MA5121: Nonparametric Statistics

Satvik Saha, 19MS154 October 3, 2023

1 Introduction

We examine the Glass Identification Dataset, and attempt to regress Refractive Index (RI) against Aluminium content (Al). To do so, we employ nonparametric methods such as kernel, local linear, and spline regression.

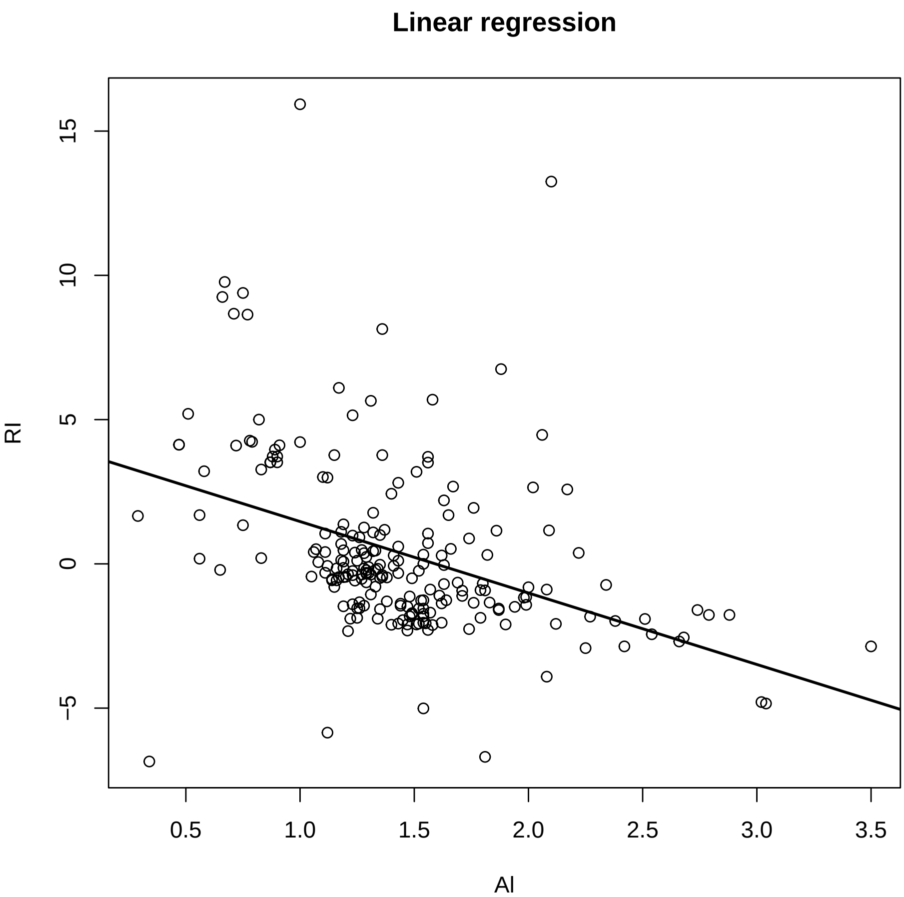

Figure 1: A scatter plot of the given data. A linear regression line has been drawn to illustrate the general trend in the data, and to emphasize the need for nonparametric methods.

2 Kernel regression

The Nadaraya-Watson estimator is of the form

$$\widehat{m}(x) = \sum_{i=1}^{n} \ell_i(x)\, y_i, \qquad \ell_i(x) = \frac{K((x - x_i)/h)}{\sum_{j=1}^{n} K((x - x_j)/h)}.$$

Here, we choose a Gaussian kernel $K(t) \propto e^{-t^2/2}$.

Indeed, it can be shown that

$$\widehat{m}(x) = \arg\min_{a} \sum_{i=1}^{n} w_i(x)(y_i - a)^2,$$

where we choose weights $w_i(x) = K((x - x_i)/h)$.

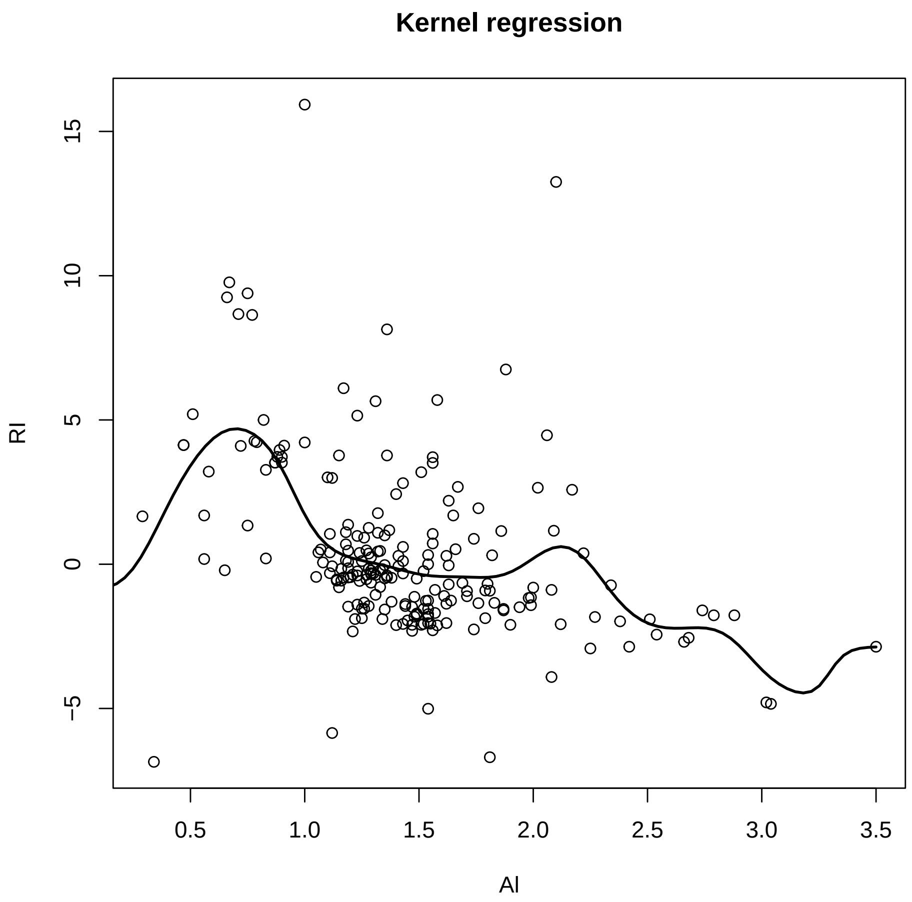

Figure 2: Kernel regression, using the Gaussian kernel. Here, $h = 0.135$. We estimate $\widehat{\sigma}^2 \approx 6.53$.

3 Local linear regression

We estimate

$$\widehat{m}(x) = \widehat{a}_0(x),$$

where

$$\widehat{\boldsymbol{a}}(x) = \arg\min_{\boldsymbol{a}} \sum_{i=1}^{n} w_i(x)(y_i - a_0 - a_1(x_i - x))^2.$$

Note that by setting

$$X = \begin{bmatrix} 1 & x_1 - x \\ 1 & x_2 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix}, \qquad W = \operatorname{diag}(w_1(x), \ldots, w_n(x)),$$

we have

$$\widehat{\boldsymbol{a}} = (X^\top W X)^- X^\top W \boldsymbol{y}.$$

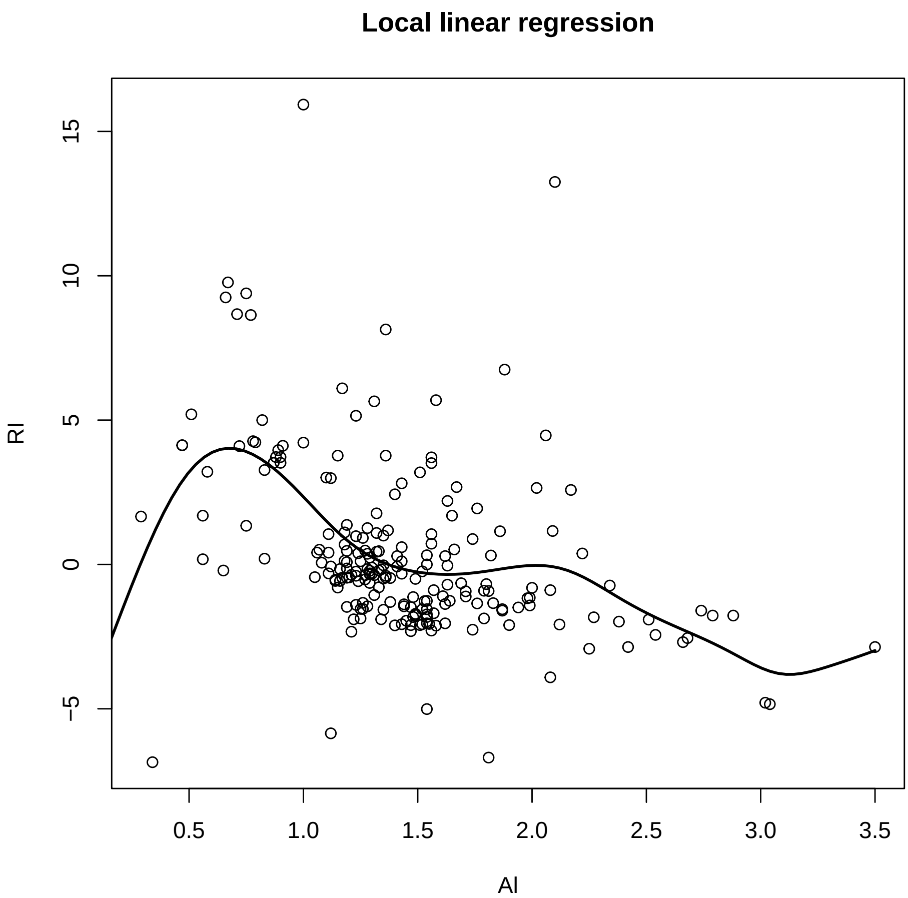

Figure 3: Local linear regression, using the Gaussian kernel. Here, $h = 0.260$. We estimate $\widehat{\sigma}^2 \approx 6.41$.

4 Cubic spline regression

Given a B-spline basis $\{B_i\}$ where all unique x_i's are chosen as knots, we estimate

$$\widehat{m}(x) = \sum_{j=1}^{n+4} \widehat{\beta}_j B_j(x).$$

Here, $\widehat{\boldsymbol{\beta}}$ is chosen so as to minimize a penalized loss, i.e.

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - B\boldsymbol{\beta}\|^2 + \lambda \boldsymbol{\beta}^\top \Omega \boldsymbol{\beta},$$

where

$$B_{ij} = B_j(x_i), \qquad \Omega_{ij} = \int B_i'' B_j''.$$

The parameter $\lambda > 0$ introduces a roughness penalty on $\widehat{m}$.

Indeed,

$$\widehat{\boldsymbol{\beta}} = (B^\top B + \lambda\Omega)^- B^\top \boldsymbol{y}.$$

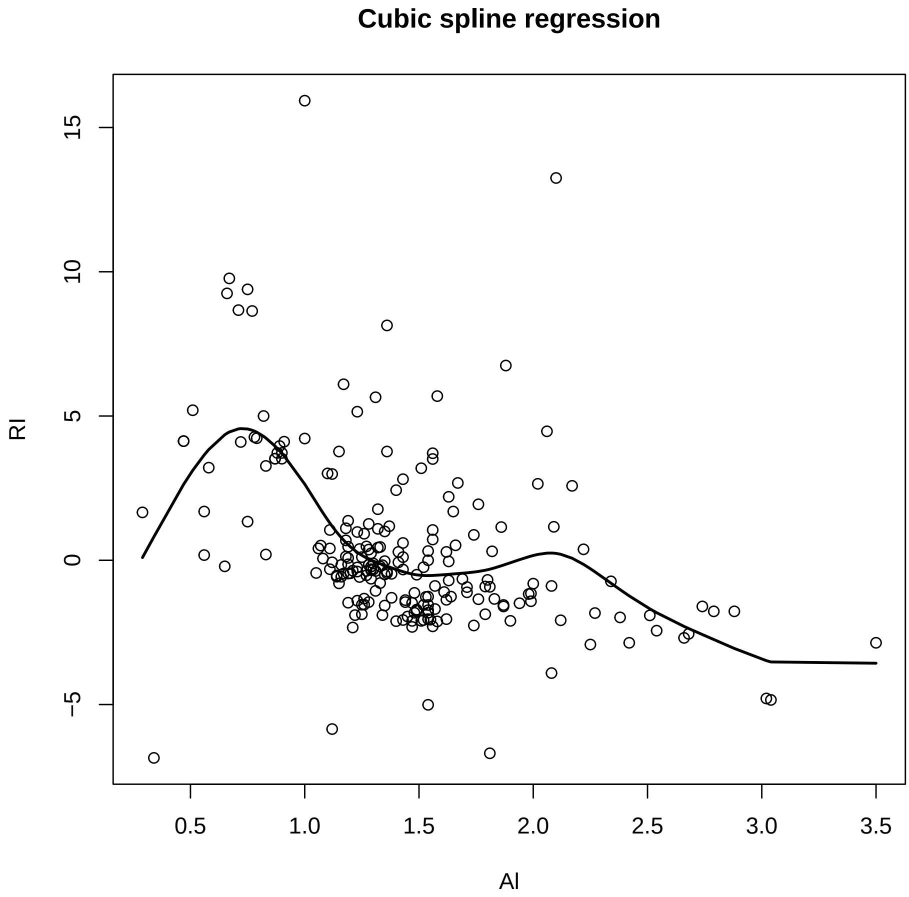

Figure 4: Cubic spline regression, using all unique values of 'Al' as knots. Here, the smoothing parameter $\lambda = 3.5 \times 10^{-6}$. We estimate $\widehat{\sigma}^2 \approx 6.42$.

5 Tuning parameters

In order to choose the parameters h and λ, we use Ordinary Cross-Validation (OCV).

Note that in all three cases, we have used *linear smoothers*, in the sense that

$$\widehat{m}(x) = L(x)\,\boldsymbol{y} = \sum_{i=1}^{n} \ell_i(x) y_i.$$

With this, the cross-validation score can be expressed as

$$OCV = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \widehat{m}(x_i)}{1 - L_{ii}} \right)^2,$$

where $L_{ii} = \ell_i(x_i)$. The parameter in question is chosen by minimizing the above.

6 Estimation of variance

With the setup of linear smoothers, a consistent estimator of the variance σ^2 (under certain conditions) is given by

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \widehat{m}(x_i))^2}{n - 2\nu + \tilde{\nu}}, \qquad \nu = \mathrm{trace}(L), \qquad \tilde{\nu} = \mathrm{trace}(L^\top L).$$

The quantity ν serves as an effective degrees of freedom.

7 Discussion

Visually, the kernel, local linear, and cubic spline regression curves look very similar. The kernel regression curve is slightly rougher than the rest.

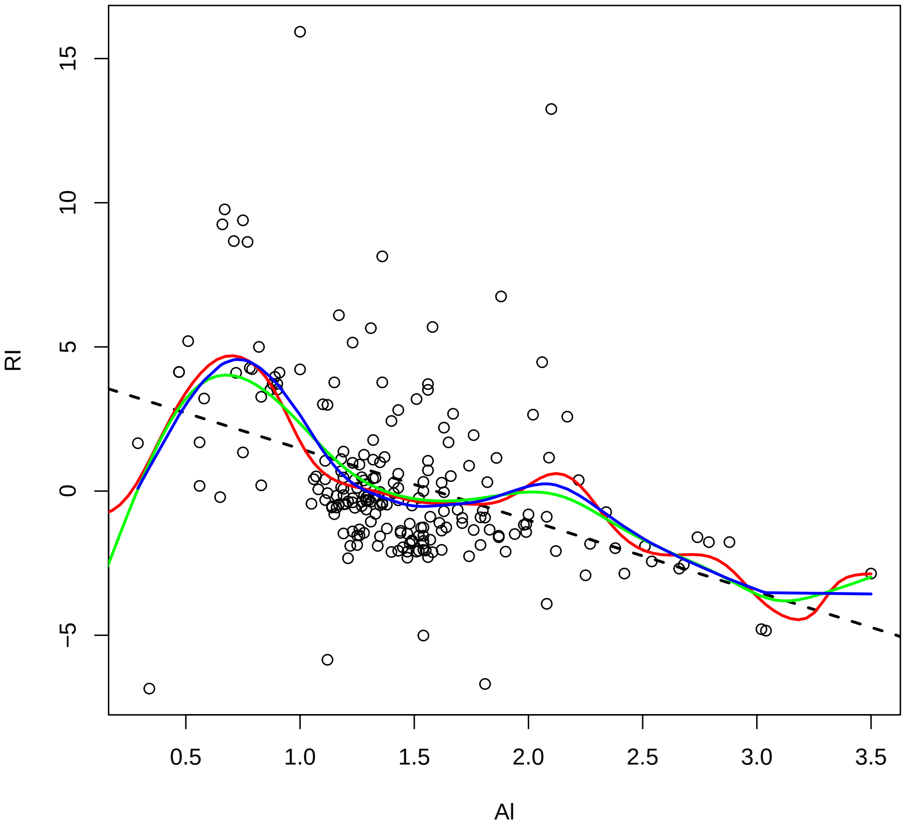

Figure 5: The linear, kernel, local linear, and spline regression curves.

On the other hand, the Ordinary Cross-Validation (OCV) score is minimum for the kernel regression (6.85), followed by the local linear regression (7.17) and finally the cubic spline regression (8.74). Thus, we choose the kernel regression method.