

Examen : Introduction à l'apprentissage automatique

3 juillet 2020

Tous les documents et les ordinateurs connectés sont autorisés.

Les exercices sont indépendants les uns des autres.

Les questions peuvent être traitées de manière indépendante en admettant le résultat des questions précédentes.

Le barème (sur 20 points, auxquels s'ajoutent 3 points bonus) n'est donné qu'à titre indicatif.

Notation

Dans tout le sujet, on notera

$$\text{sign} : x \in \mathbb{R} \mapsto \begin{cases} 1 & \text{si } x > 0 \\ -1 & \text{sinon.} \end{cases}$$

Exercice 1 (Classifieur de Bayes, 9 points)

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \{\pm 1\}$. On appelle $\pi = \mathbb{P}(Y = 1)$.

1. (1 point) Rappeler la définition du classifieur de Bayes.
2. Supposons que $d = 2$, c'est-à-dire $X = (X_1, X_2)$ avec X_1 et X_2 les composantes de X , qui sont des variables aléatoires réelles. Supposons de plus que pour tout $y \in \{\pm 1\}$, $\exists m_y \in \mathbb{R} : [X_1|Y = y] \sim \mathcal{N}(m_y, 1)$, $[X_2|Y = y] \sim \mathcal{U}([0, 1])$ et $X_1 \perp\!\!\!\perp X_2$.
 - a) (1 point) Dessiner un schéma illustrant le problème de classification (on pourra représenter les densités).
 - b) (1 point) Quelle est la densité de $[X|Y = 1]$?
 - c) (1 point) Calculer le classifieur de Bayes associé au problème.
 - d) (1 point) Sans détailler les calculs, quel aurait été le classifieur de Bayes si nous avions supposé $d = 3$ et pour tout $y \in \{\pm 1\}$, $\exists m_y \in \mathbb{R}^2 : [(X_1, X_2)|Y = y] \sim \mathcal{N}(m_y, C)$ (où C une matrice 2×2 symétrique définie positive), $[X_3|Y = y] \sim \mathcal{U}([0, 1])$ et $(X_1, X_2) \perp\!\!\!\perp X_3$?
3. On appelle $\eta : x \in \mathbb{R}^d \mapsto \mathbb{E}[Y|X = x]$ la fonction de régression du problème et $h^* : x \in \mathbb{R}^d \mapsto \text{sign}(\eta(x))$.

- a) (1 point) En développant l'espérance, montrer que h^* est le classifieur de Bayes du problème.
- b) (1 point) Soit $h : \mathbb{R}^d \rightarrow \{\pm 1\}$. Montrer que $\mathbb{P}(Y \neq h(X)) = \frac{1 - \mathbb{E}[Yh(X)]}{2}$.
- c) (1 point) Soit $x \in \mathbb{R}^d$. En remarquant que pour tout $u \in \mathbb{R}^d$: $\text{sign}(u) = \frac{|u|}{u}$, déterminer le signe de $\mathbb{E}[Y(h^*(X) - h(X)) | X = x]$.
- d) (1 point) En déduire que h^* est un minimiseur de $h \mapsto \mathbb{P}(Y \neq h(X))$ sur l'ensemble $\{\pm 1\}^{\mathbb{R}^d}$ des fonctions de \mathbb{R}^d dans $\{\pm 1\}$.

Exercice 2 (Analyse linéaire discriminante, 6 points)

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \{\pm 1\}$. On appelle $\pi = \mathbb{P}(Y = 1)$ et on suppose qu'il existe une matrice $\Sigma \in \mathbb{R}^{d \times d}$ symétrique définie positive telle que pour tout $y \in \{\pm 1\}$, $\exists \mu_y \in \mathbb{R}^d : [X | Y = y] \sim \mathcal{N}(\mu_y, \Sigma)$ et $\mu_1 \neq \mu_{-1}$.

- 1. (1 point) Expliciter la densité de probabilité du vecteur aléatoire X en fonction des densités respectives $p_1 : \mathbb{R}^d \rightarrow \mathbb{R}_+$ et $p_{-1} : \mathbb{R}^d \rightarrow \mathbb{R}_+$ de $[X | Y = 1]$ et $[X | Y = -1]$.
- 2. (1 point) Soit $h^* : x \in \mathbb{R}^d \mapsto \text{sign}(\pi p_1(x) - (1 - \pi)p_{-1}(x))$. Montrer l'équivalence, pour tout $x \in \mathbb{R}^d$:

$$h^*(x) = 1 \iff (\mu_1 - \mu_{-1})^\top \Sigma^{-1} \left(x - \frac{\mu_1 + \mu_{-1}}{2} \right) > \log \left(\frac{1 - \pi}{\pi} \right).$$

- 3. (1 point) Donner une interprétation géométrique de la condition précédente lorsque $\Sigma = I_d$ (la matrice identité de taille d) et $\pi = \frac{1}{2}$.
- 4. (1 point) Soit $d_\Sigma : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \sqrt{(x - y)^\top \Sigma^{-1} (x - y)}$ la distance de Mahalanobis induite par Σ . Donner une interprétation géométrique de cette distance.
- 5. (1 point) Soit $Z \sim \mathcal{N}(\mu_{-1}, \Sigma)$. Montrer que

$$(\mu_1 - \mu_{-1})^\top \Sigma^{-1} (Z - \mu_{-1}) \sim \mathcal{N}(0, d_\Sigma(\mu_1, \mu_{-1})^2).$$

- 6. (1 point) Soit $\Phi : \mathbb{R} \rightarrow]0, 1[$ la fonction de répartition de $\mathcal{N}(0, 1)$. Supposons que $\pi = \frac{1}{2}$. Montrer que

$$\mathbb{P}(h^*(X) = 1 | Y = -1) = 1 - \Phi \left(\frac{d_\Sigma(\mu_1, \mu_{-1})}{2} \right).$$

- 7. (1 point (bonus)) En déduire une expression de $\mathbb{P}(h^*(X) \neq Y)$ lorsque $\pi = \frac{1}{2}$.

Exercice 3 (Une variante des SVM, 5 points)

Soient $\{(X_1, Y_1), \dots, (X_n, Y_n)\} \subset \mathbb{R}^d \times \{\pm 1\}$ un échantillon de n couples aléatoires et

$C > 0$. On considère le problème d'optimisation :

$$\begin{aligned} & \underset{w \in \mathbb{R}^d, b \in \mathbb{R}, s \in \mathbb{R}^n}{\text{minimiser}} \quad \frac{1}{2} \|w\|_{\ell_2}^2 + \frac{C}{2} \sum_{i=1}^n s_i^2 \\ & \text{s. c.} \quad \begin{cases} \forall i \in \llbracket 1, n \rrbracket, s_i \geq 0 \\ Y_i(w^\top X_i + b) \geq 1 - s_i. \end{cases} \end{aligned} \quad (\text{P1})$$

- (1 point) La résolution de (P1) permet-il de construire un régresseur ou un classifieur ? Celui-ci est-il linéaire ou non-linéaire ? Expliquer votre réponse en réalisant une figure illustrant le problème en jeu.
- (1 point) Montrer que tout couple (\hat{w}, \hat{b}) solution de (P1) est aussi solution du problème

$$\underset{w \in \mathbb{R}^d, b \in \mathbb{R}}{\text{minimiser}} \quad \frac{1}{2} \|w\|_{\ell_2}^2 + C \sum_{i=1}^n \ell_i(w^\top X_i + b),$$

où $\{\ell_i : \mathbb{R} \rightarrow \mathbb{R}_+\}_{1 \leq i \leq n}$ sont des fonctions à préciser.

Quelle est la différence entre le problème d'intérêt et une machine à vecteurs supports telle que vue en cours ?

- (1 point) En notant $y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$, $A = \begin{pmatrix} Y_1 X_1^\top \\ \vdots \\ Y_n X_n^\top \end{pmatrix} \in \mathbb{R}^{n \times d}$ et $\mathbf{1}$ le vecteur rempli de 1 (de taille adéquate), montrer que (P1) peut se réécrire

$$\begin{aligned} & \underset{w \in \mathbb{R}^d, b \in \mathbb{R}, s \in \mathbb{R}^n}{\text{minimiser}} \quad \frac{1}{2} \|w\|_{\ell_2}^2 + \frac{C}{2} \|s\|_{\ell_2}^2 \\ & \text{s. c.} \quad \begin{cases} Aw + by \succcurlyeq \mathbf{1} - s \\ s \succcurlyeq 0. \end{cases} \end{aligned} \quad (\text{P2})$$

- (2 points) Montrer que le problème d'optimisation dual à (P2) est

$$\begin{aligned} & \underset{\alpha \in \mathbb{R}^n}{\text{maximiser}} \quad \mathbf{1}^\top \alpha - \frac{1}{2} \|A^\top \alpha\|_{\ell_2}^2 - \frac{1}{2C} \|\alpha\|_{\ell_2}^2 \\ & \text{s. c.} \quad \alpha \succcurlyeq 0, \quad y^\top \alpha = 0. \end{aligned}$$

- (2 points (bonus)) Soit $f : w \in \mathbb{R}^d \mapsto \ell_1(w^\top X_1)$. Déterminer ∇f et montrer que ∇f est une application L -lipschitzienne pour une constante L à préciser.