HOMEWORK ASSIGNMENT 1

1 Bayes predictor for mean absolute error loss

We consider the task of supervised learning with mean absolute error loss : $\mathcal{Y} = \mathbb{R}$ and $\ell(y, y') = |y - y'|$. Assume that for every $x \in \mathcal{X}$, conditionally to $X = x$, the random variable Y has a density $f_{Y|X}(y|x)$ w.r.t. the Lebesgue measure on $\mathbb{R}$.

1. What is the Bayes predictor g_P^* in this case ?
2. We define the empirical risk of a predictor $g : \mathcal{X} \to \mathcal{Y}$ by
$$\widehat{R}_n[g] = \frac{1}{n}\sum_{i=1}^{n}|Y_i - g(X_i)|.$$
Let us consider the class of constant predictors : $\mathcal{G} = \{g : \mathcal{X} \to \mathbb{R} : \exists c \in \mathbb{R} \text{ t.q. } g(x) \equiv c\}$. Determine the explicit form of the minimiser of the empirical risk, denoted by $\widehat{g}_{n,\mathcal{G}}$, over the class $\mathcal{G}$.
3. We are now given a partition of $\mathcal{X}$ into K parts $A_1, \ldots, A_K$. We set
$$\mathcal{G}_A = \Big\{g : \mathcal{X} \to \mathbb{R} : g \text{ is constant on every } A_k\Big\}.$$
Determine the predictor $\widehat{g}_{n,A}$ minimizing the empirical risk over $\mathcal{G}_A$.

2 Consistency of the kNN classifier

The aim of this exercise is to show that the kNN classifier (k nearest neighbor) is not consistent if k is fixed independently of the sample size n. To this end, we consider the classification problem with $\mathcal{X} = [0, 1]$ et $\mathcal{Y} = \{0; 1\}$. We denote by P_X the marginal distribution of X_is and assume that

$$\eta^*(x) = P(Y_1 = 1|X_1 = x) \equiv \frac{3}{4}, \qquad \forall x \in \mathcal{X}.$$

The goal of the next questions is to compute the risk of the Bayes classifier g_P^* as well as the expected risk of the kNN-classifier $\widehat{g}_{n,k}$. We will see that this latter does not depend on the sample size and is always strictly larger than the former.

We recall that in order to compute $\widehat{g}_{n,k}(x)$ for a given $x \in \mathcal{X}$ endowed with some metric,

- we first find $i_1, \ldots, i_k$ such that $\{X_{i_1}, \ldots, X_{i_k}\}$ are the k closest points to x among $\{X_1, \ldots, X_n\}$,
- then, we assign to $\widehat{g}_{n,k}(x) = 1$ the most frequent label in the family $\{Y_{i_1}, \ldots, Y_{i_k}\}$. In other terms,
$$\widehat{g}_{n,k}(x) = \mathbb{1}(Y_{i_1} + \ldots + Y_{i_k} > k/2).$$

1. Check that for any deterministic mapping $g : \mathcal{X} \to \{0; 1\}$, it holds that :
$$R_P(g) = \mathbf{E}_{P_X}[\eta^*(X)] + \mathbf{E}_{P_X}[g(X)(1 - 2\eta^*(X))].$$

2. Deduce from the previous question that if $\eta^* \equiv 3/4$, then the Bayes classifier is given by $g_P^* \equiv 1$ and its risk is $R_P(g_P^*) = 1/4$.
3. Show that for every $g : \mathcal{X} \to \{0;1\}$, it holds that :

$$R_P(g) = \frac{3}{4} - \frac{1}{2}\int_{\mathcal{X}} g(x)\, P_X(dx).$$

4. Let $\mathcal{D}_n = \{(X_i, Y_i); i = 1, \ldots, n\}$ and let $\widehat{g}_{n,1}(x) = \widehat{g}_{\mathrm{NN}}(x, \mathcal{D}_n)$ be the nearest neighbor classifier (1NN). For a fixed $x \in \mathcal{X}$, we are looking for the value of $\mathbf{E}_P[\widehat{g}_{\mathrm{NN}}(x, \mathcal{D}_n)]$, where the expectation is w.r.t. $\mathcal{D}_n$. For every $i = 1, \ldots, n$, let us set

$$Z_i = \begin{cases} 1, & \text{si } X_i \text{ is the NN of } x \\ 0, & \text{otherwise.} \end{cases}$$

Show that

$$\mathbf{E}_P[\widehat{g}_{\mathrm{NN}}(x, \mathcal{D}_n)] = \sum_{i=1}^{n} \mathbf{E}_P[Y_i Z_i]. \tag{1}$$

5. Check that for every i, Y_i is independent of $(X_1, \ldots, X_n)$, and that Y_i and Z_i are independent.
6. Using the previous question and the obvious relation $\sum_{i=1}^{n} Z_i = 1$ show that :

$$\mathbf{E}_P[R_P(\widehat{g}_{\mathrm{NN}})] = \frac{3}{8}.$$

Conclude.

7. Consider the case of 3 nearest neighbors : $\widehat{g}_{3-\mathrm{NN}}$. Show that its expected risk $\mathbf{E}_P[R_P(\widehat{g}_{3-\mathrm{NN}})]$ is equal to $21/64$.
8. Let us turn now to the general case of kNN predictor, $\widehat{g}_{k-\mathrm{NN}}$. Let $V_1, \ldots, V_k$ be i.i.d. Bernoulli random variables with parameter $3/4$. Show that

$$\mathbf{E}_P[\widehat{g}_{k-\mathrm{NN}}(x, \mathcal{D}_n)] = \mathbf{P}(\overline{V}_k > 1/2).$$

Deduce from this relation that the expectation above tends to 1 when $k \to \infty$ and, as a consequence, the expected risk $\mathbf{E}_P[R_P(\widehat{g}_{k-\mathrm{NN}})]$ tends to the risk of the Bayes classifier, *i.e.*, to $1/4$.