STAT 5460: Homework III (Technically II)

Sam Olson

Problem 1

Consider the kernel density estimator with $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X$:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i),$$

and denote

$$(f * g)(x) = \int f(x - y) g(y) \, dy.$$

(a)

Show that the exact bias of the kernel density estimator is given by

$$\mathbb{E}[\hat{f}(x)] - f(x) = (K_h * f)(x) - f(x).$$

b)

Show that the exact variance of the kernel density estimator equals

$$\text{Var}(\hat{f}(x)) = \frac{1}{n} \Big[(K_h^2 * f)(x) - (K_h * f)^2(x) \Big].$$

c)

Calculate the exact mean squared error (MSE) of the kernel density estimator.

d)

Calculate the exact mean integrated squared error (MISE) of the kernel density estimator.

Problem 2

a)

Use Hoeffding's inequality to bound the probability that the kernel density estimator $\hat{f}_h$ deviates from its expectation at a fixed point x, i.e., find an upper bound for

$$P\Big(\big|\hat{f}_h(x) - \mathbb{E}[\hat{f}_h(x)]\big| > \epsilon\Big)$$

for some ϵ, and show how the bound depends on n, h, ϵ and $\|K\|_\infty = \sup_{u\in\mathbb{R}} |K(u)| < \infty$.

Hint: Hoeffding's inequality states that for i.i.d. random variables Y_i such that $a \le Y_i \le b$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]\right| > \epsilon\right) \le 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$

b)

Suppose you want to construct a uniform bound over a compact interval $[a, b]$. Show that

$$P\left(\sup_{x\in[a,b]} \big|\hat{f}(x) - \mathbb{E}[\hat{f}_h(x)]\big| > \epsilon\right) \le \text{something small.}$$

Write down all the assumptions you're making in the process.

Hint: For a given $\delta > 0$, construct a finite set $N_\delta \subset [a, b]$ such that:

- For every $x \in [a, b]$, there exists $x' \in N_\delta$ with $|x - x'| \le \delta$
- $|N_\delta| \le \lceil \frac{b-a}{\delta} \rceil + 1$

c)

From Question b), construct a nonparametric uniform $1-\alpha$ confidence band for $\mathbb{E}[\hat{f}_h(x)]$, i.e., find $L(x)$ and $U(x)$ such that

$$P\big(L(x) \le \mathbb{E}[\hat{f}_h(x)] \le U(x),\ \forall x\big) \ge 1 - \alpha.$$