

TRAVAUX DIRIGÉS N° 2 : Concentration, théorie de VC

Stephan CLÉMENÇON <stephan.clemencon@telecom-paris.fr>
Ekhine IRUROZKI <irurozki@telecom-paris.fr>

EXERCICE 1. On se place dans le cadre de la classification binaire : soient un descripteur aléatoire X à valeurs dans un espace mesurable $\mathcal{X} \subset \mathbb{R}^d$ ($d \in \mathbb{N}^*$) et un label aléatoire Y valant -1 ou 1 . On considère une classe finie \mathcal{G} de classifieurs $\mathcal{X} \rightarrow \{-1, 1\}$ telle que les deux labels sont parfaitement séparables par un élément de \mathcal{G} , i.e. $\min_{g \in \mathcal{G}} L(g) = 0$ pour le risque $L : g \in \mathcal{G} \mapsto \mathbb{P}(g(X) \neq Y) \in [0, 1]$.

Soit $n \in \mathbb{N}^*$. On suppose que l'on dispose d'un échantillon i.i.d. $\{(X_i, Y_i)\}_{1 \leq i \leq n}$ suivant la même loi que (X, Y) et on note \hat{g}_n un minimiseur de l'erreur empirique de classification :

$$\hat{g}_n \in \min_{g \in \mathcal{G}} L_n(g) \quad \text{où} \quad L_n : g \in \mathcal{G} \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(X_i) \neq Y_i\}}.$$

- 1) Montrer que $\min_{g \in \mathcal{G}} L_n(g) = 0$ presque-sûrement.
- 2) Montrer que $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}|(1 - \epsilon)^n$ pour tout $\epsilon \in [0, 1]$.
En déduire que $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}|e^{-n\epsilon}$ pour tout $\epsilon > 0$.

Indication. Utiliser $\mathcal{G}_B := \{g \in \mathcal{G} : L(g) > \epsilon\}$ ainsi qu'une borne d'union.

- 3) Dédurre de la question précédente que $\mathbb{E}(L(\hat{g}_n)) \leq \frac{\log(e|\mathcal{G}|)}{n}$.

Indication. Pour toute variable aléatoire Z positive, $\mathbb{E}(Z) = \int_0^{+\infty} \mathbb{P}(Z > t) dt$.

EXERCICE 2. On se place dans le cadre de la classification binaire. On utilisera les mêmes notations que dans l'exercice précédent. On pose $L^* := L(g^*)$ avec $g^* : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta(x) \geq 1/2\}} - 1$ et on note $\eta : x \in \mathcal{X} \mapsto \mathbb{P}(Y = 1 | X = x) \in [0, 1]$ la fonction de régression. Soit $(\eta_n)_{n \in \mathbb{N}^*}$ une suite de fonctions définies sur \mathcal{X} à valeurs dans $]0, 1[$. Pour tout $n \in \mathbb{N}^*$ on considère le classifieur $g_n : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta_n(x) \geq 1/2\}} - 1$.

- 1) On suppose qu'il existe $\delta > 0$ tel que $|\eta(x) - 1/2| \geq \delta$ pour tout $x \in \mathcal{X}$. Montrer que

$$L(g_n) - L^* \leq \frac{2 \mathbb{E}((\eta_n(X) - \eta(X))^2)}{\delta}.$$

- 2) Montrer que si $L^* = 0$, alors quel que soit $q \in [1, +\infty[$

$$L(g_n) \leq 2^q \mathbb{E}(|\eta_n(X) - \eta(X)|^q).$$

Soient maintenant $\eta' : \mathcal{X} \rightarrow]0, 1[$ et $g : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta'(x) \geq 1/2\}} - 1$.

- 3) On suppose que $\mathbb{P}\{\eta'(X) = 1/2\} = 0$ et que $\mathbb{E}(|\eta_n(X) - \eta'(X)|) \rightarrow 0$ lorsque $n \rightarrow +\infty$. Montrer que $L(g_n) \rightarrow L(g)$ lorsque $n \rightarrow +\infty$.

- 4) On suppose que le label Y n'est plus observable, mais qu'une variable Z à valeurs dans $\{-1, +1\}$ l'est, telle que :

$$\begin{aligned}\mathbb{P}(Z = 1 \mid Y = -1, X) &= \mathbb{P}(Z = 1 \mid Y = -1) = a < 1/2, \\ \mathbb{P}(Z = -1 \mid Y = 1, X) &= \mathbb{P}(Z = -1 \mid Y = 1) = b < 1/2.\end{aligned}$$

On pose à présent $\eta' : x \in \mathcal{X} \mapsto \mathbb{P}(Z = +1 \mid X = x)$. Montrer que :

$$L(g) \leq L^* \left(1 + \frac{2|a-b|}{1-2\max(a,b)} \right).$$

Que peut-on en déduire lorsque $a = b$?

EXERCICE 3. Calculer la VC dimension des classes \mathcal{A} d'ensembles suivantes :

- 1) $\mathcal{A} = \{]-\infty, x_1] \times \dots \times]-\infty, x_d] : (x_1, \dots, x_d) \in \mathbb{R}^d \}$,
- 2) \mathcal{A} est constituée des rectangles de \mathbb{R}^d .

EXERCICE 4. Donner une borne supérieure de la VC dimension de la classe des boules fermées dans \mathbb{R}^d :

$$\mathcal{A} = \left\{ \left\{ x = (x_1, \dots, x_d) \in \mathbb{R}^d : \sum_{i=1}^d |x_i - a_i|^2 \leq b \right\} : a_1, \dots, a_d, b \in \mathbb{R} \right\}.$$

EXERCICE 5. Soit \mathcal{A} une classe d'ensembles de \mathbb{R}^d de VC dimension $V < +\infty$ et de coefficients d'éclatement $s(\mathcal{A}, n), \forall n \in \mathbb{N}^*$.

- 1) Montrer que : $\forall n \geq 1, s(\mathcal{A}, n) \leq (n+1)^V$.
- 2) Montrer que : $\forall n \geq V, s(\mathcal{A}, n) \leq (ne/V)^V$.

Indication. On utilisera le lemme de Sauer : $\forall n \geq 1, s(\mathcal{A}, n) \leq \sum_{k=0}^V \binom{n}{k}$.

Exercice 3, solution. Commençons par quelques rappels de cours. Soient \mathcal{A} une classe d'ensembles de \mathbb{R}^d et $n \in \mathbb{N}^*$.

Quels que soient $x_1, \dots, x_n \in \mathbb{R}^d$, on note

$$N_{\mathcal{A}}(x_1, \dots, x_n) := \text{Card} \{ \{x_1, \dots, x_n\} \cap A : A \in \mathcal{A} \}$$

le nombre d'ensembles différents que l'on peut former en intersectant $\{x_1, \dots, x_n\}$ avec les éléments de \mathcal{A} . Par construction, il y en a au plus 2^n , et si $N_{\mathcal{A}}(x_1, \dots, x_n) = 2^n$, on dit que \mathcal{A} pulvérise ou éclate $\{x_1, \dots, x_n\}$.

Le n -ème coefficient de pulvérisation ou d'éclatement de \mathcal{A} est alors défini comme la quantité

$$S_{\mathcal{A}}(n) := \max_{x_1, \dots, x_n \in \mathbb{R}^d} N_{\mathcal{A}}(x_1, \dots, x_n),$$

qui donne le plus grand nombre d'ensembles différents que l'on peut obtenir en intersectant les éléments de \mathcal{A} avec n'importe quel ensemble de $n \in \mathbb{N}^*$ éléments de \mathbb{R}^d . Par construction, il jouit des propriétés suivantes :

- (i) $S_{\mathcal{A}}(n) \in \llbracket 1, 2^n \rrbracket$ (on a toujours soit l'ensemble vide soit au moins un des n points considérés),
- (ii) $S_{\mathcal{A}}(n) = 2^n$ ssi il existe un sous-ensemble de n éléments de \mathbb{R}^d éclaté par \mathcal{A} ,
- (iii) si $S_{\mathcal{A}}(n) < 2^n$, alors pour tout $k \geq n$ on a aussi $S_{\mathcal{A}}(k) < 2^k$ (si \mathcal{A} ne peut pas éclater un ensemble à n éléments, alors elle ne peut pas éclater d'ensemble plus grand encore).

Les coefficients d'éclatement permettent ainsi de mesurer la richesse de la classe \mathcal{A} .

En cherchant l'entier n tel que à partir duquel $S_{\mathcal{A}}(n+k) < 2^{n+k}$ pour tout $k \in \mathbb{N}^*$, on définit la dimension de Vapnik-Chervonenkis (VC dimension) de la classe \mathcal{A} :

$$V_{\mathcal{A}} := \max\{n \in \mathbb{N}^* : S_{\mathcal{A}}(n) = 2^n\} \in \mathbb{N}^* \cup \{+\infty\}.$$

Elle donne le plus grand nombre de points que l'on peut éclater avec \mathcal{A} .

- 1) Quel que soit $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, notons $A_{\mathbf{x}} :=]-\infty, x_1] \times \dots \times]-\infty, x_d]$, de telle manière que $\mathcal{A} = \{A_{\mathbf{x}} : \mathbf{x} \in \mathbb{R}^d\}$. Nous allons montrer que $V_{\mathcal{A}} = d$. Pour cela, il faut d'abord trouver un ensemble de d points de \mathbb{R}^d éclatés par \mathcal{A} . Considérons les vecteurs e_1, \dots, e_d de la base canonique de \mathbb{R}^d , tels que pour tout $i \in \{1, \dots, d\}$, toutes les composantes du vecteur e_i valent 0, sauf la i -ème qui vaut 1.

On remarque que quels que soient $i \in \llbracket 1, d \rrbracket$ et $x \in \mathbb{R}^d$,

$$e_i \in A_{\mathbf{x}} \quad \text{ssi} \quad \begin{cases} x_1, \dots, x_n \geq 0 \text{ and} \\ x_i \geq 1. \end{cases}$$

Par conséquent, pour tout $x \in \mathbb{R}^d$, on a d'abord $\{e_1, \dots, e_d\} \cap A_{\mathbf{x}} = \emptyset$ ssi $x_1, \dots, x_n < 1$ puis pour tous $m \in \llbracket 1, d \rrbracket$ et $i_1, \dots, i_m \in \llbracket 1, d \rrbracket$,

$$\{e_1, \dots, e_d\} \cap A_{\mathbf{x}} = \{e_{i_1}, \dots, e_{i_m}\} \quad \text{ssi} \quad \begin{cases} \forall k \in \llbracket 1, d \rrbracket \setminus \{i_1, \dots, i_m\} : 0 \leq x_k < 1, \\ x_{i_1}, \dots, x_{i_m} \geq 1. \end{cases}$$

En d'autres termes, pour tout sous-ensemble $\{e_{i_1}, \dots, e_{i_m}\}$, on est capable de trouver au moins un $A_{\mathbf{x}} \in \mathcal{A}$ qui permet de l'isoler. Cela signifie exactement que \mathcal{A} éclate $\{e_1, \dots, e_d\}$, d'où $S_{\mathcal{A}}(d) = 2^d$. Il nous faut maintenant vérifier que \mathcal{A} ne peut pas éclater $d+1$ points. Pour cela, prenons $\mathbf{z}_1, \dots, \mathbf{z}_{d+1} \in \mathbb{R}^d$ quelconques. Quel que soit $i \in \llbracket 1, d \rrbracket$, notons $\mathbf{z}_k = (z_{k,1}, \dots, z_{k,d})$ et définissons \bar{z}_i comme l'élément de $\{\mathbf{z}_1, \dots, \mathbf{z}_{d+1}\}$ ayant la plus grande i -ème coordonnée, c'est-à-dire $\bar{z}_{i,i} = \max\{z_{1,i}, \dots, z_{d+1,i}\}$. Alors, quel que soit $\mathbf{x} \in \mathbb{R}^d$, on a

$$\begin{aligned} \{\bar{z}_1, \dots, \bar{z}_d\} \subset A_{\mathbf{x}} &\iff \forall i \in \llbracket 1, d \rrbracket, \bar{z}_{1,i}, \dots, \bar{z}_{d,i} \leq x_i \\ &\iff \forall i \in \llbracket 1, d \rrbracket, \bar{z}_{i,i} = \max\{z_{1,i}, \dots, z_{d+1,i}\} \leq x_i \\ &\iff \forall i \in \llbracket 1, d \rrbracket, z_{1,i}, \dots, z_{d+1,i} \leq x_i \\ &\iff \{\mathbf{z}_1, \dots, \mathbf{z}_{d+1}\} \subset A_{\mathbf{x}}. \end{aligned}$$

En d'autres termes, on ne peut pas trouver $A_{\mathbf{x}} \in \mathcal{A}$ qui isole le sous-ensemble de points $\{\bar{z}_1, \dots, \bar{z}_d\}$, d'où $S_{\mathcal{A}}(d+1) < 2^{d+1}$.

Par définition de la dimension VC, on en conclut que $V_{\mathcal{A}} = d$. La Figure 2 donne une illustration des résultats précédents quand $d = 2$.

Remarque. Pour avoir l'intuition de la démonstration, on peut regarder d'abord les cas $d = 1$ et $d = 2$, qui amènent à penser que $V_{\mathcal{A}} = d$ et donnent l'idée de travailler avec les maxima des coordonnées.

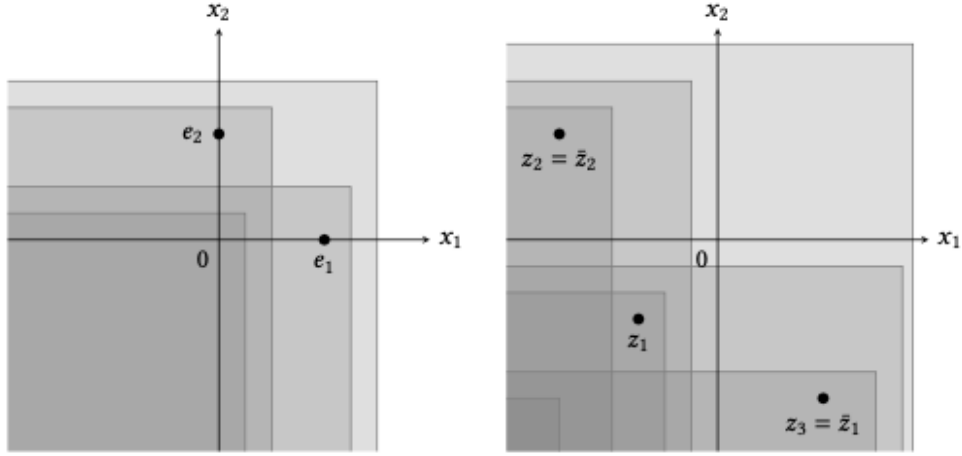


FIGURE 1 – Illustration of VC dimension for $d = 2$.

2) Quels que soient $\mathbf{x} = (x_1, \dots, x_d), \mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$, on note

$$\mathbf{x} \wedge \mathbf{y} := (x_1 \wedge y_1, \dots, x_d \wedge y_d) \quad \text{et} \quad \mathbf{x} \vee \mathbf{y} := (x_1 \vee y_1, \dots, x_d \vee y_d)$$

les minimum et maximum composante par composante. Lorsque $\mathbf{x} = \mathbf{x} \wedge \mathbf{y}$ et $\mathbf{y} = \mathbf{x} \vee \mathbf{y}$, on note

$$[\mathbf{x}, \mathbf{y}] := [x_1, y_1] \times \dots \times [x_d, y_d].$$

Nous allons montrer que $V_{\mathcal{A}} = 2d$.

Pour cela, il faut d'abord trouver un ensemble de $2d$ points de \mathbb{R}^d éclatés par \mathcal{A} . Considérons les vecteurs e_1, \dots, e_d de la base canonique de \mathbb{R}^d , comme à la question précédente, ainsi que leurs symétriques $-e_1, \dots, -e_d$. On remarque que pour tous $i \in \llbracket 1, d \rrbracket$ et $\varepsilon \in]0, 1[$,

$$\pm e_i \in [\pm e_i - \varepsilon, \pm e_i + \varepsilon].$$

Par conséquent, pour tout sous-ensemble de $\{-e_1, \dots, -e_d, e_1, \dots, e_d\}$, on peut trouver un hyper-rectangle qui l'isole des autres points : quel que soit $\varepsilon \in]0, 1[$,

$$(a) \quad \{-e_1, \dots, -e_d, e_1, \dots, e_d\} \cap [-\varepsilon, \varepsilon]^d = \emptyset,$$

$$(b) \quad \text{pour tous } m \in \llbracket 1, d \rrbracket \text{ et } i_1, \dots, i_m \in \llbracket 1, d \rrbracket,$$

$$\{-e_1, \dots, -e_d, e_1, \dots, e_d\} \cap \left[\bigwedge_{k=1}^m \pm e_{i_k} - \varepsilon, \bigvee_{k=1}^m \pm e_{i_k} + \varepsilon \right] = \{\pm e_{i_1}, \dots, \pm e_{i_m}\}.$$

On a donc bien $S_{\mathcal{A}}(2d) = 2^{2d}$.

Il nous faut maintenant vérifier que \mathcal{A} ne peut pas éclater $2d + 1$ points. Pour cela, prenons $\mathbf{z}_1, \dots, \mathbf{z}_{2d+1} \in \mathbb{R}^d$ quelconques. Quel que soit $i \in \{1, \dots, d\}$, définissons \bar{z}_i et \underline{z}_i comme les éléments de $\{\mathbf{z}_1, \dots, \mathbf{z}_{2d+1}\}$ ayant respectivement la plus grande et la plus petite i -ème coordonnée, i.e.

$$\bar{z}_{i,i} = \max\{z_{1,i}, \dots, z_{2d+1,i}\} \quad \text{et} \quad \underline{z}_{i,i} = \min\{z_{1,i}, \dots, z_{2d+1,i}\}.$$

Alors, quels que soient $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ tels que $\mathbf{x} = \mathbf{x} \wedge \mathbf{y}$ et $\mathbf{y} = \mathbf{x} \vee \mathbf{y}$, on a

$$\begin{aligned} \{\bar{z}_1, \dots, \bar{z}_d, \underline{z}_1, \dots, \underline{z}_d\} \subset [\mathbf{x}, \mathbf{y}] &\iff \forall i, k \in \llbracket 1, d \rrbracket, \bar{z}_{i,k}, \underline{z}_{i,k} \in [x_k, y_k] \\ &\iff \forall i \in \llbracket 1, d \rrbracket, \bar{z}_{i,i}, \underline{z}_{i,i} \in [x_i, y_i] \\ &\iff \forall i \in \llbracket 1, d \rrbracket, z_{1,i}, \dots, z_{2d+1,i} \in [x_i, y_i] \end{aligned}$$

$$\iff \{z_1, \dots, z_{2d+1}\} \subset [\mathbf{x}, \mathbf{y}].$$

En d'autres termes, on ne peut pas trouver d'hyper-rectangle qui isole le sous-ensemble de points $\{\bar{z}_1, \dots, \bar{z}_d, z_1, \dots, z_d\}$, d'où $S_{\mathcal{A}}(2d+1) < 2^{2d+1}$.

Par définition de la dimension VC, on en conclut que $V_{\mathcal{A}} = 2d$. La Figure 2 donne une illustration des résultats précédents quand $d = 2$.

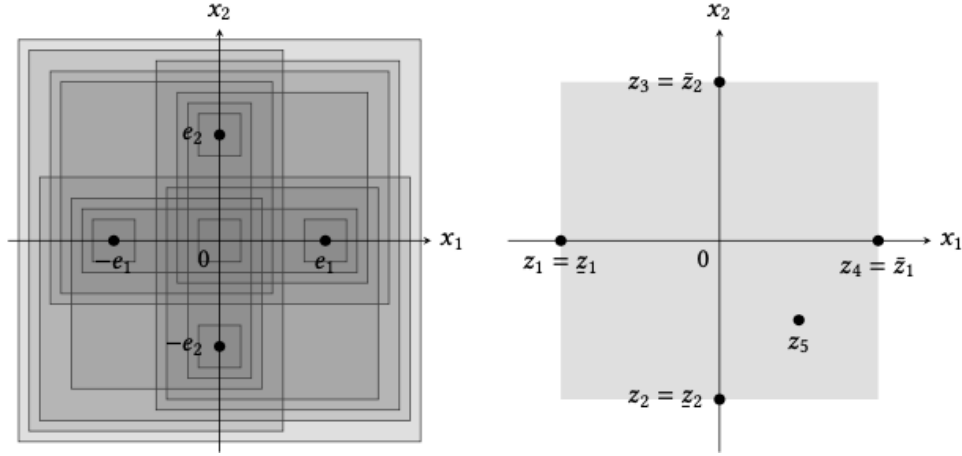


FIGURE 2 – Illustration of VC dimension for $d = 2$.

Exercice 4, solution. Pour répondre à la question posée, nous allons d'abord démontrer le théorème suivant.

Soit G un espace vectoriel de fonctions $\mathbb{R}^d \rightarrow \mathbb{R}$ ($d \in \mathbb{N}^*$), de dimension finie $m \in \mathbb{N}^*$. Alors la classe d'ensembles

$$\mathcal{A} := \{\{x \in \mathbb{R}^d : g(x) \geq 0\} : g \in G\}$$

a pour VC dimension $V_{\mathcal{A}} \leq m$.

Il suffit pour cela de prouver qu'aucun ensemble de $m+1$ points de \mathbb{R}^d ne peut être éclaté par \mathcal{A} . Prenons donc $x_1, \dots, x_{m+1} \in \mathbb{R}^d$ quelconques, puis considérons l'application linéaire

$$h : g \in G \mapsto (g(x_1), \dots, g(x_{m+1})) \in \mathbb{R}^{m+1}.$$

Puisque G est un espace vectoriel, $h(G)$ est un sous-espace vectoriel de \mathbb{R}^{m+1} , de dimension inférieure ou égale à celle de G , i.e. m . Il existe donc un vecteur non nul $\gamma = (\gamma_1, \dots, \gamma_{m+1}) \in \mathbb{R}^{m+1}$ orthogonal à $h(G)$, c'est-à-dire tel que pour tout $g \in G$, on ait

$$\langle \gamma, h(g) \rangle = \gamma_1 g(x_1) + \dots + \gamma_{m+1} g(x_{m+1}) = 0. \quad (4)$$

Soient $P := \{i \in \llbracket 1, m+1 \rrbracket : \gamma_i \geq 0\}$ et $N := \{i \in \llbracket 1, m+1 \rrbracket : \gamma_i < 0\}$, alors pour tout $g \in G$, on a

$$\sum_{i \in P} \gamma_i g(x_i) = - \sum_{i \in N} \gamma_i g(x_i). \quad (1)$$

Quitte à remplacer, si besoin, γ par $-\gamma$ (qui satisfait aussi l'Eq. 4), on peut supposer que $N \neq \emptyset$.

Supposons maintenant par l'absurde qu'il existe $g \in G$ telle que $\{x \in \mathbb{R}^d : g(x) \geq 0\}$ contient $\{x_i : i \in P\}$ mais pas $\{x_i : i \in N\}$. Alors la partie de gauche dans l'Eq. 1 est positive ou nulle, tandis que

celle de droite est forcément strictement négative, ce qui amène à une contradiction. Il n'existe donc pas d'élément de \mathcal{G} qui permette de séparer $\{x_i : i \in P\}$ et $\{x_i : i \in N\}$. Par conséquent, $\mathcal{S}_A(m+1) < 2^{m+1}$ puis $V_{\mathcal{A}} < m+1$, i.e. $V_{\mathcal{A}} \leq m$.

Revenons maintenant à la question posée. Quels que soient $a_1, \dots, a_d, b \in \mathbb{R}$ et $(x_1, \dots, x_d) \in \mathbb{R}^d$, on remarque que

$$\sum_{i=1}^d |x_i - a_i|^2 = \sum_{i=1}^d x_i^2 - 2 \sum_{i=1}^d a_i x_i + \left(\sum_{i=1}^d a_i^2 - b \right).$$

Ainsi, si l'on considère \mathcal{G} , l'espace vectoriel engendré par les fonctions $\varphi_1, \dots, \varphi_{d+2} : \mathbb{R}^d \rightarrow \mathbb{R}$ qui à tout $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ associent

$$\varphi_1(x) = \sum_{i=1}^d x_i^2, \quad \varphi_2(x) = x_1, \dots, \varphi_{d+1}(x) = x_d, \quad \varphi_{d+2}(x) = 1,$$

on a $\mathcal{A} \subset \{\{x \in \mathbb{R}^d : g(x) \leq 0\} : g \in \mathcal{G}\} =: \mathcal{B}$.

Or, comme \mathcal{G} est un espace vectoriel, $g \in \mathcal{G}$ ssi $-g \in \mathcal{G}$, et

$$\mathcal{B} = \{\{x \in \mathbb{R}^d : -g(x) \geq 0\} : g \in \mathcal{G}\} = \{\{x \in \mathbb{R}^d : g(x) \geq 0\} : g \in \mathcal{G}\}.$$

En appliquant le théorème que l'on vient de démontrer, on a donc $V_{\mathcal{A}} \leq V_{\mathcal{B}} \leq d+2$.

Exercice 5, solution.

1) Soit $n \in \mathbb{N}^*$, alors

$$\begin{aligned} (n+1)^V &= \sum_{k=0}^V \binom{V}{k} n^k = \sum_{k=0}^V \frac{V!}{(V-k)!k!} n^k = 1 + \sum_{k=1}^V \frac{(V-k+1) \cdots V}{k!} n^k \\ &\geq 1 + \sum_{k=1}^V \frac{n^k}{k!} = 1 + \sum_{k=1}^V \binom{n}{k} n^k \frac{(n-k)!}{n!} = 1 + \sum_{k=1}^V \binom{n}{k} \frac{n^k}{(n-k+1) \cdots n} \\ &\geq 1 + \sum_{k=1}^V \binom{n}{k} = \sum_{k=0}^V \binom{n}{k} \geq \mathcal{S}_{\mathcal{A}}(n) \end{aligned}$$

par le lemme de Sauer.

2) Soit un entier $n \geq V$, alors $1 \geq \frac{V}{n}$. On cherche à montrer que

$$\sum_{k=0}^V \binom{n}{k} \leq \left(\frac{ne}{V} \right)^V = \left(\frac{n}{V} \right)^V e^V$$

ou de manière équivalente que

$$\left(\frac{V}{n} \right)^V \sum_{k=0}^V \binom{n}{k} \leq e^V,$$

pour pouvoir ensuite appliquer le lemme de Sauer et conclure. Or

$$\left(\frac{n}{V} \right)^V \sum_{k=0}^n \binom{n}{k} \leq \sum_{k=0}^V \binom{n}{k} \left(\frac{V}{n} \right)^V \leq \sum_{k=0}^n \binom{n}{k} \left(\frac{V}{n} \right)^k = \left(\frac{V}{n} + 1 \right)^n \leq \left(e^{\frac{V}{n}} \right)^n = e^V.$$