Homework 6 - due Wed. Mar. 11th
High-Dimensional Approximation, Probability, and Statistical Learning

Instructor: Mauro Maggioni
Office: 302D Whitehead Hall
Web page: https://mauromaggioni.duckdns.org
E-mail: mauro.maggioni *at* youknowwhat.edu

Homework Policies

As in the first homework set.

Assignment

Study/review the parts of Rigollet's lecture notes covered in class, in particular the "Introduction chapter", and the "Unconstrained least squares estimator" in section 2.2 (and Theorem 1.19 which we used in the proof, and we already knew from materials from Verhsynin's lecture notes covered in class).

Exercises

Exercise 1 (100pt).

1. Let $\mathcal{P}_d$ be the space of polynomials on $[0,1]$ of degree up to $d-1$. Show that it is a d-dimensional subspace of $L^2([0,1],dx)$ (dx denotes the uniform, or Lebesgue, measure, albeit this doesn't matter much here). For any set of points $x_1,\dots,x_n$ in $[-1,1]$, $n\geq d$, and denote by $\mathbb{X}$ the $n\times d$ matrix defined by $(\mathbb{X})_{ij}=(x_i)^{j-1}$, $i=1,\dots,n; j=1,\dots,d$. The columns of $\mathbb{X}$ are monomials of degree 0 to $d-1$ evaluated at the n points. What is the rank of $\mathbb{X}$?
2. Let $f_1(x)=e^{-x}$ and $f_2(x)=1/(1+x^2)$. Let $y_i=f(x_i)+\epsilon_i$, where f is either f_1 or f_2 (run what follows for each of these two choices of f), the x_i's are n equi-spaced points in $[-1,1]$, and ϵ_i's are i.i.d. $\mathcal{N}(0,\sigma^2)$. Construct the estimator $\hat{f}_{n,d}$ (as above, $n\geq d$) defined as the minimizer of the empirical Mean Squared Error over $\mathcal{P}_d$:

$$\hat{f}_{n,d}=\arg\min_{g\in\mathcal{P}_d}\frac{1}{n}\sum_{i=1}^{n}|y_i-g(x_i)|^2\,.$$

To solve this problem, write $(g(x_i))_{i=1}^n=\mathbb{X}\beta$, where $\mathbb{X}\in\mathbb{R}^d$ is as above (with the x_i's being the equi-spaced points) and $\beta\in\mathbb{R}^d$ is a vector of coefficients expressing g as a linear coefficients of the monomials spanning $\mathcal{P}_d$. Show that $\beta=\mathbb{X}^\dagger(y_i)_{i=1}^n$ solves the problem when $\mathbb{X}$ is full rank (i.e. rank d), where $\mathbb{X}^\dagger=(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ is Moore-Penrose inverse of $\mathbb{X}$. [Computational note: be careful with how you compute $\mathbb{X}^\dagger(y_i)_{i=1}^n$, or $\mathbb{X}^\dagger$; in Matlab see the function pinv and the operator \).
3. Look at the following MSE's:

$$MSE_1:=\frac{1}{n}\sum_{i=1}^{n}|f(x_i)-\hat{f}_{n,d}(x_i)|^2\quad,\quad MSE_2:=\frac{1}{N}\sum_{i=1}^{N}|f(x_i^{test})-\hat{f}_{n,d}(x_i^{test})|^2\approxeq\frac{1}{2}\int_{-1}^{+1}|f(x)-\hat{f}_{n,d}(x)|^2dx$$

where N is large (how large) and x_i^{test}'s are i.i.d. uniform in $[-1,1]$, representing a large test set (the $\approxeq$ is because the discrete sum is a Monte Carlo estimate of the continuous integral, the discrete and continuous quantities being $O(N^{-\frac{1}{2}})$ close to each other). MSE_1 is measuring the de-noising effect of the estimator (discussed in class), while MSE_2 is measuring the error in what we called $L^2(\rho_X)$ in class ($\rho_X=\text{Unif}([-1,1])$). You may need n rather large in your experiments, and you may want to repeat each several times to get a mean and std for the random quantities involved. For each $f\in\{f_1,f_2\}$, study the behavior of MSE_1 and MSE_2 as $n\to+\infty$, for d fixed, for various values of d and of σ; then do the same with $d=n^{1/3}$ (you may try other powers as well).
4. Sample n points $x_1,\dots,x_n$ uniformly on $[-1,1]$, and repeat 2 and 3 above. In this case $\mathbb{X}$ is a random matrix, unlike the fixed design case we considered above (and in class).