2023-2024 **Université Gustave Eiffel**

Mathematical fundations of data science, 3 hours

Exercice 1. Kernel density estimation
Let $X_1, \dots, X_n \in \mathbb{R}^d$ be *i.i.d.* random variables drawn from a measure with density function f. We assume f is L-Lipschitz continuous on $\mathbb{R}^d$.

We regress f using the following kernel estimator

$$\forall x \in \mathbb{R}^d, \hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(\|x - X_i\|),$$

where

- $h > 0$;
- K is a function with compact support in $[0, 1]$ and such that
$$\int_{\mathbb{R}^d} K(\|x\|)dx = 1.$$
-
$$\int_{\mathbb{R}^d} K(\|x\|)^2 dx = V.$$
- $K_h(u) = \frac{1}{h^d} K\left(\frac{u}{h}\right)$.

1. Show that $\hat{f}$ is a density function.
2. Show that
$$|\mathbb{E}[\hat{f}(x) - f(x)]| \leq Lh.$$
3. Show that
$$Var(\hat{f}(x)) \leq \frac{Vf(x)}{h^d n} + \frac{VL}{h^{d-1}n}.$$
4. Compute the MSE of $\hat{f}(x)$. How should h behave with respect to n to minimize this quantity?
5. Suppose f is differentiable with a L'-Lipschitz gradient. Improve the bias and MSE bounds.

Exercice 2. Logistic regression
Let $X_1, \dots, X_n \in \mathbb{R}^d$ (size 1xd) for some positive integer d. For any $i \in \{1, \dots, n\}$, let Y_i be a Rademacher random variable (i.e. taking values in $\{-1, 1\}$ and such that $P(Y_1 = 1 \mid X_1) = f(X_1\theta)$, where $\theta \in \mathbb{R}^d$ (size d x 1) and $f : u \to \frac{1}{1+e^{-u}}$.

1. Show that the Maximum likelihood estimator of θ verifies
$$\hat{\theta}_{MLE} = \arg\min_{\alpha} \sum_{i=1}^{n} l(-Y_i X_i \alpha)),$$
where $l : z \to \log_2(1 + exp(z))$ is the logistic loss. In other words, the MLE estimator under this modelization is equivalent to an empirical risk minimization approach with loss function l.
2. We now want to check that this loss function verifies the assumptions of Zhang's lemma. Show that l is non-decreasing, convex and non-negative.
3. For any $\eta > 0$. Let $H_\eta : z \to \eta l(z) + (1 - \eta) l(-z)$ and $\tau(\eta) = \inf_z H_\eta(z)$. Find two constants $c > 0$ and $\gamma \in [0, 1]$ such that
$$\forall \eta \in [0, 1], |\eta - \frac{1}{2}| \leq c(1 - \tau(\eta))^\gamma.$$
Hint: show that
$$1 - \tau(\eta) = \frac{1}{\log(2)}\left(\eta \log\left(\frac{\eta}{1/2}\right) + (1 - \eta)\log\left(\frac{1 - \eta}{1/2}\right)\right)$$
and use the following lemma

Lemma 1 (Discrete Pinsker's inequality) *For any two probability measures P, Q on the discrete set $\mathcal{X}$, we have*

$$\frac{1}{2}\left(\sum_{x\in\mathcal{X}} |P(x) - Q(x)|\right)^2 \leq \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right).$$

Exercice 3. Estimation error with finite number of predictors
We consider the problem of classification. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random variables with $X_1, \dots, X_n \in \mathcal{X}$ and $Y_1, \dots, Y_n \in \{-1, 1\}$. We denote by $R_P(f)$ the risk associated to a classifier f:

$$R_P(f) = \mathbb{E}[l(-Y_1 f(X_1))],$$

where l is a loss function. The associated empirical risk is

$$R_n(f) = \frac{1}{n}\sum_{i=1}^{n} l(-Y_i f(X_i)).$$

Let $\mathcal{F}$ be a set of classifiers (that is functions from $\mathcal{X}$ and suppose $\inf_{f\in\mathcal{F}} R_P(f)$ is attained by $f^\star$. Finally, we denote by $\hat{f}_n$ the estimator minimizing the empirical risk

$$\hat{f}_n = \arg\min_{f\in\mathcal{F}} R_n(f)$$

The objective of this exercise is to bound the excess of risk of the estimator defined by

$$\mathbb{E}[R_P(\hat{f}_n) - R_P(f^\star)].$$

1. Show that
$$R_P(\hat{f}_n) - R_P(f^\star) \leq 2\sup_{f\in\mathcal{F}} |R_P(f) - R_n(f)|.$$

2. Let $Z_1, \ldots Z_k$ be k random variables in $[0,1]$. For $\lambda > 0$, show that
$$\max_j (Z_j - \mathbb{E}[Z_j]) \leq \frac{1}{\lambda}\log\left(\sum_{i=1}^{k} e^{\lambda(Z_i - \mathbb{E}[Z_i])}\right)$$
and deduce
$$\mathbb{E}[\max_j (Z_j - \mathbb{E}[Z_j])] \leq \frac{1}{\lambda}\log\left(\sum_{i=1}^{k} \mathbb{E}[e^{\lambda(Z_i - \mathbb{E}[Z_i])}]\right).$$

3. Using a Taylor expansion of
$$\lambda \to \log(\mathbb{E}[e^{\lambda(Z_i - \mathbb{E}[Z_i])}])$$
show that
$$\mathbb{E}[e^{\lambda(Z_i - \mathbb{E}[Z_i])}] \leq e^{\lambda^2 Var(Z_i)/2} \leq e^{\lambda^2/8}.$$

4. Show that
$$\mathbb{E}[\max_j (Z_j - \mathbb{E}[Z_j])] \leq \frac{\sqrt{2\log k}}{2}.$$

5. We now assume $|\mathcal{F}| = k$ and
$$\forall x \in \mathcal{X}, \forall f \in \mathcal{F}, \max(l(f(X_i)), l(-f(X_i))) \leq l_\infty.$$
Prove
$$\mathbb{E}[R_P(\hat{f}_n) - R_P(f^\star)] \leq l_\infty \sqrt{\frac{2\log(2k)}{n}}.$$