Exercises - Local Methods : KNN, Nadaraya Watson regression

Exercice 1 (K-Nearest Neighbours)**:**
Consider the KNN rule with equal weights for classification. One observes n i.i.d. replications $D_n = \{(X_i, Y_i),\, i \le n\}$ of a random pair (X, Y) with $X \in \mathbb{R}^d$ and $Y \in \{0, 1\}$. Recall that the KNN classifier given the dataset D_n is

$$g_{k,n}(x) = \begin{cases} +1 & \text{if } \sum_{i=1}^{k} Y_{(i)}(x) \ge 1/2 \\ 0 & \text{otherwise,} \end{cases}$$

where $Y_{(i)}$ is the label of the i^{th} nearest neighbour $X_{(i)}(x)$ of point x. Recall also that the regression function is defined as $\eta(x) = \mathbb{P}[Y = 1|X]_{X=x}$. We assume here that X is uniformly distributed over $[0, 1]$ and that $\eta(x) = 1/3$ for all $x \in [0, 1]$.

1. Compute the joint probability $\mathbb{P}[X \in [a, b], Y = 1]$ for all $0 < a < b$.
2. Give the expression for the Bayes classifier g^* For a generic regression function $\eta(x)$ and for this particular problem. Give its error risk $\mathbb{P}[g^*(X) \neq Y]$ (the Bayes error) in general and in this particular case.
3. Consider the training set

$$(X_1 = 0.4, Y_1 = 0), (X_2 = 0.2, Y_2 = 0), (X_3 = 0.7, Y_3 = 1)$$

   Compute the k-NN classifier (for all x) in the case $k = 1$ and $k = 3$.
4. Compute in each case the expected classification error $R(g_{knn}) = \mathbb{P}[g_{knn}(X) \neq Y]$ for a new observation (X, Y). Compare with the Bayes error
5. Consider now another model : X is again uniformly distributed over $[0, 1]$ but now $\eta(x) = \mathbb{1}_{(1/2,1]}(x)$. Repeat questions 1,2,3,4 with the same dataset as in Q3.

Exercice 2 (Key lemma explaining K-nn consistency)**:**
Let P_X be the law of X on $\mathbb{R}^d$. Let x in the support of P_X, that is for all $\epsilon > 0$, for all ball $B(x, \epsilon)$ of radius ϵ centered at x, $P_X(B(x, \epsilon)) > 0$. Let k_n be a sequence of integers such that $k_n/n \to 0$. Let $(X_i, i \in \mathbb{N})$ be an iid sequence distributed as X.

- Show that, almost surely , $\|X_{(k_n,n)} - x\| \to 0$ where $X_{(k_n)}$ is the k_n'th nearest neighbour of x among $(X_1, \dots, X_n)$.

Exercice 3 (Nadaraya-Watson Regression)**:**
Consider a regression problem for a random pair $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ with target Y and covariate X. We assume that $\mathbb{E}[Y^2] < \infty$. The goal is to approach $m(x) = \mathbb{E}[Y|X = x]$. We assume that the pair (X, Y) has a density $f(x, y)$ with respect to the Lebesgue measure on $\mathbb{R}^{d+1}$.

1. Write $m(x)$ as an integral involving the f and the marginal density f_X of X.
2. Given kernels K_x, K_y of order 1 for density estimation of X and Y, (thus $\int uK(u)\mathrm{d}u = 0$ and $\int K(u)\mathrm{d}u = 1$) define the product kernel density estimate as

$$\widehat{f}(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K_x\Big(\frac{X_i - x}{h}\Big) K_y\Big(\frac{Y_i - y}{k}\Big).$$

(a) Recall the expression for the Kernel density estimate $\widehat{f}_X$ of f_X based on K_x and a dataset $X_{1:n}$.

(b) Show that the Nadaraya-Watson estimate $\widehat{m}(x)$ based on K_x is the plug-in estimate of $\mathbb{E}[Y|X=x]$ based on the expression found in Question 1) up to replacing f_X, f with the kernel density estimates $\widehat{f}_X$ and $\widehat{f}$.

3. Recall why $m(x)$ can be seen as the minimizer of some quadratic risk at point x.
4. Show that Nadaraya-Watson estimator is the solution of following wheighted optimization problem :

$$\widehat{m}(x) = \underset{c\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1} K\left(\frac{x-X_i}{h}\right)(Y_i - c)^2.$$

Exercice 4 (Consistency of the Nadaraya-Watson estimator)**:**
The goal is to prove the following result, where the notations are the same as in the previous exercise.

Théorème 0.1 (Consistency of Nadaraya-Watson)
Consider the regression model $Y = m(X) + \epsilon$ where $\mathbb{E}(\epsilon^2) = \sigma^2$ and $\mathbb{E}[\epsilon] = 0$. Let $h_n \to 0$, $nh_n \to \infty$ as $n \to \infty$. Let f_X be the density of X and assume that $\mathbb{E}[Y^2] < \infty$. Let x such that f_X and m are continuous at x and $f_X(x) > 0$. Then the N-W estimate $\widehat{m}_n$ is weakly consistent for estimating $m(x)$, that is

$$\widehat{m}_n(x) \xrightarrow[n\to\infty]{P} m(x).$$

The proof uses the following

Lemme 0.2 (Parzen, 1962, *On Estimation of a probability density function and mode*)
Suppose that $K(x)$ is a bounded and integrable function such that $\lim_{\|x\|\to\infty} \|x\|K(x) = 0$. Let g be an integrable function. Then

$$\lim_{h\to 0} \frac{1}{h} \int K\left(\frac{u-x}{h}\right) g(u)\mathrm{d}u = g(x) \int K(u)\mathrm{d}u.$$

Notice that the N-W estimator at x writes

$$\widehat{m}(x) = \frac{\widehat{\phi}_n(x)}{\widehat{f}_n(x)}$$

where

$$\widehat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad ; \quad \widehat{\phi}_n(x) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i.$$

1. Show that $\mathbb{E}[\widehat{\phi}_n(x)] \to m(x)f(x)$
2. Show that $nh\mathbb{V}\mathrm{ar}(\widehat{\phi}_n(x)) \to (m^2(x) + \sigma^2) f_X(x) \int K^2(u)\mathrm{d}u$
3. Conclude that $\widehat{\phi}_n(x) \xrightarrow{P} m(x)f_X(x)$.
4. By a similar argument show that $\widehat{f}_n(x) \xrightarrow{P} f_X(x)$
5. Conclude the proof.