K(.,.)$ is a kernel with values in $\mathbf{R}$ and $(H, \langle .,. \rangle_{RK})$ is a Reproducing Kernel Hilbert Space for K then
a) for all $f \in H$, $\langle K(x,.), f\rangle_{RK} = f(x)$
b) $K(x_1,x_2)^2 \leq K(x_1,x_1)K(x_2,x_2)$
c) for all $f \in H, f(x)^2 \leq K(x,x)\langle f, f\rangle_{RK}$
Answers : a,b,c

18. in the discount factor curve construction problem studied in the course
a) the regularisation term is linked to a scalar product on the function space
b) the regularisation term transforms a non parametric problem into a parametric problem
c) without the regularisation term the problem would not be well defined
Answers : a,b,c

19. in the discount factor curve construction problem studied in the course
a) the spline functions are of degrees 2
b) the spline functions have knots on the cash flow dates of the instruments used as inputs
c) the coefficients of the spline functions are the solutions of a ridge regression problem
d) the coefficients of the spline functions are the solutions of a Lasso regression problem
Answers : b,c

20. in the Smith and Nelson model the regularisation term is of the form
a) $\int f''^2(s)ds$
b) $\int f''^2(s) + \sigma^2 f'^2(s)ds$
c) $\int f''(s) + \sigma^2 f'(s)ds$
Answers : b

Problem [14pts]

Let $f(x, \lambda)$ be a function of (x, λ). (x^*, λ^*) is called a saddle point of f iff

$$\forall (x, \lambda), f(x^*, \lambda) \leq f(x^*, \lambda^*) \leq f(x, \lambda^*)$$

1. **[2pts]** show that $\sup_\lambda \inf_x f(x, \lambda) \leq \inf_x \sup_\lambda f(x, \lambda)$
Proof : $\forall (x, \lambda), f(x, \lambda) \leq \sup_\lambda f(x, \lambda)$
$\Longrightarrow \forall \lambda, \inf_x f(x, \lambda) \leq \inf_x \sup_\lambda f(x, \lambda)$
$\Longrightarrow \sup_\lambda \inf_x f(x, \lambda) \leq \inf_x \sup_\lambda f(x, \lambda)$ Q.E.D.

2. **[2pts]** show that if a saddle point (x^*, λ^*) exists then

$$\sup_\lambda \inf_x f(x, \lambda) = \inf_x \sup_\lambda f(x, \lambda)$$

Proof : $\inf_x \sup_\lambda f(x, \lambda) \leq \sup_\lambda f(x^*, \lambda)$ and using the assumptions $\sup_\lambda f(x^*, \lambda) \leq \inf_x f(x, \lambda^*) \leq \sup_\lambda \inf_x f(x, \lambda)$ we get,
$\inf_x \sup_\lambda f(x, \lambda) \leq \sup_\lambda \inf_x f(x, \lambda)$ which is the second part of the inequality Q.E.D.

3. **[2pts]** using 2., show that if a saddle point (x^*, λ^*) exists then

$$f(x^*, \lambda^*) = \inf_x \sup_\lambda f(x, \lambda) = \inf_x f(x, \lambda^*) = \sup_\lambda f(x^*, \lambda)$$

Proof : $\forall (x, \lambda)$,

$$\inf_x \sup_\lambda f(x, \lambda) \leq \sup_\lambda f(x^*, \lambda) \leq f(x^*, \lambda^*) \leq \inf_x f(x, \lambda^*) \leq \sup_\lambda \inf_x f(x, \lambda)$$

and as the two extreme terms are equal from the question 2. then all terms are equal Q.E.D.

In this second part which is mostly independent from the first part, we consider the problem $\begin{cases} \min\limits_{x \in \mathbf{R}^n} \|x\|^2 \\ Ax = p \end{cases}$
where A is full row rank, and (as a consequence) $p \in \mathrm{Im}(A)$

4. **[1pt]** for λ fixed, find $x(\lambda)$, as a function of A and λ which satisfies

$$\mathcal{L}(x(\lambda), \lambda) = \inf_x \mathcal{L}(x, \lambda)$$

Proof : The function is strictly convex in x so the point where the derivative cancels is the minimum and
$\frac{\partial \mathcal{L}}{\partial x}(x, \lambda) = 0 \iff 2x' - \lambda' = 0 \iff x(\lambda) = \frac{1}{2}A'\lambda$

5. **[1pt]** calculate $\mathcal{L}(x(\lambda), \lambda)$ as a function of A, λ and p
Proof : $\mathcal{L}(x(\lambda), \lambda) = \|\frac{1}{2}A'\lambda\|^2 - \frac{1}{2}\langle \lambda, AA'\lambda \rangle + \langle \lambda, p \rangle = -\frac{1}{4}\|A'\lambda\|^2 + \langle \lambda, p \rangle$

6. **[2pts]** find λ^*, as a function of A and λ which satisfies

$$\mathcal{L}(x(\lambda^*), \lambda^*) = \sup_\lambda \mathcal{L}(x(\lambda), \lambda)$$

Proof : let $\phi(\lambda) = \mathcal{L}(x(\lambda), \lambda)$. ϕ is strictly concave in λ so the point of minimum is the point where the derivative cancels.
Now, $\frac{\partial \phi}{\partial \lambda}(\lambda) = -\frac{1}{2}\lambda' AA' + p'$ so $\frac{\partial \phi}{\partial \lambda}(\lambda) = 0 \iff \lambda' AA' = 2p'$.
As A has full raw rank AA' is invertible and AA' is symmetric and so is its inverse, therefore,
$\frac{\partial \phi}{\partial \lambda}(\lambda) = 0 \iff \lambda' = 2p'(AA')^{-1} \iff \lambda^* = 2(AA')^{-1}p$

7. **[1pt]** verify that $Ax(\lambda^*) = p$
Proof : $Ax(\lambda^*) = A\frac{1}{2}A' 2(AA')^{-1}p = AA'(AA')^{-1}p = p$

8. **[1pt]** show that

$$\mathcal{L}(x(\lambda^*), \lambda^*) = \sup_\lambda \mathcal{L}(x(\lambda^*), \lambda)$$

Proof : $\mathcal{L}(x(\lambda^*), \lambda) = \|x(\lambda^*)\|^2 - \langle \lambda, Ax(\lambda^*) - p \rangle = \|x(\lambda^*)\|^2$ as $Ax(\lambda^*) - p = 0$. So, $\sup\limits_\lambda \mathcal{L}(x(\lambda^*), \lambda) = \|x(\lambda^*)\|^2 = \mathcal{L}(x(\lambda^*), \lambda^*)$ which proves the result.

9. **[1pt]** show that $(x(\lambda^*), \lambda^*)$ is a saddle point for $\mathcal{L}$.
Proof :
first, $\mathcal{L}(x(\lambda^*), \lambda^*) = \sup_\lambda \mathcal{L}(x(\lambda^*), \lambda) \Longrightarrow \forall \lambda, \mathcal{L}(x(\lambda^*), \lambda) \leq \mathcal{L}(x(\lambda^*), \lambda^*)$
second, $\forall \lambda, \mathcal{L}(x(\lambda), \lambda) = \inf_x \mathcal{L}(x, \lambda) \Longrightarrow \mathcal{L}(x(\lambda^*), \lambda^*) = \inf_x \mathcal{L}(x, \lambda^*)$
$\Longrightarrow \forall x, \mathcal{L}(x(\lambda^*), \lambda^*) \leq \mathcal{L}(x, \lambda^*)$
so, $\forall (x, \lambda), \mathcal{L}(x(\lambda^*), \lambda) \leq \mathcal{L}(x(\lambda^*), \lambda^*) \leq \mathcal{L}(x, \lambda^*)$ which proves the result.