Homework 3 - Due ~~Wed. Sep 26th~~ Fri. Sep. 28th
High-Dimensional Approximation, Probability, and Statistical Learning

Instructor: Mauro Maggioni | **Office**: 405 Krieger Hall
Web page: www.math.jhu.edu/˜ mauro | **E-mail**: myfirstname.maggioni@jhu.edu

Homework Policies

As for the first homework set.

Assignment

Review: Go over the proof we did in class about convergence of the piecewise constant estimator (on equal subintervals of $[0,1]$) for a Lipschitz function f on $[0,1]$. Make sure you understand both the big picture and the details.

This is the simplest example of regression that I could think of, but still contains many features of the general case. References for the general case include *A distribution-free theory of nonparametric regression*, L. Gyorfi, M. Kohler, A. Krzyzak, H Walk and *Introduction to nonparametric estimation*, A. Tsybakov.

Exercises. Make sure you motivate your answers as rigorously as possible. (*) denotes questions that do not count towards the total score, but the number of correct answers to (*) will be tracked through the semester and considered in the final grade.

Exercise 1 (30pts). Suppose now f is only Hölder α on $[0,1]$, for some $\alpha \in (0,1)$, meaning that there exists a constant C such that for all $x, y \in [0,1]$ we have $|f(x) - f(y)| \leq C|x-y|^\alpha$. What changes in the proof and in the result? Does the bias term change? Does the variance term change? Does the rate of convergence change and how?

Exercise 2 (30pts). Consider the same setting as in class but on the unit cube $[0,1]^D$ (and f Lipschitz, as in class). Describe the piecewise constant estimator you would be using then, and that we briefly mentioned in class (piecewise constant on the natural collection of cubes of side $1/m$). What will change in the proof? Which terms (bias, variance) change and how? Does the convergence rate change and how?

Exercise 3 (40pts). Familiarize yourself with the simple template code for regression available on the course website. Modify it (or write your own code from scratch) to derive a convergence plot (i.e. $\mathbb{E}[||\hat{f}_{n,m} - f||^2_{L^2([0,1])}]$ as a function of n): use a logarithmic scale on the vertical (error) axis, fit a line to the error plot and check the consistency with what is predicted by the theorem.

Change the choice $m = m(n)$, by making m smaller or larger than what the Theorem suggests: comment on what you see.

Please note that in this exercise you may need to choose many values of n, many of which large.