Fitting data by localized least squares

12191885 김재겸
12211917 김한별

Functional Data Analysis

4.7

1. Review
2. Introduction
3. Local Regression
4. Kernel smoothing
5. Localized basis function estimator
6. Summary of localized basis methods

Review

- fit the discrete observations y_j, j = 1, … , n using the model $y_j = x(t_j) + \epsilon_j$
- $$x(t) = \sum_k^K c_k \phi_k(t) = \mathbf{c}'\boldsymbol{\phi}.$$
- $$\hat{x}(t_j) = \sum_{\ell=1}^n S_j(t_\ell) y_\ell \ , \qquad \hat{x}(\mathbf{t}) = \mathbf{S}\mathbf{y} \ , \qquad \mathbf{S} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{W}$$
- In this section, $x(t_j) = \sum_\ell^n w_\ell y_\ell$

Introduction

- For a smoothing method to make any sense at all, the value of the function estimate at a point t must be influenced mostly by the observations near t.
- In this section, we consider estimators where the local dependence is made more explicit by means of local weight functions.
- Local weight functions? Local?

localized least squares

- What is the localized least squares ?
  - 회귀 분석에서 사용되는 방법 중 하나
  - 주어진 데이터 전체가 아닌, 어떠한 데이터 포인트 주변의 작은 지역(로컬 영역)을 고려하여 회귀 모델을 적합하는 기법
  - idea: 특정 데이터 포인트의 주변 데이터 포인트들에 가중치를 부여하여 회귀 모델을 구성

    -> 데이터 전체를 이용하는 것보다 해당 지역의 데이터 패턴을 더 정확하게 반영

Local regression non-linear regression 중 하나의 방법

- target point 주변에 있는 관측지들을 이용하여 그 point에서의 적합을 계산.

알고리즘 7.1 Local Regression At $X = x_0$

1. x_0에 가까운 x_i 훈련용 관측치의 일부 $s = k/n$를 모아라.

2. 이 이웃하는 각 점에 가중치 $K_{i0} = K(x_i, x_0)$을 할당해서, x_0로부터 가장 먼 점은 가중치가 0 이되고, 가장 가까운 점은 가장 높은 가중치를 갖게 한다. 이들 k개의 가장 가까운 이웃을 제외한 모든 점은 가중치 0을 갖는다.

3. 식 (7.14)를 최소화하는 $\hat{\beta}_0$와 $\hat{\beta}_1$를 찾아서 앞서 언급한 가중치를 이용해 x_i 상에서 y_i의 가중최소제곱회귀(weighted least squares regression)를 적합시켜라.

$$\sum_{i=1}^{n} K_{i0}(y_i - \beta_0 - \beta_1 x_i)^2. \tag{7.14}$$

4. x_0에서 적합된 값은 $\hat{f}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$에 의해 주어진다.

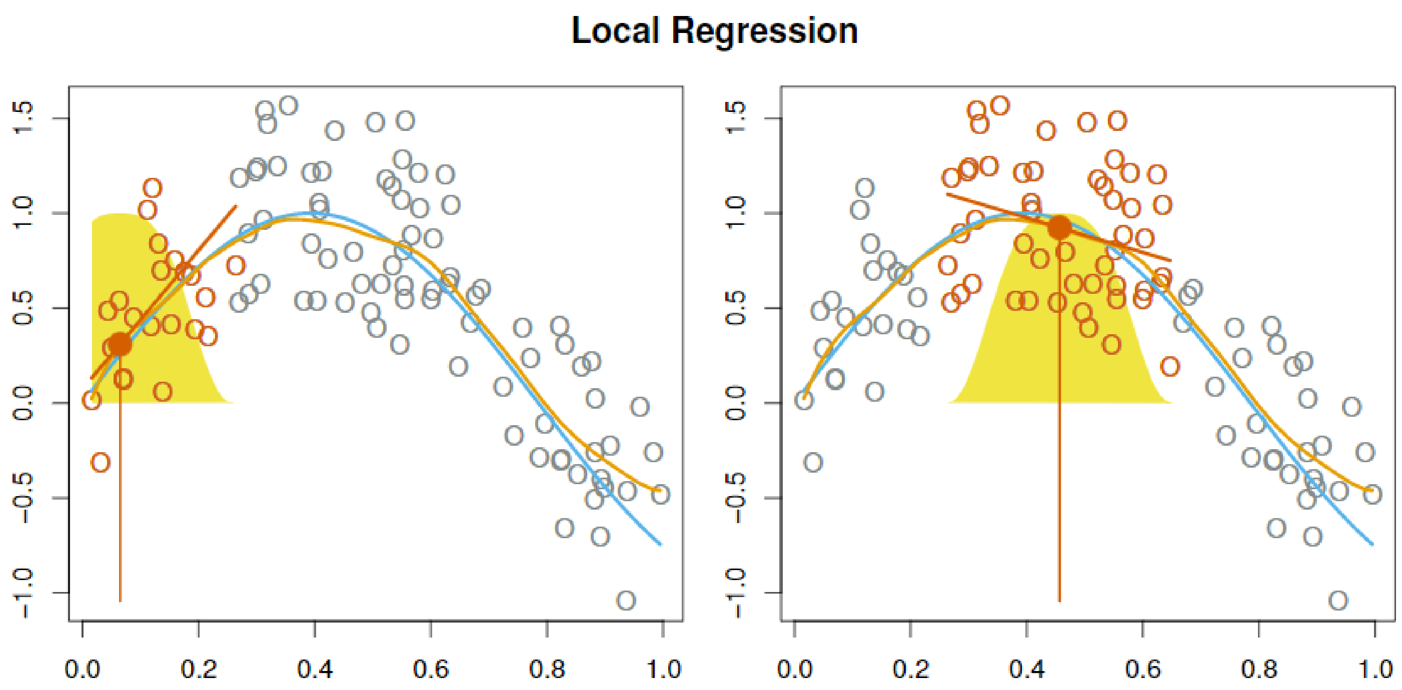


그림 7.9 국소적 회귀가 몇몇 모의 자료에서 설명된다. 여기서 파랑색 곡선은 모의 자료가 생성되게 하는 $f(x)$를 나타내고, 주황색 곡선은 국소적 회귀 추정치 $\hat{f}(x)$이다. 주황색 점들은 주황색 수직선으로 표현되는 목표점 x_0에 대해 근방의 점(local)들이다. 플랏에 겹쳐진 노랑색 종모양은 목표점으로부터 거리를 가지고 0까지 감소하는 각 점에 할당된 가중치를 나타낸다. x_0에서 적합 $\hat{f}(x_0)$는 가중 선형 회귀 (주황색 선분)를 적합하고, 추정치 $\hat{f}(x_0)$로서 x_0 (주황색으로 채워진 점)에서 적합된 값을 이용해서 얻어진다.

Kernel function

- a kernel function with values Kern(u).

Uniform:	$\mathtt{Kern}(u) = 0.5$ for $\lvert u \rvert \le 1,$	0 otherwise
Quadratic:	$\mathtt{Kern}(u) = 0.75(1-u^2)$ for $\lvert u \rvert \le 1,$	0 otherwise
Gaussian:	$\mathtt{Kern}(u) = (2\pi)^{-1/2}\exp(-u^2/2).$	

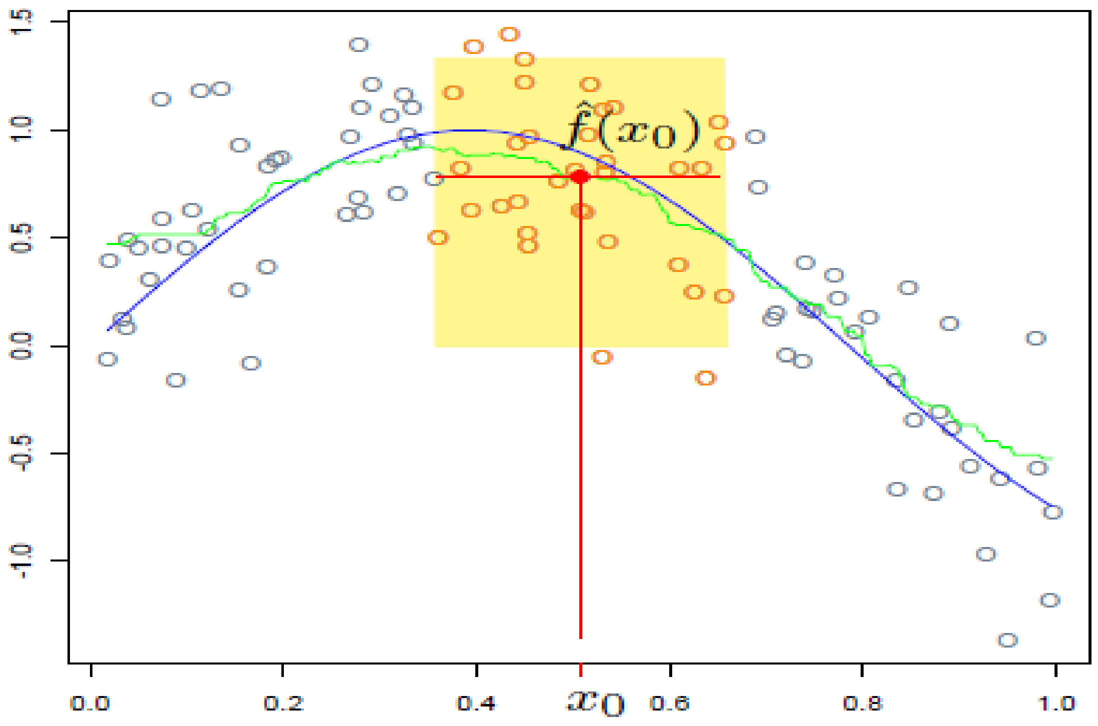
If we then define weight values to be

$$w_\ell(t) = \mathtt{Kern}\left(\frac{t_\ell - t_j}{h}\right) ,$$

Kernel smoothing

- k**-Nearest Neighbor Average**

$$\hat{f}(x_0) = Ave(y_i \mid x_i \in N_k(x_0))$$



Kernel smoothing

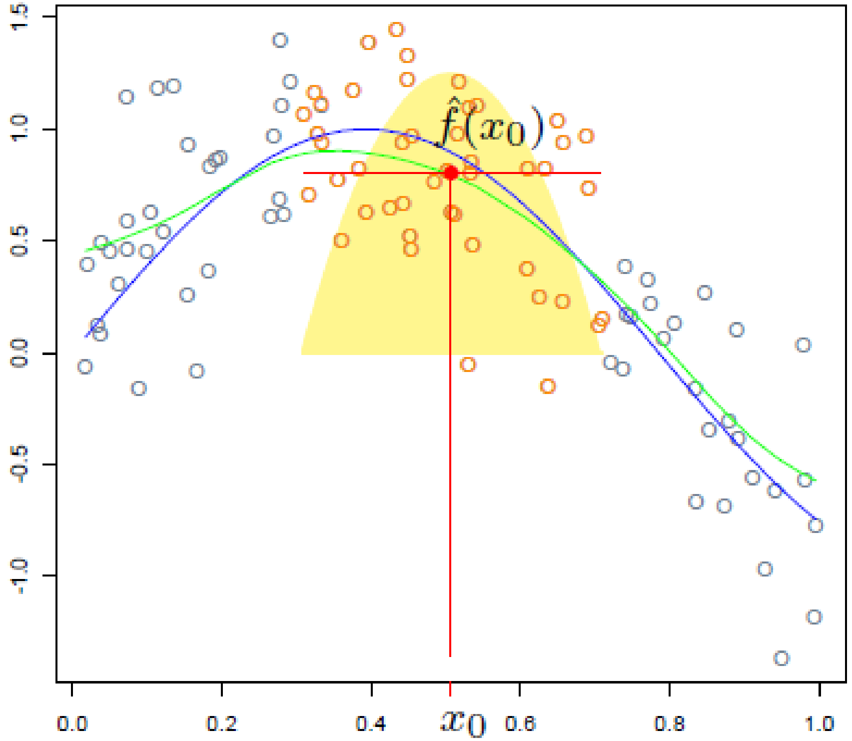
Nadaraya-Watson Kernel-weighted Average

$$\hat{f}(x_0) = \sum_{i=1}^{N} \frac{K_\lambda(x_0, x_i)}{\sum_{i=1}^{N} K_\lambda(x_0, x_i)} y_i$$

$$K_\lambda(x_0, x) = D\left(\frac{|x - x_0|}{\lambda}\right)$$

$$D(t) = \begin{cases} \frac{3}{4}(1 - t^2) & \text{if } |t| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\left|\frac{|x - x_0|}{\lambda}\right| \le 1$$



How to choose the bandwidth λ(h) ?

- The bandwidth controls the balance between two considerations

  : bias and variance in the estimate.

- Small values of λ imply that the expected value of the estimate $\hat{x}(t)$ must be close to the true value x(t).

- The price we pay is in terms of the high variability of the estimate, since it is be based on comparatively few observations.

- On the other hand, variability can always be decreased by increasing λ, although this is inevitably at the expense of higher bias.

- Unfortunately, none of these can always be trusted.

- Our own view is that trying out a variety of values of h and inspecting the consequences graphically remains a suitable means of resolving the bandwidth selection problem for most practical problems.

Localized basis function estimators

- combination of kernel estimators and basis function estimators
- basis idea : least squares criterion $$\text{SMSSE}(\mathbf{y}|\mathbf{c}) = \sum_{j=1}^{n}[y_j - \sum_{k}^{K} c_k\phi_k(t_j)]^2.$$
- $$\text{SMSSE}_t(\mathbf{y}|\mathbf{c}) = \sum_{j=1}^{n} w_j(t)[y_j - \sum_{k=1}^{K} c_k\phi_k(t_j)]^2$$
- $$\text{SMSSE}_t(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})'\mathbf{W}(t)(\mathbf{y} - \mathbf{\Phi c})$$

  -> W(t) is a diagonal matrix containing the weight values $w_j(t)$

- In Weighted least squares fits (4.2) : $\text{SMSSE}(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})'\mathbf{W}(\mathbf{y} - \mathbf{\Phi c})$

Localized basis function estimators

- $\hat{\mathbf{c}}(t) = [\mathbf{\Phi}'\mathbf{W}(t)\mathbf{\Phi}]^{-1}\mathbf{\Phi}'\mathbf{W}(t)\mathbf{y}$
- $\hat{x}(t) = \Phi\hat{c}(t) = \Phi(\Phi' W(t)\Phi)^{-1}\Phi' W(t)y = S(t)y$
- $S(t) = \mathbf{W}(t)\mathbf{\Phi}[\mathbf{\Phi}'\mathbf{W}(t)\mathbf{\Phi}]^{-1}\phi(t)$
- Consequently, $\hat{x}(t)$ is essentially a linear combination of only the observations in the neighborhood of t

Summary of localized basis methods

- The role of the bandwidth parameter h is obvious, and as a consequence it is even possible to allow h to adapt to curvature variation.
- However, these methods have the disadvantage of being unstable near the boundary of the section.

Q&A