Examen de Machine learning
Cycle pluridisciplinaire d'études supérieures
Université Paris sciences et lettres

mercredi 1er juin 2022

Les exercices sont indépendants. Toute affirmation devra être justifiée, à l'exception du premier exercice. Il n'est pas interdit d'admettre certains éléments de démonstration (voire des questions entières) afin de ne pas rester bloqué, mais cela doit être mentionné. Il n'est pas nécessaire de traiter l'ensemble du sujet pour obtenir une bonne note.

Exercice 1 (*Questions de cours*). — Répondre aux questions sans justification.

1) Donner la définition d'une classe de prédicteurs linéaires. Donner deux exemples d'algorithmes d'apprentissage donnant des prédicteurs linéaires.
2) Donner la définition de la minimisation du risque empirique régularisé. Donner un exemple.
3) Donner une définition informelle du sur-apprentissage et du sous-apprentissage.
4) Expliquer comment diminuer le sur-apprentissage avec l'algorithme donné en exemple à la question 2) ci-dessus.
5) Comment diminuer le sur-apprentissage lorsqu'on utilise l'algorithme des k plus proches voisins ?

Exercice 2 (*k plus proches voisins*). — On se place dans un cadre de classification avec l'ensemble d'entrées $\mathscr{X} = [0, 10]$ et l'ensemble de sorties $\mathscr{Y} = \{0, 1\}$. On considère sur $\mathscr{X}$ la distance :

$$\rho(x, x') = |x - x'| .$$

On dispose de l'échantillon d'apprentissage suivant :

$$S = ((0, 0), (3, 1), (4, 1), (6, 0), (9, 0)).$$

1) Représenter graphiquement les données d'apprentissage.

2) Pour $1 \leqslant k \leqslant 5$, on note $\hat{f}^{(k)}$ le prédicteur kNN (pour la distance ρ) construit avec S. Pour les valeurs $k \in \{1, 2, 5\}$, donner (sans justifier) l'expression de $\hat{f}^{(k)}$.

3) Pour chaque $k \in \{1, 2, 5\}$, calculer l'erreur d'apprentissage de $\hat{f}^{(k)}$.

Exercice 3 (*Détection de sous-mots*). — Soit Σ un ensemble fini appelé *alphabet*. Soit $d \geqslant 1$ un entier et on définit :

$$\mathcal{X}_d = \bigcup_{1 \leqslant k \leqslant d} \Sigma^d.$$

$\mathcal{X}_d$ est appelé ensemble des *mots* sur l'alphabet Σ de longueur au plus d. Pour chaque mot $v \in \mathcal{X}_d$, on note $|v|$ la *longueur* de v, autrement dit, l'entier tel que $v \in \Sigma^{|v|}$, et on note :

$$v = (v_1, \dots, v_{|v|}).$$

Pour $v, v' \in \mathcal{X}_d$, on dit que v est un *sous-mot* de v' s'il existe un entier $k \geqslant 0$ tel que :

$$v' = (v'_1, \dots, v'_k, v_1, \dots, v_{|v|}, v'_{k+|v|+1}, \dots, v'_{|v'|}).$$

On note alors $v \prec v'$.

On considère un problème d'apprentissage où l'ensemble d'entrées est $\mathcal{X}_d$ et l'ensemble de sorties $\mathcal{Y} = \{-1, 1\}$. Pour chaque mot $v \in \mathcal{X}_d$, on définit le prédicteur f_v par :

$$\forall x \in \mathcal{X}_d, \quad f_v(x) = \begin{cases} 1 & \text{si } v \prec x \\ -1 & \text{sinon.} \end{cases},$$

et on considère la classe de prédicteurs :

$$\mathcal{F} = \{f_v\}_{v \in \mathcal{X}_d}.$$

On définit l'application $\mathrm{K} \colon \mathcal{X}_d \times \mathcal{X}_d \to \mathbb{R}$ par :

$$\forall x, x' \in \mathcal{X}_d, \quad \mathrm{K}(x, x') = \mathrm{Card}\,\{v \in \mathcal{X}_d \mid v \prec x \text{ et } v \prec x'\},$$

où Card désigne le cardinal de l'ensemble.

1) Trouver un espace préhilbertien $(\tilde{\mathcal{X}}, \langle \cdot\,, \cdot \rangle)$ et une application $\psi \colon \mathcal{X}_d \to \tilde{\mathcal{X}}$ tels que K est le noyau associé, c'est-à-dire :

$$\forall x, x' \in \mathcal{X}_d, \quad \mathrm{K}(x, x') = \langle \psi(x), \psi(x') \rangle.$$

2) Montrer que pour tout $v \in \mathcal{X}_d$, il existe un prédicteur linéaire $\tilde{f}_v \colon \tilde{\mathcal{X}} \to \mathbb{R}$ tel que $f_v = \tilde{f}_v \circ \psi$.

Exercice 4 (*Caractérisation des noyaux*). — Pour tout entier $m \geqslant 1$, on rappelle qu'une matrice réelle M de taille $m \times m$ est dite *semi-définie positive* si pour tout $u \in \mathbb{R}^m$,

$$u^\top \mathrm{M} u \geqslant 0.$$

Soit $\mathcal{X}$ un ensemble quelconque.

1) On suppose dans cette question qu'on a $(\tilde{\mathscr{X}}, \langle \cdot\,, \cdot \rangle)$ espace préhilbertien, $\psi : \mathscr{X} \to \tilde{\mathscr{X}}$, et $\mathrm{K} : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ le noyau associé. Soit $x_1, \dots, x_m \in \mathscr{X}$. Montrer que la matrice :

$$\mathrm{G} = (\mathrm{K}(x_i, x_k))_{1 \leqslant i,k \leqslant m}$$

est symétrique semi-définie positive.

On souhaite à présent établir la réciproque. Soit $\mathrm{K} : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ une application telle que pour tout entier $m \geqslant 1$ et tous $x_1, \dots, x_m \in \mathscr{X}$, la matrice $(\mathrm{K}(x_i, x_k))_{1 \leqslant i,k \leqslant m}$ est symétrique semi-définie positive.

Le but de l'exercice est de montrer que K est un noyau, autrement dit qu'il existe une application de redescription à valeurs dans un espace préhilbertien dont le noyau associé est K.

Pour $x \in \mathscr{X}$, on note K_x la fonction $\mathrm{K}_x : x' \mapsto \mathrm{K}(x, x')$. Soit $\tilde{\mathscr{X}}$ le sous-espace vectoriel de $\mathbb{R}^{\mathscr{X}}$ engendré par les fonctions $(\mathrm{K}_x)_{x \in \mathscr{X}}$. Soit A l'ensemble des vecteurs $\alpha = (\alpha_x)_{x \in \mathscr{X}} \in \mathbb{R}^{\mathscr{X}}$ ayant un nombre fini de coefficients non-nuls. $\tilde{\mathscr{X}}$ peut donc s'écrire :

$$\tilde{\mathscr{X}} = \left\{ \sum_{x \in \mathscr{X}} \alpha_x \mathrm{K}_x \right\}_{\alpha \in \mathrm{A}},$$

où la somme est bien définie car seul un nombre fini de termes est non-nul.

Soit $\tilde{x}, \tilde{x}' \in \tilde{\mathscr{X}}$. Il existe $\alpha, \beta \in \mathrm{A}$ tels que

$$\tilde{x} = \sum_{x \in \mathscr{X}} \alpha_x \mathrm{K}_x \qquad \text{et} \qquad \tilde{x}' = \sum_{x \in \mathscr{X}} \beta_x \mathrm{K}_x.$$

On pose alors :

$$\langle \tilde{x}, \tilde{x}' \rangle = \sum_{x, x' \in \mathscr{X}} \alpha_x \beta_{x'} \mathrm{K}(x, x'),$$

où la somme a bien un sens car seul un nombre fini de termes est non-nul. L'objectif des questions ci-dessous est de montrer que $\langle \cdot\,, \cdot \rangle$ est bien un produit scalaire sur $\tilde{\mathscr{X}}$ et qu'il existe une application $\psi : \mathscr{X} \to \tilde{\mathscr{X}}$ telle que K est le noyau associé.

2) Montrer que la valeur de $\langle \tilde{x}, \tilde{x}' \rangle$ ne dépend pas du choix des coefficients $(\alpha_x)_{x \in \mathscr{X}}$ et $(\beta_x)_{x \in \mathscr{X}}$ (qui peuvent ne pas être uniques).

3) Montrer que $\langle \cdot\,, \cdot \rangle$ est bilinéaire et symétrique.

4) Soit $\tilde{x} \in \tilde{\mathscr{X}}$. Montrer que $\langle \tilde{x}, \tilde{x} \rangle \geqslant 0$.

5) Soit $\tilde{x} \in \tilde{\mathscr{X}}$ et $x \in \mathscr{X}$. Pour $\lambda \in \mathbb{R}$, en considérant la quantité

$$\langle \lambda \tilde{x} + \tilde{x}(x) \mathrm{K}_x, \lambda \tilde{x} + \tilde{x}(x) \mathrm{K}_x \rangle,$$

montrer que :

$$\tilde{x}(x)^4 \leqslant \tilde{x}(x)^2 \langle \tilde{x}, \tilde{x} \rangle \, \mathrm{K}(x, x).$$

6) En déduire que si $\tilde{x} \in \tilde{\mathscr{X}}$ est tel que $\langle \tilde{x}, \tilde{x} \rangle = 0$, alors $\tilde{x} = 0$.

7) Montrer que K est le noyau associé à une application de redescription $\psi : \mathscr{X} \to \tilde{\mathscr{X}}$ à préciser.