ÉCOLES NORMALES SUPÉRIEURES

CONCOURS D'ADMISSION 2014 **FILIÈRE MP SPÉCIALITÉ INFO**

COMPOSITION D'INFORMATIQUE-MATHÉMATIQUES – A – (ULCR)

(Durée : 4 heures)

L'utilisation des calculatrices **n'est pas autorisée** pour cette épreuve.
Le langage de programmation choisi par le candidat doit être spécifié en tête de la copie.

⋆ ⋆ ⋆

Dimension VC d'un hypergraphe

Ce sujet comporte six pages et quatre parties.

La partie I introduit la notion de dimension de Vapnik-Chervonenkis, en abrégé *dimension VC*, d'un hypergraphe et établit une borne polynomiale sur la complexité d'un hypergraphe de dimension VC bornée. La partie II étudie l'effet d'une méthode de normalisation d'hypergraphes sur la dimension VC. La partie III détermine la dimension VC de certains hypergraphes définis géométriquement en établissant au préalable quelques résultats classiques de géométrie convexe. La partie IV examine un algorithme d'approximation d'un hypergraphe géométrique de dimension VC non bornée.

Les notions introduites en partie I sont utilisées dans les parties II et III. La partie IV s'appuie sur certains résultats de la partie III. Les résultats d'une question pourront être admis dans la suite du sujet.

Pour tout ensemble S on note $P(S)$ l'ensemble des parties de S, c'est-à-dire :

$$P(S) = \{T \mid T \subseteq S\}.$$

Un *hypergraphe de sommets* S est un sous-ensemble de $P(S)$. Si H est un hypergraphe de sommets S alors on appelle *hyperarête* de H tout élément de H. Par exemple si $S = \{0, 1, 2\}$ alors $P(S) = \{\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$ et $H = \{\{0\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$ est un hypergraphe de sommets S ; l'ensemble $\{1, 2\}$ est une hyperarête de H.

Pour tout ensemble *fini* X on note $|X|$ son *cardinal*, c'est-à-dire son nombre d'éléments. En particulier, quand H est un hypergraphe, $|H|$ désigne le nombre d'hyperarêtes de H.

Tout au long du sujet, n et d désignent des entiers strictement positifs. Pour tous entiers $p \geq 0$ et $q \geq 0$ on note $\binom{p}{q}$ le nombre de sous-ensembles distincts de cardinal q d'un ensemble de cardinal p. En particulier $\binom{p}{0} = 1$ pour tout p.

Partie I. Dimension VC d'un hypergraphe

Question 1 Soit $S = \{0, 1, \ldots, n-1\}$ l'ensemble des entiers compris entre 0 et $n-1$.

a. Quel est, en fonction de n, le nombre maximum d'hyperarêtes que peut avoir un hypergraphe de sommets S ? On justifiera la réponse.

b. Combien d'hypergraphes de sommets S différents existe-t-il ? On justifiera la réponse.

c. Démontrer :

$$\forall p \geq q \geq 1, \qquad \binom{p}{q} = \binom{p-1}{q} + \binom{p-1}{q-1}.$$

Si H est un hypergraphe de sommets S et U est un sous-ensemble de S, on définit la *trace* de H sur U, notée $H_{|U}$, par :

$$H_{|U} = \{U \cap \alpha \mid \alpha \in H\}.$$

Question 2 Soient S un ensemble fini, H un hypergraphe de sommets S et $U \subseteq V \subseteq S$.

a. Donner un exemple de S, H, U et V pour lesquels $H_{|U} \not\subseteq H_{|V}$.

b. Démontrer que pour toutes hyperarêtes $\alpha, \beta \in H$, on a :

$$\alpha \cap U \neq \beta \cap U \quad \Rightarrow \quad \alpha \cap V \neq \beta \cap V.$$

c. En déduire que $|H_{|U}| \leq |H_{|V}|$.

Soit H un hypergraphe de sommets S. On dit qu'un sous-ensemble U de S est *pulvérisé* par H si $H_{|U} = P(U)$. Si S est fini, on définit la *dimension VC* de H, notée $\dim_{\mathrm{VC}}(H)$, comme le cardinal maximum d'un sous-ensemble $U \subseteq S$ pulvérisé par H, augmenté de 1 :

$$\dim_{\mathrm{VC}}(H) = 1 + \max\{|U| \mid U \subseteq S \text{ et } H_{|U} = P(U)\}.$$

On ne définit pas de notion de dimension VC pour les hypergraphes dont l'ensemble de sommets est infini.

Question 3 Dans cette question on suppose $n \geq 3$.

a. Soient S un ensemble de n réels deux à deux distincts et H l'hypergraphe de sommets S défini par $H = \{S \cap [a,b] \mid a, b \in \mathbb{R}\}$. Calculer $\dim_{\mathrm{VC}}(H)$.

b. On note $\mathbb{S} = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. Soit $S' \subseteq \mathbb{S}$ un sous-ensemble de $\mathbb{S}$ de cardinal n. On définit :

$$f : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{S} \\ \theta & \mapsto & (\cos\theta, \sin\theta) \end{array} \right. \qquad \text{et} \quad H' = \{S' \cap f([a,b]) \mid a, b \in \mathbb{R}\}.$$

H' est donc un hypergraphe de sommets S'. Calculer $\dim_{\mathrm{VC}}(H')$.

Question 4 Soient S un ensemble fini de cardinal n et H un hypergraphe de sommets S.

a. Montrer que si $\dim_{\mathrm{VC}}(H) = 1$ alors $|H| \leq 1$.

b. Soit $x \in S$. On définit deux hypergraphes de sommets $S \setminus \{x\}$:

$$H' = H_{|S\setminus\{x\}} \quad \text{et} \quad H'' = \{\alpha \in H' \mid \alpha \in H \text{ et } \alpha \cup \{x\} \in H\}.$$

Montrer que $|H| = |H'| + |H''|$.

c. On considère à nouveau les hypergraphes H' et H'' définis à la question 4(b). Montrer que $\dim_{\mathrm{VC}}(H') \leq \dim_{\mathrm{VC}}(H)$ et que $\dim_{\mathrm{VC}}(H'') \leq \dim_{\mathrm{VC}}(H) - 1$.

d. Montrer que si $\dim_{\mathrm{VC}}(H) = d$, alors :

$$|H| \leq \sum_{i=0}^{d-1} \binom{n}{i}.$$

e. Donner, pour tout $n \geq d$, un exemple d'hypergraphe à n sommets pour lequel l'inégalité de la question 4(d) devient une égalité. On justifiera la réponse.

Partie II. Compression d'hypergraphe et dimension VC

Soit H un hypergraphe de sommets S. Pour tout $x \in S$, on définit l'hypergraphe *décalé* de H par x, noté $D_x(H)$, par :

$$D_x(H) = \{\alpha \setminus \{x\} | \alpha \in H\} \cup \{\alpha | \alpha \in H \text{ et } \alpha \setminus \{x\} \in H\}.$$

Par exemple, si $S = \{0, 1, 2\}$ et $H = \{\{0\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$, alors :

$$D_0(H) = \{\emptyset, \{2\}, \{1, 2\}, \{0, 1, 2\}\}.$$

Dans le reste de cette partie on suppose que $S = \{0, 1, \ldots, n-1\}$ et que H est un hypergraphe de sommets S.

Question 5 Montrer que pour tout $x \in S$ on a $|D_x(H)| = |H|$. Indication : on pourra chercher à expliciter une bijection entre H et $D_x(H)$.

Question 6 Soient $U \subseteq S$ et $x \in S$.

a. Montrer que si $x \notin U$ alors $(D_x(H))_{|U} \subseteq D_x(H_{|U})$.

b. Montrer que si $x \in U$ alors $(D_x(H))_{|U} \subseteq D_x(H_{|U})$.

Question 7 Montrer que pour tout $x \in S$ on a $\dim_{\mathrm{VC}}(D_x(H)) \leq \dim_{\mathrm{VC}}(H)$.

Question 8 Pour tout $i \in \mathbb{N}$ on note $i \bmod n$ le reste de la division entière de i par n, c'est-à-dire que $i \bmod n$ est l'unique entier $r \in \{0, 1, \ldots, n-1\}$ tel qu'il existe $q \in \mathbb{N}$ satisfaisant $i = qn + r$. On considère la suite $(H_i)_{i \in \mathbb{N}}$ d'hypergraphes définie par :

$$\begin{cases} H_0 = H \\ H_i = D_{i \bmod n}(H_{i-1}) \text{ pour tout } i \geq 1 \end{cases}$$

a. Démontrer qu'il existe un entier $i_0 \in \mathbb{N}$ et un hypergraphe H^* de sommets S tel que :

$$\forall i \geq i_0, \qquad H_i = H^*.$$

b. Démontrer que $\max_{\alpha \in H^*} |\alpha| \leq \dim_{\mathrm{VC}}(H) - 1$.

c. En déduire une nouvelle preuve de l'identité établie à la question 4(d) :

$$|H| \leq \sum_{i=0}^{d-1} \binom{n}{i}, \qquad \text{où} \quad d = \dim_{\mathrm{VC}}(H).$$

Partie III. Dimension VC d'hypergraphes géométriques

On munit l'espace vectoriel $\mathbb{R}^d$ du produit scalaire usuel :

$$(x_1, x_2, \ldots, x_d) \cdot (y_1, y_2, \ldots, y_d) = \sum_{i=1}^{d} x_i y_i,$$

et on rappelle que $\mathbb{R}^d$ muni de ce produit scalaire est un espace euclidien. Un ensemble $A \subseteq \mathbb{R}^d$ est dit *convexe* si et seulement si :

$$\forall x, y \in A, \forall \lambda \in [0,1], \quad \lambda x + (1-\lambda) y \in A.$$

Par convention, l'ensemble vide est convexe. On définit l'*enveloppe convexe* d'un ensemble $A \subseteq \mathbb{R}^d$, notée $\mathrm{conv}(A)$, comme l'intersection de tous les sous-ensembles convexes de $\mathbb{R}^d$ qui contiennent A :

$$\mathrm{conv}(A) = \bigcap_{B \subseteq \mathbb{R}^d; A \subseteq B; B \text{ convexe}} B.$$

On note $\mathcal{C}_{\mathbf{d}}$ l'ensemble des sous-ensembles convexes de $\mathbb{R}^d$.

Question 9

a. Montrer que toute intersection de sous-ensembles convexes de $\mathbb{R}^d$ est convexe.

b. Démontrer que pour tout sous-ensemble $A \subseteq \mathbb{R}^d$ les trois propositions suivantes sont équivalentes :

(1) A est convexe,

(2) $A = \mathrm{conv}(A)$,

(3) $\forall n \geq 2, \forall x_1, x_2, \ldots, x_n \in A, \forall \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}^+$ tels que $\sum_{i=1}^n \lambda_i = 1$, $\sum_{i=1}^n \lambda_i x_i \in A$.

c. Démontrer que pour tout sous-ensemble $A \subseteq \mathbb{R}^d$ on a :

$$\mathrm{conv}(A) = \left\{ \sum_{i=1}^{n} \lambda_i x_i \mid n \in \mathbb{N}^*, x_1, x_2, \ldots, x_n \in A, \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}^+ \text{ tels que } \sum_{i=1}^{n} \lambda_i = 1 \right\}$$

Un sous-ensemble $D \subseteq \mathbb{R}^d$ est un *demi-espace fermé* (resp. *demi-espace ouvert*) s'il existe $c \in \mathbb{R}^d$ et $t \in \mathbb{R}$ tels que $D = \{x \in \mathbb{R}^d \mid x \cdot c \leq t\}$ (resp. $D = \{x \in \mathbb{R}^d \mid x \cdot c < t\}$). On note $\mathcal{D}_d$ l'ensemble des demi-espaces fermés de $\mathbb{R}^d$.

Question 10

a. Montrer que tout demi-espace fermé de $\mathbb{R}^d$ est convexe.

b. Montrer que pour tout sous-ensemble X de $\mathbb{R}^d$ et tout demi-espace fermé D de $\mathbb{R}^d$:

$$D \cap \mathrm{conv}(X) \neq \emptyset \quad \Leftrightarrow \quad D \cap X \neq \emptyset$$

Une *partition* d'un ensemble S en deux parties est un ensemble $\{I_1, I_2\}$ de sous-ensembles de S qui sont disjoints et dont l'union égale S : $I_1 \cap I_2 = \emptyset$ et $I_1 \cup I_2 = S$.

Question 11 Soient $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ des vecteurs deux à deux distincts.

a. Montrer que si $n \geq d+2$ alors il existe n réels $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ non tous nuls tels que $\sum_{i=1}^{n} \lambda_i x_i = \vec{0}$ et $\sum_{i=1}^{n} \lambda_i = 0$.

b. En déduire que si $n \geq d+2$ alors l'ensemble $\{x_1, x_2, \ldots, x_n\}$ admet une partition en deux sous-ensembles non-vides V_1 et V_2 tels que $\mathrm{conv}(V_1) \cap \mathrm{conv}(V_2)$ est non vide.

Question 12 Soit S un sous-ensemble fini de $\mathbb{R}^d$ et soit H l'hypergraphe de sommets S défini par $H = \{S \cap D \mid D \in \mathcal{D}_d\}$. Montrer que $\dim_{\mathrm{VC}}(H) \leq d+2$.

Question 13 Soit $n \geq 2$. Soit $S = \{x_1, x_2, \ldots, x_n\}$ où $x_1, x_2, \ldots, x_n \in \mathbb{R}^2$ sont des vecteurs de norme 1 deux à deux distincts. Calculer la dimension VC de l'hypergraphe H de sommets S défini par $H = \{S \cap C \mid C \in \mathcal{C}_2\}$. On justifiera la réponse.

Partie IV. Approximation d'une famille d'hypergraphes géométriques

Dans cette partie, ε désigne un réel strictement compris entre 0 et 1.

Soient $S \subseteq \mathbb{R}^d$ un ensemble de cardinal n et $\mathcal{R}$ un ensemble (éventuellement infini) de sous-ensembles de $\mathbb{R}^d$. Un ensemble $T \subset \mathbb{R}^d$ est *ε-fin* pour S relativement à $\mathcal{R}$ si

$$\forall R \in \mathcal{R} \text{ tel que } |R \cap S| \geq \varepsilon |S|, \quad R \cap T \neq \emptyset.$$

Les ensembles ε-fins jouent un rôle important dans les algorithmes d'approximation en géométrie algorithmique. Quand la dimension VC de l'hypergraphe $\{S \cap R \mid R \in \mathcal{R}\}$ est bornée indépendamment de $|S|$, il existe toujours un ensemble ε-fin de petite taille et des méthodes simples et générales permettent d'en calculer un efficacement. Dans le cas $\mathcal{R} = \mathcal{C}_d$, la dimension VC de cet hypergraphe n'est pas bornée indépendamment de $|S|$ (cf question 13) et il faut recourir à des méthodes spécialisées. L'objectif de cette partie est d'étudier l'une de ces méthodes.

Question 14 Soient $C_1, C_2, \dots, C_n \in \mathcal{C}_d$.

a. On suppose que $n \geq d+2$ et que pour tout sous-ensemble $I \subseteq \{1, 2, \dots n\}$ de cardinal $n-1$, l'ensemble $\cap_{i \in I} C_i$ est non vide. En déduire que l'ensemble $\cap_{i=1}^{n} C_i$ est non-vide.

> **Indication :** on pourra fixer, de manière arbitraire, pour tout $1 \leq k \leq n$, un vecteur $x_k \in \cap_{i \in \{1,2,\dots,n\} \setminus \{k\}} C_i$ et utiliser la question 11 (b).

b. Montrer que si $n \geq d+2$ et que pour tout sous-ensemble $I \subseteq \{1, 2, \dots n\}$ de cardinal $d+1$, l'ensemble $\cap_{i \in I} C_i$ est non vide, alors l'ensemble $\cap_{i=1}^{n} C_i$ est non-vide.

Question 15 Soit $S \subseteq \mathbb{R}^d$ un ensemble de cardinal n. Un *point central* de S est un élément $c \in \mathbb{R}^d$ (pas nécessairement dans S) tel que tout demi-espace fermé contenant c contient au moins $\frac{n}{d+1}$ éléments de S.

a. Montrer qu'un vecteur $c \in \mathbb{R}^d$ est un point central de S si et seulement si c appartient à tout demi-espace ouvert D contenant strictement plus de $\frac{dn}{d+1}$ vecteurs de S.

b. On définit

$$\mathcal{C}(S) = \left\{ \operatorname{conv}(D \cap S) \mid D \text{ demi-espace ouvert tel que } |D \cap S| > \frac{dn}{d+1} \right\}$$

c'est-à-dire que les éléments de $\mathcal{C}(S)$ sont les enveloppes convexes de sous-ensembles de S contenus dans un demi-espace ouvert D contenant strictement plus que $\frac{dn}{d+1}$ vecteurs de S. Montrer que pour tous $C_1, C_2, \dots, C_{d+1} \in \mathcal{C}(S)$ l'intersection $\cap_{i=1}^{d+1} C_i$ est non-vide.

c. Montrer que tout ensemble fini $S \subseteq \mathbb{R}^d$ de cardinal $n \geq d+1$ admet un point central.

Un principe général de construction d'un ensemble ε-fin relativement à $\mathcal{C}_2$ est le suivant :

$(\mathcal{M})$

Entrée : un sous-ensemble fini $S \subseteq \mathbb{R}^2$ et $0 < \varepsilon < 1$.
Sortie : $T \subseteq \mathbb{R}^2$ qui est ε-fin pour S relativement à $\mathcal{C}_2$.

$T \leftarrow \emptyset$
Tant qu'il existe $C \in \mathcal{C}_2$ tel que $|C \cap S| \geq \varepsilon |S|$ et $C \cap T = \emptyset$,
| Calculer un point central c de $C \cap S$
| Ajouter c à T
Retourner T.

Soit $S \subseteq \mathbb{R}^2$ un ensemble de cardinal n et soit c un point central de S. On suppose que $c \notin S$. Un *triangle de* S est un sous-ensemble de S de cardinal 3. On dit qu'un triangle T de S *entoure* un vecteur x si $x \in \operatorname{conv}(T)$. La *profondeur simpliciale* d'un vecteur x par rapport à S est le nombre de triangles de S qui entourent x.

Question 16 On admet le résultat suivant.

> **Lemme de sélection** : si $S \subseteq \mathbb{R}^2$ est de cardinal n et c est un point central de S alors la profondeur simpliciale de c par rapport à S est supérieure ou égale à $\frac{2}{9}\binom{n}{3}$.

Montrer que la méthode $(\mathcal{M})$ termine et majorer le cardinal de l'ensemble T retourné.

* *
*