Co-funded by the Erasmus+ Programme of the European Union

Project funded by

European Commission Erasmus + Programme –Jean Monnet Action

Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Advanced Spatial Nonparametric Techniques - Nonparametric regression -

Margherita Gerolimetto and Stefano Magrini

Department of Economics - Ca'Foscari University of Venice

Pisa 23 October 2017

Nonparametric regression: model

Margherita Gerolimetto and Stefano Magrini

The underlying model for local regression is:

$$y_i = M(x_i) + \epsilon_i, \quad i = 1, \ldots, n$$

Note that:

- the distribution of the y_i's are unknown
- the functional form $M(x_i)$ is unknown

Assumptions:

- ϵ_i's are zero mean, uncorrelated and homoskedastic
- M can be **locally** approximated by a member of a parametric class, usually chosen to be a polynomial of certain degree p. This is defined **parametric localization**

Least squares polynomials

It is possible to estimate $M(x)$ in a neighborhood of x_0 by minimizing squared errors for pairs $(x_i, y_i), i = 1, \ldots, n$

$$min_{\alpha} \sum_{i=1}^{N} \{y_i - \alpha_0 - \alpha_1(x_i - x_0) - ... - \alpha_p(x_i - x_0)^p\}^2 \, w_i$$

Note that:

- this is a family of estimators parametrized by p that is the degree of polynomials
- if $p = 0$, then the estimator is the simple mean $\bar{y}$
- w_i are weights, whose mechanism is analogous to that for the kernel density estimate

Choosing the polynomial degree

In case $p = 0$:

$$\hat{M}(x) = \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h}) y_i}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})}$$

this is called **local constant** regression (aka kernel regression or Nadaraya-Watson regression)

In case $p = 1$:

$$\hat{M}(x) = \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h}) y_i}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})} + (x - \bar{X}_w) \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h})(x_i - \bar{X}_w) y_i}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})(x_i - \bar{X}_w)^2}$$

where

$$\bar{x}_w = \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h}) x_i}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})}$$

this is called **local linear** regression

Effect of different polynomial degrees

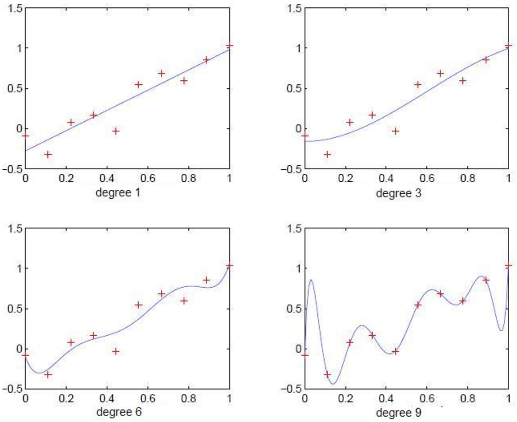

Figure: Different values of the polynomial degree p

Bandwidth

As for kernel density estimation the choice of the bandwidth parameter:

- is made with the goal of producing an estimate that is as smooth as possible without distorting the underlying pattern of dependence
- must find a balance between bias and variance

Common approaches for bandwidth selection are:

- plug-in
- cross validation (for fixed and variable bandwidth)
- k-neighbors

Optimal bandwidth, plug in and cross-validation

As for kernel density estimation the optimal smoothing is the value h that minimizes

$$MISE = E(\hat{M}_h - M)^2$$

where $\hat{M}_h$ is the estimated function with a certain h, i.e.

$$h_{opt} = \left[\frac{R(K)\sigma^2}{\left\{ n\mu_2(K)^2 \int M''(u)^2 f_X(u)du \right\}} \right]^{\frac{1}{5}}$$

where $R(K) = \int K(u)^2 du$, $\mu_2(K) = \int u^2 k(u)du$ and $f_X(u)$ is the density function for the predictors, σ^2 is the error term variance.

The plug-in bandwidth moves from h_{opt} and it is based on plugging in the optimal bandwidth formula a preliminary estimate for $M''(X)$

The cross-validation bandwidth is the value h that minimizes

$$CV(h) = \frac{1}{n} \sum \hat{e}_{i,i-1}^2$$

where $\hat{e}_{i,i-1}$ are leave-one-out residuals