Computer-Intensive Statistics and Applications
Individual 3: Kernel Estimation and Regression

Chenglong Li

Problem 1

We consider a bin with bandwidth $h > 0$ as

$$B_j = [x_0 + (j-1)h, x_0 + jh],$$

where j is an integer, and x_0 is the origin of the histogram. For $x \in B_j$, we approximate $f(x)$ by $h^{-1}P(X \in B_j)$ and estimate $P(X \in B_j)$ by $n^{-1}\sum_{i=1}^{n} \mathbf{1}(X_i \in B_j)$. So the histogram is given by

$$\begin{aligned}
\hat{f}_h(x) &= \sum_j \mathbf{1}(x \in B_j)\frac{1}{h}[\sum_{i=1}^{n} \frac{1}{n}\mathbf{1}(X_i \in B_j)] \\
&= \frac{1}{nh}\sum_{i=1}^{n}\sum_j \mathbf{1}(X_i \in B_j)\mathbf{1}(x \in B_j).
\end{aligned}$$

It easy to know that

$$\begin{aligned}
\hat{f}_h(x) &\geq 0 \\
\int \hat{f}_h(x)dx &= 1.
\end{aligned}$$

Next, we consider a fixed x, then if $x \in B_j$

$$\begin{aligned}
\mathbb{E}[\hat{f}(x)] &= \frac{P(X \in B_j)}{h} \\
\mathrm{Var}[\hat{f}(x)] &= \frac{P(X_j \in B_j[1 - P(X_j \in B_j)])}{nh^2}.
\end{aligned}$$

We suppose that the true density $f(x)$ is smooth enough, let b_j be the midpoint of B_j. Taylor expansion for $x \in B_j$ yields

$$f(x) = f(b_j) + f'(b_j)(x - b_j) + o(h) \quad \text{as } h \to 0.$$

Hence,

$$\begin{aligned}
\mathrm{Bias} = \mathbb{E}[\hat{f}(x)] - f(x) &= f'(b_j)(b_j - x) + o(h), \\
\mathrm{Var}[\hat{f}(x)] &= \frac{1}{nh}f(x) + o\left(\frac{1}{nh}\right),
\end{aligned}$$

as $h \to 0$ and $nh \to \infty$. For a fixed x, the mean squared error satisfies

$$\begin{aligned}
\mathrm{MSE}(\hat{f}(x)) &= \text{Variance} + \text{Bias}^2 \\
&= \mathrm{Var}[\hat{f}(x)] + \left\{\mathbb{E}[\hat{f}(x)] - f(x)\right\}^2 \\
&= \frac{1}{nh}f(x) + [f'(b_j)]^2(b_j - x)^2 + o\left(\frac{1}{nh}\right) + o(h^2).
\end{aligned}$$

The mean integrated squared error (MISE) for the godness of estimation.

$$\text{MISE}(\hat{f}) = \mathbb{E}\left[\int [\hat{f}(x) - f(x)]^2 dx\right].$$

So we can know the MISE of the histogram is

$$\begin{aligned}\text{MISE}(\hat{f}) &= \int \mathbb{E}\left[(\hat{f}(x) - f(x))^2\right] dx \\ &= \int \text{MSE}(\hat{f}(x))dx\end{aligned}$$

So

$$\begin{aligned}\text{MISE} &= \int \text{Variance} + \text{Bias}^2 dx \\ &= \int \frac{f(x)}{nh}dx + \int [\text{Bias}]^2 dx + o\left(\frac{1}{nh}\right) + o(h)\end{aligned}$$

We know $\int f(x)dx = 1$, so we get

$$\int \frac{f(x)}{nh}dx = \frac{1}{nh}$$

For Bias part,

$$\begin{aligned}[\text{Bias}]^2 &= [f'(b_j)(x - b_j)]^2 + o(h) \\ &= f'(b_j)(x - b_j)^2 + o(h)\end{aligned}$$

Integrate within this bin and we set $u = x - b_j$

$$\int_{b_j - h/2}^{b_j + h/2} [Bias]^2 dx = \int_{-h/2}^{h/2} f'(b_j)u^2 du + o(h)$$

so we can know

$$\begin{aligned}\int_{-h/2}^{h/2} u^2 du &= \left[\frac{u^3}{3}\right]_{-h/2}^{h/2} \\ &= \frac{(h/2)^3 - (-h/2)^3}{3} \\ &= \frac{2 \cdot (h^3/8)}{3} \\ &= \frac{h^3}{12}.\end{aligned}$$

So for single bin

$$\int_{b_j - h/2}^{b_j + h/2} [Bias]^2 dx = f'(b_j)^2 \frac{h^3}{12} + o(h)$$

For the whole bins, we can get

$$\begin{aligned}\int [Bias]^2 dx &= \sum_j f'(b_j)^2 \frac{h^3}{12} + o(h) \\ &= \frac{h^3}{12} \sum_j f'(b_j)^2 + o(h) \\ &= \frac{h^3}{12} \frac{1}{h} \int f'(x)^2 dx + o(h) \\ &= \frac{h^2}{12} \|f'(x)\|_2^2 + o(h)\end{aligned}$$

where

$$\|f'(x)\|_2 = \sqrt{\int [f'(x)]^2 \, dx}.$$

So the MISE of the histogram is

$$\text{MISE}(\hat{f}) = \frac{1}{nh} + \frac{h^2}{12}\|f'(x)\|_2^2 + o\left(\frac{1}{nh}\right) + o(h),$$

where

$$\|f'(x)\|_2 = \sqrt{\int [f'(x)]^2 \, dx}.$$

The leading term of MISE is called the asymptotic MISE

$$\begin{aligned} \text{AIMSE}(\hat{f}) &= \frac{1}{nh} + \frac{h^2}{12}\|f'(x)\|_2^2 \\ &= \frac{1}{nh} + \frac{h^2}{12}\int [f'(x)]^2 \, dx. \end{aligned}$$

Problem 2

We have seen in the slides that, for each h in a pre-specified grid of candidate bandwidths,

$$CV(h) = \int \hat{f}_h^2(x)\, dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{h,-i}(X_i),$$

where

$$\hat{f}_{h,-i}(x) = \frac{1}{(n-1)h}\sum_{j\neq i} K\left(\frac{x-X_j}{h}\right).$$

We know

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

where $K(\cdot)$ is a suitably chosen function.
For the first term

$$\begin{aligned}\hat{f}^2(x) &= \left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\right]^2 \\ &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right)\end{aligned}$$

So we can know that

$$\begin{aligned}\int \hat{f}^2(x)dx &= \int \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right) dx \\ &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right) dx\end{aligned}$$

Now we consider $\int K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right) dx$, let $u = \frac{x-X_i}{h}$, so $x = hu + X_i$ and $dx = hdu$. So we can get that

$$\int K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right) dx = \int K(u)K(u + \frac{X_i - X_j}{h})hdu.$$

So

$$\begin{aligned}\int \hat{f}^2(x)dx &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right) dx \\ &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int K(u)K(u + \frac{X_i - X_j}{h})hdu \\ &= \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n} \int K(u)K(u + \frac{X_i - X_j}{h})du\end{aligned}$$

We set $\delta = \frac{X_i - X_j}{h}$, so

$$\begin{aligned}\int \hat{f}^2(x)dx &= \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n} \int K(u)K(u + \frac{X_i - X_j}{h})du \\ &= \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n} \int K(u)K(u + \delta)du\end{aligned}$$

For this trem $\int K(u)K(u+\delta)du$ we can use Fourier transform.

Problem 3

Suppose that we have observed a random sample $X_1, X_2, \cdots, X_n \in \mathbb{R}^d$. We consider a ball $B(x, \rho)$ with the center x and radius $\rho(x)$. For $x \in B(x, \rho)$, we still approximate $f(x)$ by

$$\frac{1}{\mathrm{Vol}(B(x,\rho))} P(X \in B(x,\rho)),$$

where $\mathrm{Vol}(B(x,\rho)) = \frac{\pi^{d/2}}{\Gamma(1+2^{-1}d)}\rho^d$ is the volume of $B(x,\rho)$. The k-nearest neighbor (kNN) estimator is

$$\begin{aligned}\hat{f}_k(x) &= \frac{1}{n\mathrm{Vol}(B(x,R_k(x)))}\left[\sum_{i=1}^{n}\frac{1}{n}\mathbf{1}(X_i \in B(x,R_k(x)))\right] \\ &= \frac{k}{n\mathrm{Vol}(B(x,R_k(x)))},\end{aligned}$$

where $R_k(x)$ be the distance of x to its kth nearest point. Now we consider a ball $B(0,1)$ with the center 0 and radius 1. So we can know

$$\mathrm{Vol}(B(x,\rho(x))) = \rho(x)^d \mathrm{Vol}(B(0,1)).$$

$$\mathrm{Vol}(B(x,R_k(x))) = R_k^d(x)\mathrm{Vol}(B(0,1))$$

So the k-nearest neighbor (kNN) estimator is

$$\begin{aligned}\hat{f}_k(x) &= \frac{k}{n\mathrm{Vol}(B(x,R_k(x)))} \\ &= \frac{k}{nR_k^d(x)\mathrm{Vol}(B(0,1))} \\ &= \frac{1}{nR_k^d(x)\mathrm{Vol}(B(0,1))}\sum_{i=1}^{n}\mathbf{1}\{\|X_i - x\| \le R_k(x)\}\end{aligned}$$

We compare with the formula of kernel kNN estimator is

$$\hat{f}_k(x) = \frac{1}{nR_k^d(x)}\sum_{i=1}^{n} K\left(\frac{x - X_i}{R_k(x)}\right).$$

So we can let $\frac{1}{\mathrm{Vol}(B(0,1))}\mathbf{1}\{\|X_i - x\| \le R_k(x)\} = K(\cdot)$ is a kernel function and let $u = \frac{x - X_i}{R_k(x)}$.

$$\begin{aligned}K(u) &= \frac{1}{\mathrm{Vol}(B(0,1))}\mathbf{1}\{\|u\| \le 1\} \\ K\left(\frac{x - X_i}{R_k(x)}\right) &= \frac{1}{\mathrm{Vol}(B(0,1))}\mathbf{1}\{\|X_i - x\| \le R_k(x)\}\end{aligned}$$

So the kernel kNN estimator is

$$\hat{f}_k(x) = \frac{1}{nR_k^d(x)}\sum_{i=1}^{n} K\left(\frac{x - X_i}{R_k(x)}\right).$$

Problem 4

If $X_i \in \mathbb{R}^1$, the kernel regression is

$$m(x) = E(Y \mid X = x) = \int y \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy$$

So we can estimate $f_X(x)$ and $\hat{f}_{(X,Y)}(x,y)$ by

$$\hat{f}_X(x) = \frac{1}{nh_x} \sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x})$$

$$\hat{f}_{(X,Y)}(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K_x\left(\frac{x - X_i}{h_x}\right) K_y\left(\frac{y - Y_i}{h_y}\right)$$

So

$$\begin{aligned}\hat{m}(x) &= \int y \frac{\hat{f}_{(X,Y)}(x,y)}{\hat{f}_X(x)} dy \\ &= \frac{\sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right) Y_i}{\sum_{i=1}^{n} K_x\left(\frac{x-X_i}{h_x}\right)}.\end{aligned}$$

We extend it to the multidimensional case $X_i \in \mathbb{R}$, so the multivariate kernel regression is

$$m(x) = E(Y \mid X = x) = \int y \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy$$

We can estimate $f_X(x)$ and $f_{(X,Y)}(x,y)$ by

$$\hat{f}_X(x) = \frac{1}{n \det(H)} \sum_{i=1}^{n} K_x\left[H^{-1}(x - X_i)\right]$$

$$\hat{f}_{(X,Y)}(x,y) = \frac{1}{nh \det(H)} \sum_{i=1}^{n} K_y\left(\frac{y - Y_i}{h}\right) K_x\left[H^{-1}(x - X_i)\right].$$

So

$$\begin{aligned}\hat{m}(x) &= E(Y \mid X = x) \\ &= \int y \frac{\hat{f}_{(X,Y)}(x,y)}{\hat{f}_X(x)} dy \\ &= \frac{\int y \hat{f}_{(X,Y)}(x,y) dy}{\hat{f}_X(x)}\end{aligned}$$

Let's look at the first half of the equation,

$$\begin{aligned}\int y \hat{f}_{(X,Y)}(x,y) dy &= \int y \left[\frac{1}{nh \det(H)} \sum_{i=1}^{n} K_y(\frac{y - Y_i}{h}) K_x[H^{-1}(x - X_i)]\right] dy \\ &= \frac{1}{nh \det(H)} \sum_{i=1}^{n} [K_x[H^{-1}(x - X_i)] \boxed{\int y K_y(\frac{y - Y_i}{h}) dy}].\end{aligned}$$

K_y is a kernel function, so we can know that

$$\int K_y(u) du = 1$$

$$\int u K_y(u) du = 0.$$

We let

$$K_y\left(\frac{y-Y_i}{h}\right)=\frac{1}{h}K\left(\frac{y-Y_i}{h}\right),$$

we set $u=\frac{y-Y_i}{h}$, so we can know $y=Y_i+hu$ and $dy=hdu$. Then

$$\begin{aligned}\int yK_y(\frac{y-Y_i}{h})dy &= \int (Y_i+hu)\frac{1}{h}K(u)hdu \\ &= \int (Y_i+hu)K(u)du \\ &= Y_i\end{aligned}$$

Since K_y is eliminated by intergration

$$\int y\hat{f}_{(X,Y)}(x,y)dy=\frac{1}{n\det(H)}\sum_{i=1}^{n}K_x[H^{-1}(x-X_i)]Y_i.$$

now we also know that

$$\hat{f}_X(x)=\frac{1}{n\det(H)}\sum_{i=1}^{n}K_x\left[H^{-1}(x-X_i)\right]$$

So the multivariate kernel regression is

$$\begin{aligned}\hat{m}(x) &= \int y\frac{\hat{f}_{(X,Y)}(x,y)}{\hat{f}_X(x)}\,dy \\ &= \frac{\frac{1}{n\det(H)}\sum_{i=1}^{n}K_x[H^{-1}(x-X_i)]Y_i}{\frac{1}{n\det(H)}\sum_{i=1}^{n}K_x\left[H^{-1}(x-X_i)\right]} \\ &= \frac{\sum_{i=1}^{n}K_x[H^{-1}(x-X_i)]Y_i}{\sum_{i=1}^{n}K_x\left[H^{-1}(x-X_i)\right]}\end{aligned}$$

$$\hat{m}_H(x)=\frac{\sum_{i=1}^{n}K\left[H^{-1}(x-X_i)\right]Y_i}{\sum_{i=1}^{n}K\left[H^{-1}(x-X_i)\right]}$$

Problem 5

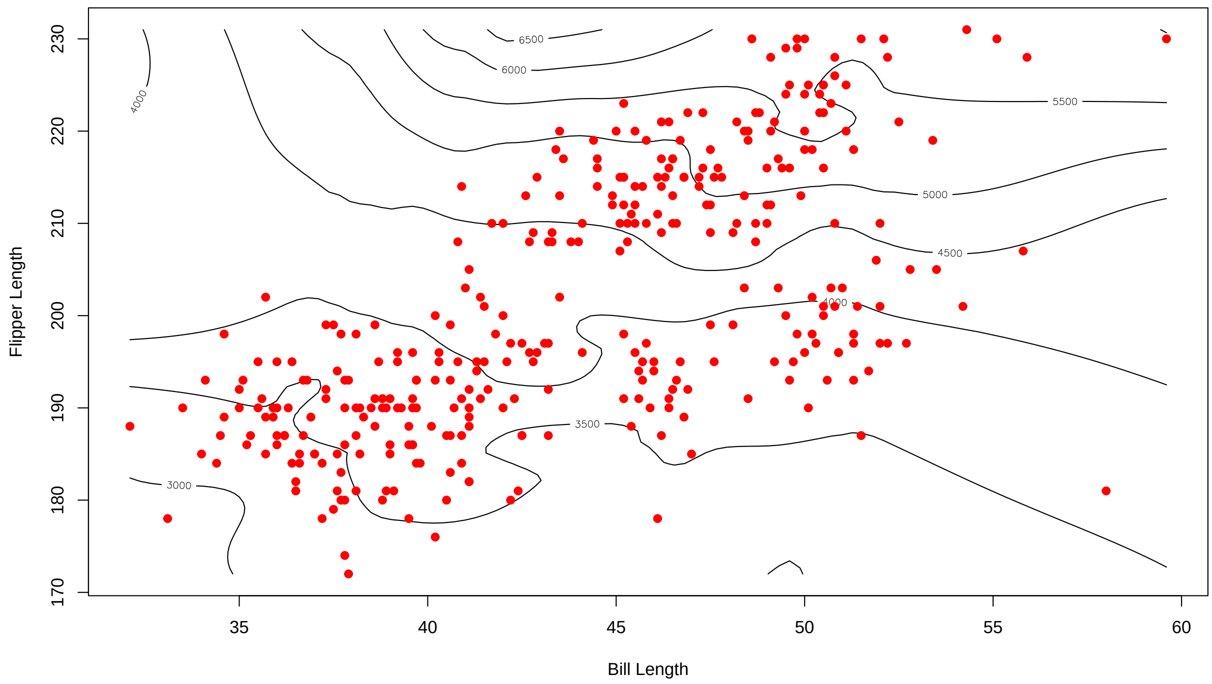

Figure 1: Contour Plot of Predicted Body Mass

```
library(ggplot2)
library(plotly)
library(akima)
# loading data
data <- read.csv("/Users/max/CISA_Work/Penguin.csv")
body_mass <- data$body_mass
bill_length <- data$bill_length
flipper_length <- data$flipper_length

# Local Linear Regression
local_linear_regression <- loess(body_mass ~ bill_length + flipper_length ,data=data, degree = 1, span = 0.1) # degree 1 is linear

# create new data
bill_seq <- seq(min(bill_length), max(bill_length), length.out = 100)
flipper_seq <- seq(min(flipper_length), max(flipper_length), length.out = 100)
grid_points <- expand.grid(bill_length = bill_seq, flipper_length = flipper_seq)
# prediction
prediction <- predict(local_linear_regression, newdata = grid_points)

z_matrix <- matrix(prediction,
                   nrow = length(bill_seq),
                   ncol = length(flipper_seq))

contour(
  x = bill_seq,
  y = flipper_seq,
  z = z_matrix,
  xlab = "Bill Length",
  ylab = "Flipper Length",
)

points(bill_length, flipper_length, pch = 19, col = "red")
```