STA3000F: Homework 3

November 22, 2023

In order to reduce your workload during the final season, there are only 3 questions in Homework 3. However, we need to be consistent in terms of grading. Therefore, each question is worth 4/9 points, and the entire homework is worth 4/3 points.

1 Q1: Pointwise risk for Sobolev space

Consider the first order Sobolev class of densities

$$\mathcal{S}(1, L) := \Big\{ p : [0,1] \to \mathbb{R}_+, \text{ s.t.} \int_0^1 |p'(x)|^2 dx \le L, \ \int_0^1 p(x) dx = 1 \Big\},$$

where L is a universal constant. Let $\widehat{p}_{n,h}$ be the kernel density estimator, with a bounded kernel function K supported on $[-1, 1]$. Given $p \in \mathcal{S}(1, L)$ and let $(X_i)_{i=1}^n$ be i.i.d. samples from p, for any $x_0 \in (0, 1)$, prove that

$$\mathbb{E}\big[|\widehat{p}_{n,h}(x_0) - p(x_0)|^2\big] \le C\Big(\frac{1}{nh} + h\Big),$$

for a constant C depending only on L and the kernel function K.

By tuning the bandwidth $h = h_n$ optimally, we can conclude that $|\widehat{p}_{n,h_n}(x_0) - p(x_0)| = O(n^{-1/4})$ with high probability.

2 Q2: Multivariate KDE

Let p be a density in $\mathbb{R}^d$, and let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} p$. For a scalar $\beta \in (0, 1]$, assume that p satisfies the Hölder assumption

$$|p(x) - p(y)| \le \|x - y\|_2^{\beta}.$$

Assume furthermore that $p(x) \le p_{\max}$ for any $x \in \mathbb{R}^d$.

Given a kernel function K such that $\int_{\mathbb{R}^d} K(x)dx = 1$, we can define the kernel density estimator just as the one-dimensional case

$$\widehat{p}_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n K\Big(\frac{x - X_i}{h}\Big).$$

For an appropriate choice of kernel K (you may impose any integrability or regularity conditions as needed), given $x_0 \in \mathbb{R}^d$, derive upper bounds on the bias and variance for $\widehat{p}_{n,h}$, and conclude that

$$\mathbb{E}\big[|\widehat{p}_{n,h}(x_0) - p(x_0)|^2\big] \lesssim n^{-\frac{2\beta}{2\beta+d}},$$

for appropriate choice of h.

3 Q3: Isotonic regression

Consider the function class

$$\mathcal{F} := \big\{ f : [0,1] \to [0,1],\ f \text{ is non-decreasing} \big\}.$$

Show that

$$\log N\Big(\varepsilon; \mathcal{F}, L^2(Q)\Big) \le \frac{c}{\varepsilon} \log\Big(\frac{1}{\varepsilon}\Big),$$

for any probability measure Q on $[0,1]$, for a universal constant $c > 0$.

Conclude that for the constrained least-squares estimator $\widehat{f}$ and $f^* \in \mathcal{F}$, we have

$$\Big\| \widehat{f} - f^* \Big\|_n \lesssim \Big(\frac{\log n}{n}\Big)^{1/3},$$

with high probability.

[Hint: use a piecewise constant function to approximate a monotonic function, and control the covering number using number of pieces.]

[Remark: the logarithmic factor can be removed using highly technical arguments. See van der Vaart and Wellner, Section 2.7.2.]

[Grading: you will get partial credits if you prove a weaker bound. In particular, you will get $1/\alpha$ of the credit if you prove a covering number bound of order $\varepsilon^{-\alpha}$, ignoring logarithmic factors.]