TD1: Non parametric regression

1 Bias and variance

Consider the regression problem of dimension N:

$$X_k = \theta_k + \varepsilon_k, k = 1, \dots, N,$$

where $X_1, \dots, X_N$ are i.i.d random variables in $\mathbb{R}$,

$\theta \in \mathbb{R}^N$ the unknown parameter and $\varepsilon_1, \dots, \varepsilon_N$ are i.i.d samples following a centered Gaussian with known variance σ^2.

We want to find an estimator of θ of the form $\widehat{\theta}(\lambda) = \lambda X$ with $\lambda \in [0, 1]$.

1. Compute its bias $\mathbb{E}(\widehat{\theta}(\lambda) - \theta)$ and the sum of the variance its coordinates: $\sum_{k=1}^N \mathbb{V}(\widehat{\theta}(\lambda)_k)$. What is the effect of λ on the estimator ?

2. Compute $\lambda^\star$ the minimizer of the oracle risk of $\widehat{\theta}(\lambda)$ given by:

$$R(\lambda) = \mathbb{E}(L(\lambda))$$

   where $L(\lambda) = \sum_{k=1}^N (\widehat{\theta}(\lambda)_k - \theta_k)^2$.

3. Propose an unbiased estimator of θ_k^2, given the observed sample x_k.

4. Deduce an unbiased estimator $C(\lambda)$ of $R(\lambda) - \sum_{k=1}^N \theta_k^2$.

2 Regression

Let $f : \mathbb{R} \to \mathbb{R}$ be an unknown function. Suppose we observe noisy values of f:

$$Y_i = f(X_i) + \epsilon_i,$$

where $(X, \epsilon), (X_i, \epsilon_i)_{i=1,\dots,n} \subset \mathbb{R} \times \mathbb{R}$ is an i.i.d. collection of random variables such that $\epsilon \perp X$, $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[\epsilon^2] = \sigma^2$ and $\mathbb{E}[f(X)^2] < \infty$. To estimate the function f, we shall use a sequence of functions $(h_k)_{k\geq 1}$ where each $h_k : \mathbb{R} \to \mathbb{R}$ and $G = \mathbb{E}[h(X)h(X)^T]$ is invertible. . The procedure is as follows. We choose K, we set $h(x) = (h_1(x), \dots h_K(x))^T$ and then compute

$$\theta_{K,n} \in \underset{\theta \in \mathbb{R}^K}{\arg\min} \sum_{i=1}^n (Y_i - h(X_i)^T\theta)^2.$$

The estimate of f is given by $f_{K,n}(x) = h(x)^T\theta_{K,n}$. Define the risk $R(\theta) = \mathbb{E}[(f(X) - h(X)^T\theta)^2]$. The aim is to study the $R(\theta_{n,K})$.

1. Let $\theta_K^* \in \arg\min_{\theta \in \mathbb{R}^K} R(\theta)$. Give the normal equations satisfied by θ_K^* and deduce an expression for θ_K^*.

2. Give the expression for $\theta_{K,n}$.

3. Show that the estimated function $f_{K,n}$ is invariant by any linear invertible transform on the set of functions h, i.e. h is replaced by Ah where $A \in \mathbb{R}^{K\times K}$ is invertible. In the following we shall assume that $G = I_K$.

4. Show that for any $\theta \in \mathbb{R}^K$, $R(\theta) = R(\theta_K^*) + \|\theta - \theta_K^*\|^2$ where the norm $\|\cdot\|$ should be specified.

5. From now on, we suppose that the smallest eigenvalue of $n^{-1}\sum_{i=1}^n h(X_i)h(X_i)^T$ is lower bounded by $\lambda > 0$. Show that $\|\theta_{K,n} - \theta_K^*\| \leq \lambda^{-1/2}\|n^{-1}\sum_{i=1}^n \xi_i h(X_i)\|$ where $\xi_i \neq \epsilon_i$ should be specified.

6. Show that $n^{-1}\sum_{i=1}^n \xi_i h(X_i) \to 0$ almost surely (hint: use the Cauchy-Schwarz inequality).

7. Suppose that $f \in \mathrm{span}((h_k)_{k\geq 1})$. Conclude that choosing K large enough, $\limsup_n R(\theta_{K,n})$ can be made arbitrarily small.