Jared Yu

Computational Statistics

Problem Set 12

1. Posted on the course Blackboard site is the data set smooth.txt containing 100 bivariate observations, . (Note: and are listed in two separate groups, however their order matters). Smooth this data with a constant-span running mean smoother. Select a span of for chosen by cross validation and use truncated neighborhoods near the edges of the data. As part of your solution you should include the following:
   1. A plot of versus and an explanation of why you chose the span you did. (Hint: You can use equation 11.18 to greatly simplify these calculations.)
   2. 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 and then you can use your plotting software to interpolate the line. For example in R you can use “plot(, , type=”l”)”. (Hint: Don’t forget to sort the data as a pair and with respect to the .)
2. Consider developing a constant-span running median smoother.
   1. Give a formula for computing .
   2. Is this a linear smoother? If so, what is the smoothing matrix ? If not, explain why not.
   3. 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 for a running median smooth? (You can make your argument with pictures and do not need to compute the .)
3. …
   1. Smooth the smooth.txt data using a normal kernel smoother. Select the optimal bandwidth from visually. Again, you only need to calculate the smooth at the points and include a plot of the smooth and the data points.
   2. In class it was stated that Kernel smoothers are linear smoothers. So, for your smooth, what is row of the smoothing matrix .