

Part 2b A Review of Nonparametric Regression

Although nonparametric regression has earlier roots, the first “officially” recognized attempt is credited to Nadaraya (1964) and Watson (1964). Their estimator was given by

$$\hat{q}_h(x_i) = \frac{\sum_{j=1}^n K_h(X_j - x_i) Y_j}{\sum_{j=1}^n K_h(X_j - x_i)}$$

where $K_h(u) = \frac{K(u/h)}{h}$ is the kernel function having bandwidth h (a nonnegative number controlling the size of the local neighborhood), and $K(u)$ is the kernel density function (Fan and Gijbels (1996)). The kernel function is a real valued weight function, and in practice it is most often assigned a symmetric density function. The most common choices are the Epanechnikov (1969) kernel

$$K(u) = .75(1 - u^2)I_u(-1,1)$$

and the Gaussian kernel

$$K(u) = \frac{e^{-u^2/2}}{\sqrt{2\mathbf{p}}} I_u(\mathfrak{R}).$$

Mays (1995) gives a brief discussion of the efficiencies of various kernels. He notes that the kernel selection is not terribly important to the quality of fit.

In spite of the presence of other general types of estimators (Gasser and Muller (1979)), the Nadaraya-Watson type estimator still stands out for two primary reasons. First, it brought about the notion of a kernel function. One early interpretation of the term “kernel regression” referred to any nonparametric regression utilizing a kernel function. Second, by recognizing the Nadaraya-Watson estimator as a local constant fit one might extend the idea to local polynomial fitting.

Hence, in the situation in which we are fitting a p th degree polynomial, we minimize (at each i)

$$\sum_{j=1}^n (Y_j - \sum_{d=0}^p \mathbf{b}_d (X_j - x_i)^d)^2 K_h(X_j - x_i).$$

Consequently, in the wake of the Nadaraya-Watson development this concept was studied intensively by a number of researchers. Stone (1977), and Cleveland (1979) were among those who presented early studies in local polynomial fitting. Many have followed, and the issue of degree selection has become an integral topic of nonparametric regression. Although the issue has not been firmly decided, generally a degree selection of $p = 0, 1$, or 2 is a safe decision, and it has been shown that odd values of p outperform even values asymptotically. For more on this topic see Fan and Gijbels (1996), and Ruppert and Wand (1994).

In spite of the existence of estimators other than local polynomial smoothers (like spline and nearest neighbor estimates), we will turn our attention to a second critical topic which is germane to local polynomial regression and that is bandwidth selection. This issue has been continually addressed since 1964 and is critical to any serious investigation of nonparametric regression. The user is faced with the problem of selecting a bandwidth that is too large and inflating the estimate's bias, or too small and inflating the estimate's variance. Both will affect the mean squared error (MSE) of the estimate of mean response. There are essentially two types of bandwidth selection procedures that are considered, constant or variable bandwidth selection. In local polynomial modeling, this amounts to fitting the model using a fixed interval (at each x_i) for the kernel function, or varying the interval according to factors involving x_i or y_i or both.

Constant bandwidths are normally chosen for data sets for which the design density is uniform, and/or the mean response has no apparent great change of form. Variable bandwidths are often chosen when one of these conditions is violated. Ruppert, Sheather and Wand (1995) presented two constant bandwidth selectors, the direct plug in (DPI) selector and the solve the equation (STE) selector.

In addition Ruppert (1995) also presented a variable bandwidth selector, the empirical bias bandwidth selector (EBBS). Finally, it is worth noting that bandwidth selection really has to do with how closely the user wishes the curve to fit points on the graph. Essentially, smaller bandwidths tend to fit the data points closely, while larger bandwidths produce fits to the data more akin to those obtained via parametric regression.

In conclusion, the nonparametric regression estimate provides a positive alternative to the parametric estimate, but we must be cognizant of the fact that in the end, we have a winding path passing roughly through the middle of the ordered data points. Slower convergence to the true mean response function, and less information concerning functional form are negative artifacts of nonparametric regression. Consequently, nonparametric regression does not provide a complete answer to the regression problem.

Without reservation we recommend using a semiparametric approach; specifically MRR1 and MRR2, and shall endeavor to demonstrate that (in addition to having small sample prowess) these methods are successful asymptotically too. We will begin by outlining the development of model robust regression in the next section.