STA3000F: Homework 3

Due: Dec 24 end of day on Quercus

Q1. Nonparametric classification

Given a function $f^* : [0,1] \mapsto [0,1]$ belonging to the Hölder class $\Sigma(\beta)$ with exponent $\beta \in (0,1]$, i.e.,

$$|f^*(x) - f^*(y)| \leq |x-y|^\beta, \quad \text{for any } x, y \in [0,1].$$

Let $x_i = i/n$ for $i = 1, 2, \cdots, n$, and let the observations be

$$Y_i \sim \mathrm{Ber}\big(f^*(x_i)\big), \quad \text{indpendently for } i = 1, 2, \cdots, n.$$

Our goal is to estimate the function f^* from the random observations $(Y_i)_{i=1}^n$. Given a bandwidth h, consider the local averaging estimator

$$\widehat{f}_{n,h}(x) := \frac{1}{nh} \sum_{i=1}^n Y_i \mathbf{1}_{|x_i - x| \leq h}.$$

1. For any $x_0 \in [0,1]$, determine the minimax rate

$$\inf_{\widehat{\tau}_n} \sup_{f \in \Sigma(\beta)} \mathbb{E}\big[\, |\widehat{\tau}_n - f(x_0)|^2 \,\big]$$

2. Suppose that we have the prior information that the target function f^* satisfies

$$0 \leq f^*(x_0) \leq n^{-\alpha}, \quad \text{for any } x \in [0,1],$$

   for some known $\alpha \in (0, \beta/4)$.

   Based on this additional information, derive an improved rate of convergence for the estimator $\widehat{f}_{n,h}$ (which depends on the sample size n, and exponents α, β).

Q2. Maximal likelihood for density estimation

Throughout this question, we use the Hellinger distance

$$h^2(p_1, p_2) := \frac{1}{2} \int (\sqrt{p_1(x)} - \sqrt{p_2(x)})^2 dx$$

Given i.i.d. observations $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} p^* \in \mathcal{P}$ for some function class $\mathcal{P}$, we consider the maximal likelihood estimator

$$\widehat{p}_n := \arg\max_{p \in \mathcal{P}} \sum_{i=1}^{n} \log p(X_i).$$

Assume that the class $\mathcal{P}$ is a convex set. For simplicity, we also assume that x is one-dimensional and that the densities in $\mathcal{P}$ are supported on the interval $[0, 1]$.

1. Prove that

$$h^2(\widehat{p}_n, p^*) \leq \int \frac{2\widehat{p}_n(x)}{\widehat{p}_n(x) + p^*(x)} \Big(d\widehat{\mathbb{P}}_n(x) - d\mathbb{P}(x) \Big),$$

   where $\widehat{\mathbb{P}}_n := n^{-1} \sum_{i=1}^{n} \delta_{X_i}$ denotes the empirical distribution of the n samples.

   [Hint: you can prove and use the following technical inequality]

$$h^2(p_1, p_0) \leq \frac{1}{2} \int \frac{(p_1(x) - p_0(x))^2}{p_1(x) + p_0(x)} dx.$$

2. Let $\mathcal{F} := \{\log p : p \in \mathcal{P}\}$. Assuming that for any $\delta > 0$, we have

$$N(\delta; \mathcal{F}, \|\cdot\|_\infty) < +\infty,$$

   prove that $h^2(\widehat{p}_n, p^*) \to 0$ as $n \to +\infty$.

Q3. Two-dimensional kernel density estimation

Let p^* be a probability density function on $\mathbb{R}^2$, such that $p(x) \leq p_{\max}$ and $\|\nabla^3 p^*(x)\|_{\mathrm{op}} \leq 1$ for any $x \in \mathbb{R}^2$. Given i.i.d. samples $X_1, X_2, \cdots, X_n$ from p^*, define the kernel density estimator

$$\widehat{p}_{n,h}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big).$$

Find an appropriate choice of the kernel K and the bandwidth $h = h_n$, and establish the convergence rate

$$\mathbb{E}\big[|\widehat{p}_{n,h}(x_0) - p^*(x_0)|^2\big] \leq cn^{-\frac{3}{4}},$$

for some constant $c > 0$ independent of n.

Q4. Minimax risk for estimating under sup norm

Let $X_1, X_2, \cdots, X_n$ be i.i.d. samples from the uniform distribution over $[0,1]$. Conditionally on each X_i, we have

$$Y_i = f^*(X_i) + \varepsilon_i, \quad \text{where } \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

Let $\mathcal{F}$ be the following class

$$\mathcal{F} := \Big\{ f : [0,1] \to [0,1] \;\Big|\; |f(x) - f(y)| \le |x - y|, \quad \text{for any pair } x, y \in [0,1] \Big\}.$$

Prove that there exists a universal constant $c > 0$, such that

$$\inf_{\widehat{f}_n} \sup_{f^* \in \mathcal{F}} \mathbb{E}\Big[\Big\| \widehat{f}_n - f^* \Big\|_\infty^2 \Big] \ge c \Big(\frac{\log n}{n} \Big)^{2/3},$$

where $\|\cdot\|_\infty$ denotes the sup-norm: $\|f\|_\infty := \sup_{x \in [0,1]} |f(x)|$.