Applied Statistics 2013

Homework 6

Exercise 1. *This is an adaptation of exercise 3 of chapter 5 in the Wasserman book: Get the data on fragments of glass collected in forensic work from the course website. Let Y be refractive index and let x be aluminum content. Experiment with various linear smoothing methods, such as local polynomials and smoothing splines. Use different values of bandwidth, different kernels, etc... In particular you should try the following:*

a) Perform a nonparametric regression to the model $Y_i = r(x_i) + \epsilon_i$. Use local linear estimation with smoothing parameter $nn \in (0, 1)$. A smoothing parameter equal to 0.6 means that the bandwidth is chosen so that a local neighborhood always contains about 60% of the total number of data points. Choose the smoothing parameter using generalized cross-validation. Make a picture of the generalized cross validation score against the smoothing parameter.

b) What is the optimal nearest-neighbor fraction?

Note: *Most of the demo code on the course website can used here.*

Exercise 2. *Suppose we perform local polynomial fitting with degree p. So, for example, if $p = 2$ we do local quadratic fitting. Let Q be a polynomial of degree $\leq p$.*

a) Suppose all points (x_i, Y_i) satisfy $Q(x_i) = Y_i$. Prove that for all x, the local polynomial fit $\hat{r}_n$ satisfies $\hat{r}_n(x) = Q(x)$. We say that local polynomial fitting with degree p reproduces polynomials of degree $\leq p$.
Hint: *start with equation (5.55) from Wasserman. Then use $Y_i = Q(x_i)$ and the fact that Q is a polynomial of degree at most p.*

b) Verify this with a small example in `R`.

c) Prove that

$$Q(x) = \sum_{i=1}^{n} \ell_i(x) Q(x_i) \ .$$