

Practical exercise 3 : estimation in single-index models

Objective: Write a piece of R code implementing an estimation method in single-index models, described in Chapter 3. You can illustrate its calculation using the Warsaw real estate data (this is obtained from `data(apartments)` after having loaded package `PBImisc`), or one of the real data examples in the package `np`, or a data set that you have simulated yourself.

Level 1: write a function, whose arguments should be a sample of data, a vector \mathbf{b} and a point z , which returns the list of all leave-one-out kernel estimators of the regression function $\mathbb{E}(Y - \mathbf{X}^\top \mathbf{b} | Z = z)$ at the point z , with a given kernel (*e.g.* naive) and a given bandwidth h . [There should be as many elements in this list as there are data points.]

Level 2: implement the nonlinear least squares estimator $\hat{\beta}$ of the parametric component of the model, with this same kernel and bandwidth.

Level 3: write a function, whose arguments should be a sample of data and a point z , which returns the kernel estimator of the regression function $\mathbb{E}(Y | \mathbf{X}^\top \beta = z)$ at the point z , with this same kernel and bandwidth.

Level 4: upgrade your code in Level 3, first by including as arguments in this function the kernel function (within a reasonable list of kernels) and the bandwidth h , and then by allowing a second, optimal choice of bandwidth made simultaneously with the nonlinear least squares procedure.

Level 5: speed up your code in Level 4.

Compare your results with those of existing packages. Do you see any problem with the implementation of the estimator? (instability, unreasonably large computation time...)