STAT 460/560- Assignments

Week VI:

Feel free to give hints to your classmates (and to me) for coding related problems.

1. Show that the local polynomial kernel estimator of $\boldsymbol{\beta}_x$ (as defined in Section 8.3 in the Lecture Notes) is given by

$$\hat{\boldsymbol{\beta}}_x = (\mathbf{Z}_x^\tau \mathbf{W}_x \mathbf{Z}_x)^{-1} \mathbf{Z}_x^\tau \mathbf{W}_x \mathbf{y}_n,$$

where the weight matrix is a diagonal matrix of the kernels

$$\mathbf{W}_x = \text{diag}\{K_h(x_1 - x), K_h(x_2 - x), ..., K_h(x_n - x)\}.$$

2. (a) Use the data `cps71` from the package `np` in `R` to fit the following kernel regressions (refer to Section 8.3 in the Lecture Notes for details) with $x =$ age and $y =$ logwage .

For $x = 25, 35, 50$,

(i) obtain the local constant kernel estimator of $\boldsymbol{\beta}_x$, i.e., $p = 0$.

(ii) obtain the local linear kernel estimator of $\boldsymbol{\beta}_x$, i.e., $p = 1$.

(iii) obtain the local quadratic kernel estimator of $\boldsymbol{\beta}_x$, i.e., $p = 2$.

Report these vectors. What are your predicted values at these x?

Remark: Use the density function of the standard Normal distribution $N(0, 1)$ as the kernel and let bandwidth $h = n^{-1/5}$, where n the number of observations.

(b) Plot the data, along with the three fitted regression functions $\hat{g}(x)$ in part (a). Use grid of length 0.5 over the range x, do not include the boundaries.

Write and attach largely your own codes. Feel free to collaborate and share with your friends but indicate if you did.

3. Using the same data `cps71` as in Question #1, obtain the bandwidth recommended by the leave-one-out cross-validation (CV) and generalized cross-validation (GCV), as follow.

   (a) For local constant kernel estimator of $\boldsymbol{\beta}_x$, find the bandwidth within the range of $h \in (0.02, 5)$.

   (b) For local linear kernel estimator of $\boldsymbol{\beta}_x$, find the bandwidth within the range of $h \in (0.02, 5)$.

   You may use the `R` function *bw.cv*() or other codes if you can locate them. Explicitly show your source and codes.

4. Give the expression of the natural cubic spline $s(x)$ with the following knots and values

$$
\begin{aligned}
t_1 &= 1, \ s(t_1) = 5 \\
t_2 &= 2, \ s(t_2) = 17 \\
t_3 &= 5, \ s(t_3) = 156.5.
\end{aligned}
$$

   Plot the natural cubic spline in the domain $x \in [1, 5]$.

5. Fit a cubic spline to the data `cps17` as in Question #1 and #2 with $\lambda = 2$.

   Plot the fitted natural spline over the range of x.