

Problem Set 12

Associated Reading: Chapter 11: Introduction - 11.2.1.2, 11.2.3, 11.2.5 - 11.2.5.1, 11.4, 11.4.2, 11.6

Complete the problems either by hand or using the computer and upload your final document to the Blackboard course site. All final submittals are to be in PDF form. Please document any code used to solve the problems and include it with your submission.

1. Posted on the course Blackboard site is the data set `smooth.txt` containing 100 bivariate observations, (X, Y) . (Note: X and Y are listed in two separate groups, however their order matters). Smooth this data with a constant-span running mean smoother. Select a span of $2k + 1$ for $1 \leq k \leq 11$ chosen by cross validation and use truncated neighborhoods near the edges of the data. As part of your solution you should include the following:
 - (i) A plot of $CVRSS_k(\hat{s}_k)$ versus k and an explanation of why you chose the span you did. (Hint: You can use equation 11.18 to greatly simplify these calculations.)
 - (ii) A plot of your optimal smooth (as a line) overlaid on the data (as points). Keep in mind that you only need to calculate the smooth at the points x_i and then you can use your plotting software to interpolate the line. For example in R you can use `plot(X, \hat{s}_k , type="l")`. (Hint: Don't forget to sort the data (x_i, y_i) as a pair and with respect to the x_i 's.)
2. Consider developing a constant-span running median smoother.
 - (a) Give a formula for computing $\hat{s}_k(x_i)$.
 - (b) Is this a linear smoother? If so, what is the smoothing matrix S ? If not, explain why not.
 - (c) Smooth the `smooth.txt` data using this smoother and the optimal span you found for a constant-span running mean smoother in Problem 1. Be sure to include a plot of the smooth and the data points as before. How does this smooth compare to the running mean smooth? Is this still the optimal span in $1, \dots, 11$ for a running median smooth? (You can make your argument with pictures and do not need to compute the $CVRSS_k(\hat{s}_k)$.)
3. (a) Smooth the `smooth.txt` data using a normal kernel smoother. Select the optimal bandwidth from $h \in \{.1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0, 1.1\}$ visually. Again, you only need to calculate the smooth at the points x_i and include a plot of the smooth and the data points.
 - (b) In class it was stated that Kernel smoothers are linear smoothers. So, for your smooth, what is row i of the smoothing matrix S .