

TRAVAUX DIRIGÉS DE  
**MACHINE LEARNING**  
CYCLE PLURIDISCIPLINAIRE D'ÉTUDES SUPÉRIEURES  
UNIVERSITÉ PARIS SCIENCES ET LETTRES

Joon Kwon

vendredi 24 mars 2023



**EXERCICE 1.** — On se place dans un cadre de classification binaire avec  $\mathcal{X} = [0, 1]$  et  $\mathcal{Y} = \{0, 1\}$ . Soit  $P$  une distribution sur  $\mathcal{X} \times \mathcal{Y}$ , ainsi que des variables aléatoires  $(X, Y) \sim P$ . Un prédicteur  $f_*$  minimisant le risque, c'est-à-dire tel que :

$$R(f_*) = \min_{f \in \mathcal{F}(\mathcal{X}, \mathcal{Y})} R(f)$$

est appelé prédicteur *oracle*. On rappelle que la perte considérée par défaut en classification est la perte 0-1 :

$$\ell(y, y') = \mathbb{1}_{\{y \neq y'\}} = \begin{cases} 1 & \text{si } y \neq y', \\ 0 & \text{si } y = y'. \end{cases}$$

On considère ci-dessous deux cas distincts. Dans chacun des cas, déterminer un prédicteur oracle ainsi que son risque.

- 1) La distribution  $P$  est telle que  $X$  suit une loi uniforme sur  $[0, 1]$  et  $Y = 1$  si  $X \geq 1/2$  et  $Y = 0$  sinon. Déterminer un prédicteur oracle ainsi que son risque.
- 2) La distribution  $P$  est telle que  $X$  et  $Y$  sont indépendants et  $\mathbb{P}[Y = 1] = 2/3$ .

**EXERCICE 2.** — On se place dans un cadre de classification avec l'ensemble d'entrées  $\mathcal{X} = [0, 10]$  et l'ensemble de sorties  $\mathcal{Y} = \{0, 1\}$ . On considère sur  $\mathcal{X}$  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  $1 \leq k \leq 5$ , on note  $\hat{f}^{(k)}$  le prédicteur  $k$ NN (pour la distance  $\rho$ ) 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  $\hat{f}^{(k)}$ .

