MATH524 – Spring 2025

Problem Set: Week 3

1. (**Optimal kernel**) Consider the second-order Epanechnikov kernel defined as

$$K_E(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{x^2}{5}\right)\mathbf{1}_{\{|x| \le \sqrt{5}\}},$$

and note that $\int |u|^2|K_E(u)|\,\mathrm{d}u = 1$. Let K_0 be another non-negative second-order kernel with $\int |u|^2|K_0(u)|\,\mathrm{d}u = 1$. By considering $e(x) = K_0(x) - K_E(x)$, or otherwise, show that the Epanechnikov kernel always has lower risk than any K_0. That is, $R(K_0) \ge R(K_E)$. **Solution**: Note that

$$\int_{-\infty}^{\infty} e(x)\ \mathrm{d}x = \int_{-\infty}^{\infty} x^2 e(x)\ \mathrm{d}x = 0,$$

and $e(x) \ge 0$, for $|x| \ge \sqrt{5}$. Now, using the fact that both kernels are non-negative second-order,

$$\begin{aligned} R(K_0) - R(K_E) &= \int_{-\infty}^{\infty} K_0(x)^2 - K_E(x)^2\ \mathrm{d}x \\ &= \int_{-\infty}^{\infty} e(x)(K_0(x) + K_E(x))\ \mathrm{d}x \\ &= R(e) + 2\int_{-\infty}^{\infty} e(x)K_E(x)\ \mathrm{d}x. \end{aligned}$$

But,

$$\int_{-\infty}^{\infty} e(x)K_E(x)\ \mathrm{d}x = \frac{3}{4\sqrt{5}}\int_{-\sqrt{5}}^{\sqrt{5}}\left(1 - \frac{x^2}{5}\right)e(x)\ \mathrm{d}x = \frac{3}{4\sqrt{5}}\int_{|x|>\sqrt{5}}\left(\frac{x^2}{5} - 1\right)e(x)\ \mathrm{d}x \ge 0.$$

The result follows.

2. (**Linear Smoothers, Cross-validation**)

Let $\{(y_i, x_i) : 1 \le i \le n\}$ be a random sample taking values in $\mathbb{R}^2$. A linear smoother is given by

$$\hat{e}(x) = \sum_{i=1}^{n} w_{n,i}(x)y_i, \qquad w_{n,i}(x) = w(x_1, x_2, \cdots, x_n; x).$$

Note that $w_{n,i}(x)$ is only a function of $\{x_i : 1 \le i \le n\}$ and not of $\{y_i : 1 \le i \le n\}$. Recall that local polynomial regression takes on the following form

$$\hat{e} = \mathbf{e}_0' \arg\min_{e} \sum_{i=1}^{n}(y_i - p(x_i - x)'e)^2 K_h(x_i - x)$$

where $\mathbf{e}_0$ is the first basis unit vector and $p(x) = (1, x, x^2, \ldots, x^p)'$ is the polynomial basis up to order p.

(a) Show that local polynomial regression estimators can be written as linear smoothers and give the exact form of the "smoothing weights" $w_{n,i}(x)$.

Solution: Let

$$X = \begin{bmatrix} 1 & x_1 - x & \cdots & (x_1 - x)^p \\ 1 & x_2 - x & \cdots & (x_2 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & x_n - x & \cdots & (x_n - x)^p \end{bmatrix}, \qquad W = \mathrm{diag}\Big\{K_h(x_i - x),\ 1 \le i \le n\Big\}, \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

and let $\mathbf{e}_j$ be the $(j+1)$-th unit basis vector. Then the estimator is given by

$$\begin{aligned}\hat{e}(x) &= \mathbf{e}_0' \left(X'WX\right)^{-1} X'WY \\ &= \sum_i \mathbf{e}_0' \left(X'WX\right)^{-1} p(x_i - x) K_h(x_i - x) y_i \\ &= \sum_i w(x_i - x) y_i,\end{aligned}$$

where $w(x_i - x) = \mathbf{e}_0' \left(X'WX\right)^{-1} p(x_i - x) K_h(x_i - x)$.

(b) Show the following simplified cross-validation formula holds for local polynomial regression. [1]

$$\mathsf{CV}(c) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{e}_{(i)}(x_i)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{e}(x_i)}{1 - w_{n,i}(x_i)}\right)^2,$$

where $\hat{e}_{(i)} = \sum_{j\neq i} w_{n,j}(x) y_j$ is the leave-one-out estimator and c denotes a tuning parameter (i.e., a bandwidth h_n for local polynomials).

Solution:

Note that $\sum_k X_k' X_k = X'X$. Thus, by the hint, and the fact that W is a diagonal matrix,

$$\begin{aligned}(X'WX)^{-1} &= (X_{(i)}'W_{(i)}X_{(i)} + W_{ii}X_i'X_i)^{-1} \\ &= (X_{(i)}'W_{(i)}X_{(i)})^{-1} - \frac{W_{ii}(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}}{1 + W_{ii}X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'}.\end{aligned}$$

Then,

$$\begin{aligned}(X'WX)^{-1}X_i' &= \left((X_{(i)}'W_{(i)}X_{(i)})^{-1} - \frac{W_{ii}(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}}{1 + W_{ii}X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'}\right)X_i' \\ &= \frac{(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'}{1 + W_{ii}X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'}\end{aligned}$$

The leave-in predicted value at x is

$$\begin{aligned}\hat{e}(x_i) &= \mathbf{e}_0'(X'WX)^{-1}X'WY \\ &= \mathbf{e}_0'(X'WX)^{-1}X_i'W_{ii}Y_i + \mathbf{e}_0'(X'WX)^{-1}X_{(i)}'W_{(i)}Y_{(i)} \\ &= wy_i + \mathbf{e}_0'\left((X_{(i)}'W_{(i)}X_{(i)})^{-1} - \frac{W_{ii}(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}}{1 + W_{ii}X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'}\right)X_{(i)}'W_{(i)}Y_{(i)} \\ &= wy_i + \left(1 - \frac{\mathbf{e}_0'W_{ii}(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'X_i}{1 + W_{ii}X_i(X_{(i)}'W_{(i)}X_{(i)})^{-1}X_i'}\right)\hat{e}_{(i)} \\ &= wy_i + (1-w)\hat{e}_{(i)}(x_i).\end{aligned}$$

Hence,

$$\hat{e}(x_i) = wy_i + (1-w)\,\hat{e}_{(i)}(x_i),$$

[1] The following result is useful: for an invertible matrix $\mathbf{A}$ and a column vector $\mathbf{v}$, and $\lambda \neq -1/(\mathbf{v}'\mathbf{A}\mathbf{v})$ the following holds

$$\left(\mathbf{A} + \lambda\mathbf{v}\mathbf{v}'\right)^{-1} = \mathbf{A}^{-1} - \frac{\lambda\mathbf{A}^{-1}\mathbf{v}\mathbf{v}'\mathbf{A}^{-1}}{1 + \lambda\mathbf{v}'\mathbf{A}^{-1}\mathbf{v}}.$$

or equivalently

$$\frac{y_i - \hat{e}(x_i)}{1 - w(x_i - x)} = y_i - \hat{e}_{(i)}(x_i),$$

which justifies the simplified cross-validation formula.

(c) Providing regularity conditions, show that

$$\frac{\hat{e}(x) - e(x)}{\sqrt{\mathbb{V}[\hat{e}(x)|x_1, x_2, \cdots, x_n]}} \to_d \mathcal{N}(0, 1).$$

where $e(x) = \mathbb{E}[Y|X = x]$

Solution:

Note that the estimator is written as a weighted sum

$$\hat{e}(x) = \sum_i w(x_i - x) y_i,$$

and the weights $w(x_i - x)$ depends only on the covariates. Therefore we consider the centered quantities

$$\sum_i z_i, \qquad z_i = (\mathbb{V}[\hat{e}(x) \ |x_i, 1 \le i \le n])^{-\frac{1}{2}} \ w(x_i - x)\Big(y_i - \mathbb{E}[y_i \ |x_i, 1 \le i \le n]\Big),$$

then it is easy to see that

$$\mathbb{E}[z_i \ |x_i, 1 \le i \le n] = 0,$$

and

$$\begin{aligned}
\mathbb{V}\left[\sum_i z_i \ \mid x_i, 1 \le i \le n\right] &= \mathbb{V}\left[\sum_i (\mathbb{V}[\hat{e}(x) \ |x_i, 1 \le i \le n])^{-\frac{1}{2}} \ w(x_i - x) y_i \ |x_i, 1 \le i \le n\right] \\
&= (\mathbb{V}[\hat{e}(x) \ |x_i, 1 \le i \le n])^{-1} \sum_i w(x_i - x)^2 \mathbb{V}[y_i \ |x_i, 1 \le i \le n] \\
&= 1.
\end{aligned}$$

Assume for some $\epsilon > 0$,

$$\sum_i \mathbb{E}[z_i^{2+\epsilon} \ |x_i, 1 \le i \le n] \to 0,$$

then the Lindeberg-Feller CLT implies

$$\sum_i z_i \to_d \mathcal{N}(0, 1),$$

Finally note that

$$\frac{\hat{e}(x) - e(x)}{\sqrt{\mathbb{V}[\hat{e}(x)|x_i, 1 \le i \le n]}} = \sum_i z_i + \frac{\mathbb{E}[\hat{e}(x) \ |x_i, 1 \le i \le n] - e(x)}{\sqrt{\mathbb{V}[\hat{e}(x) \ |x_i, 1 \le i \le n]}},$$

so we need the condition to ensure bias would not show up asymptotically (this is generally achieved by undersmoothing)

$$\frac{\mathbb{E}[\hat{e}(x) \ |x_i, 1 \le i \le n] - e(x)}{\sqrt{\mathbb{V}[\hat{e}(x) \ |x_i, 1 \le i \le n]}} \to 0.$$

(d) Propose an asymptotically valid 95% confidence interval for $e(x)$, with x fixed. That is,

$$\forall x: \liminf_n \mathbb{P}\left[e(x) \in \mathsf{C.I.}(x)\right] \geq 0.95.$$

Is this derived confidence interval equivalent to the uniform confidence band? That is, does it satisfy the following probability expression?

$$\liminf_n \mathbb{P}\left[\forall x: \; e(x) \in \mathsf{C.I.}(x)\right] \geq 0.95?$$

Explain your answer.

Solution: Assume we have a consistent estimate of the conditional variance, i.e.

$$\frac{\hat{\mathbb{V}}[\hat{e}(x) \; |x_i, 1 \leq i \leq n]}{\mathbb{V}[\hat{e}(x) \; |x_i, 1 \leq i \leq n]} \to 1,$$

then a valid 95% C.I. would be

$$\mathsf{C.I.}(x) = \left[\hat{e}(x) \pm 1.96 \times \sqrt{\hat{\mathbb{V}}[\hat{e}(x) \; |x_i, 1 \leq i \leq n]}\right].$$

This is a pointwise confidence interval. In general, this is not equivalent to the uniform confidence interval that satisfies $\liminf_n \mathbb{P}\left[\forall x: \; e(x) \in \mathsf{C.I.}(x)\right] \geq 0.95$.

(e) Conduct the following Monte Carlo experiment. You are free to use inbuilt commands or libraries for matrix operations, dataframe structures, quantile calculations and plotting, but should *not* use any pre-packaged local polynomial regression implementations.

Consider the following DGP

- $x_i \sim \mathsf{Uniform}(-1, 1)$;
- $y_i = 0.3x_i^2 - 1.5x_i^3 + 0.2x_i^4 - 0.002x_i^5 + \varepsilon_i$;
- $\varepsilon_i \sim \mathcal{N}(0, 0.1^2)$,
- Consider the second-order Epanechnikov kernel, $K(u) = \frac{3}{4}(1 - u^2)$ for $-1 \leq u \leq 1$.

The dataset generated by this process is provided as a CSV on Moodle, named `exercise3.csv`.

The first column of the CSV contains the y_i's and the second column contains the x_i's.

i. Consider a degree 3 (p = 3) local polynomial estimator of $\mu(x)$, that is, $\hat{e}(x_i)$. Plot the $\mathsf{CV}(h)$, as a function of h and compute the CV estimator, denoted $\hat{h}_{\mathsf{CV}}$. Use h between 0.5 and 1.0 with 0.1 increments.

   Solution: See code in ex_03_sols.R and Figure 1.

ii. Using the data-driven tuning parameter choice $\hat{h}_{\mathsf{CV}}$, plot the following functions of $x \in [-1, 1]$ (in one single graph): (i) the true regression function; (ii) the estimated regression function $\hat{e}(x)$; (iii) the data. (Using a grid of 10 evaluation points should be enough.)

   Solution: See code in ex_03_sols.R and Figure 2

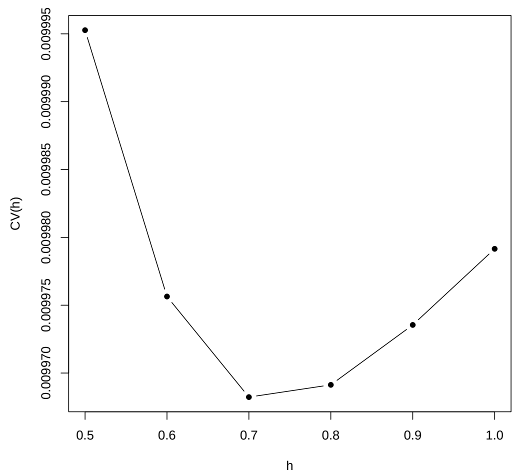

Figure 1: Cross validation for $h \in [0.5, 1.0]$

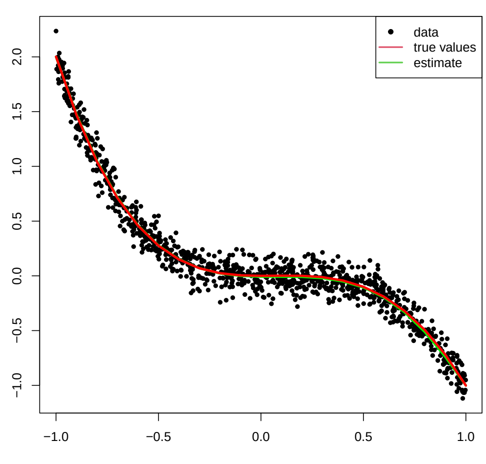

Figure 2: Local polynomial estimate with $\hat{h}_{\mathtt{CV}} = 0.7$ and true values