

# STAT3612\_4TH\_TUTORIAL

YOU Jia

2/21/2017

## Non Parametric Regression: Introduction

The goal of a regression analysis is to produce a reasonable analysis to the unknown response function  $f$ , where for  $n$  number of data points  $(X_i, Y_i)$ , the relationship can be modeled as

$$y_i = f(x_i) + \epsilon_i \text{ for } i = 1, 2, \dots, n$$

### Approaches:

- Parametric approach:

$f(\cdot)$  is known and smooth. It is fully described by a finite set of parameters. Easy to interpret. e.g.: linear model:

$$y_i = x_i^T \beta + \epsilon_i$$

- Nonparametric approach:

$f(\cdot)$  is smooth, flexible, but unknown. We need to let the data decide the shape of  $f(\cdot)$ . Difficult to interpret.

$$y_i = f(x_i) + \epsilon_i$$

- Semi-parametric approach:

$f(\cdot)$  is a combination of parametric and non-parametric function. Have some parameters to estimate, but some parts are decided by the data.

$$y_i = x_i^T \beta + f_z(z_i) + \epsilon_i$$

## Series Estimation

Series estimation is a very common nonparametric regression method.

A linear series approximation takes the form:

$$f(x) \approx \sum_{j=1}^p \beta_j \phi_j(x) = \Phi(x)^T \beta$$

where  $\phi_j(x)$  are (nonlinear) functions of  $x$ . Known as basis functions.

A series approximation to  $f(x)$  takes the general form:

$$f_j(x) = f_j(x, \beta)$$

where  $f_j(x, \beta)$  is a known parametric family and  $\beta$  is a vector of  $j$  unknowns.

Aiming to minimize the following function:

$$\min \sum_i^n [y_i - \phi(x)^T \beta]^2 = (y - \Phi \beta)^T (y - \Phi \beta)$$

Then the least squares estimate  $\hat{\beta} = (\Phi^T \Phi)^{-1} \Phi^T y$ , which becomes a linear regression format.

Several candidates for series approximation:

- **Polynomial Basis:**

The polynomial Basis  $1, x, x^2, \dots$  might be the most common bases. The polynomial basis has the significant drawback of multicollinearity, which means that  $p$  must be relatively small. This can be avoided by orthogonalizing. A more significant drawback is that polynomials generally fit well near the center of the interval of interest but fit poorly near the endpoints. And polynomials are unable to reveal local features of  $f$  unless  $p$  is large.

- **Fourier Basis**

The Fourier basis  $1, \sin\omega x, \cos\omega x, \sin 2\omega x, \cos 2\omega x$  suitable for periodic functions. Note that  $\omega$  determines the period  $2\pi/\omega$ . Derivative estimation is quite simple in a Fourier setting since

$$\sin(k\omega x)' = k\omega \cos(k\omega x)$$

$$\cos(k\omega x)' = -k\omega \sin(k\omega x)$$

This implies that we can find the Fourier expansions of derivatives by multiplying coefficients by suitable powers of  $k\omega$  changing sign and interchanging *sine* and *cosine* coefficients as necessary.

The chief drawback of the Fourier basis is that it is suitable only for functions that have no prominent local features and similar curvature everywhere.

- **Wavelet Basis**

A wavelet basis is a basis for all functions on  $\mathbb{R}$  that are square integrable. One constructs a wavelet basis by first choosing a mother wavelet function  $\psi$ , which can then be translated and dilated as:

$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$$

where  $j$  and  $k$  are integers. For example, the Haar wavelets are

$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k), j \in \mathbb{N}, 0 \leq k < 2$$

where

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

The chief benefit of the wavelet approach is that it copes well with local behavior like discontinuities and rapid variation. The spatially adaptive nature of wavelets makes them quite useful for parsimoniously fitting functions that exhibit prominent local features. The downside is that a large sample size is required.

- **SPLINES**

Suppose that  $[a, b]$  is the interval of interest. A spline function  $\psi$  with domain  $[a, b]$  is a piecewise polynomial function, where the various pieces are defined on disjoint subintervals. More specifically, divide  $[a, b]$  into  $k$  disjoint subintervals  $[\tau_{j-1}, \tau_j]$  by choosing  $k - 1$  distinct interior points such that

$$a = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{k-1} < \tau_k = b$$

The  $t_j$  are usually called breakpoints. On subinterval  $[t_{j-1}, t_j]$ , the spline function is a polynomial:

$$\psi(x) = P_j(x) \text{ for } \tau_{j-1} \leq x < \tau_j \text{ where } (j = 1, \dots, k)$$

If  $m$  is the highest order of the  $P_j$ ,  $\psi$  is said to have order  $m$ , which is equivalent to saying that  $\psi$  has degree  $m - 1$ . The  $P_j$  are constructed to ensure smooth transitions at breakpoints.

We can make  $\psi$  more flexible by increasing the number of breakpoints. Sometimes the breakpoints are evenly spaced, but ideally one should place breakpoints intelligently. For example, it makes sense to have more breakpoints over a region where  $f$  has a complicated structure, fewer over a region where  $f$  is simpler

- **B-Spline**

Define new knots  $u_1, \dots, u_m$  such that

$$u_1 \leq u_2 \leq \dots \leq u_m \leq \tau_0$$

let  $u_{j+m} = \tau_j$  for  $j \in 1, \dots, k - 1$ , and define another set of  $m$  knots such that

$$\tau_k \leq u_{k+m+1} \leq \dots \leq u_{k+2m}$$

The placement of these extra knots does not matter, and so we usually let  $u_1 = \dots = u_m = \tau_0$  and  $u_{k+m+1} = \dots = u_{k+2m} = \tau_k$ . Now, define

$$\phi_{i1} = 1\{u_i \leq x < u_{i+1}\} (i = 1, \dots, k + 2m - 1)$$

Then, for  $j \leq m$ , let

$$\phi_{ij} = \frac{x - u_i}{u_{i+j} - u_i} \phi_{i, j-1} + \frac{u_{i+j} - x}{u_{i+j} - u_{i+1}} \phi_{i+1, j-1}$$

for  $i = 1, \dots, k + 2m - j$ . If the denominator is 0, so is the function.

Please refer to the code in Lecture Notes.