Local Polynomial Regression

JEONG

August 6, 2008

Based on $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$, iid copies of (X, Y), we want to estimate

$$m(x) = E(Y|X = x).$$

Global Polynomial Regression

Model: $m(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$.

LSE: To minimize

$$\sum_{i=1}^{n} \Big(Y_i - \beta_0 - \beta_1 X_i - \cdots - \beta_p X_i^p\Big)^2$$

with respect to $(\beta_0, \beta_1, \cdots, \beta_p)$.

Local Polynomial Regression

Basic idea: For $x \approx x_0$,

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p.$$

To minimize

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(X_i - x_0) - \cdots - \beta_p(X_i - x_0)^p\}^2$$
$$\times I(|X_i - x_0| \leq h)$$

with respect to $(\beta_0, \beta_1, \cdots, \beta_p)$.

$$\widehat{m^{(r)}}(x_0) = r!\widehat{\beta}_r, \quad r = 0, 1, \cdots, p$$

where $(\widehat{\beta}_0, \cdots, \widehat{\beta}_p)$ is the solution of the above minimization problem.

Generalization

To minimize

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(X_i - x_0) - \cdots - \beta_p(X_i - x_0)^p\}^2$$
$$\times K((X_i - x_0)/h)$$

with respect to $(\beta_0, \beta_1, \cdots, \beta_p)$, where K is a weight function.

Matrix Notation

Polynomial regression: To solve

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{Y},$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 & \cdots & X_1^p \\ 1 & X_2 & \cdots & X_2^p \\ \vdots & \vdots & & \vdots \\ 1 & X_n & \cdots & X_n^p \end{pmatrix}, \; \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

and

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^\top.$$

Local polynomial regression: To solve

$$\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X} \boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{Y},$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 - x_0 & \cdots & (X_1 - x_0)^p \\ 1 & X_2 - x_0 & \cdots & (X_2 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & X_n - x_0 & \cdots & (X_n - x_0)^p \end{pmatrix}, \ \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^\top$$

and

$$\boldsymbol{W} = \mathrm{diag}\{K((X_1 - x_0)/h), \cdots, K((X_n - x_0)/h)\}.$$

Special Cases

$p = 0$ (local constant):

$$\widehat{m}_{\mathrm{NW}}(x_0) = \frac{\sum_{i=1}^{n} Y_i K((X_i - x_0)/h)}{\sum_{i=1}^{n} K((X_i - x_0)/h)}.$$

$p = 1$ (local linear):

$$\widehat{m}_{\mathsf{LL}}(x_0) = \frac{t_2(x_0)s_0(x_0) - t_1(x_0)s_1(x_0)}{s_2(x_0) - s_1(x_0)^2}$$

where

$$t_\ell(x_0) = \sum_{i=1}^{n} Y_i(X_i - x_0)^\ell K((X_i - x_0)/h)$$

and

$$s_\ell(x) = \sum_{i=1}^{n} (X_i - x_0)^\ell K((X_i - x_0)/h),$$

for $\ell = 0, 1, 2$.

Issues

Choice of K: less important

Choice of h: crucial

Statistical Properties

Notations:

- $\mu_\ell(K) = \int u^\ell K(u)du$, $\ell = 0, 1, 2, \cdots$.
- $R(K) = \int K(u)^2 du$.
- $f(x)$: the density of X supported on $[0, 1]$.
- $v(x) = Var(Y|X = x)$.

Assumptions: As $n \to \infty$, $h \to 0$ and $nh \to \infty$. And for $x_0 \in$ Interior(supp$f$),

- f is continuously differentiable at x_0 and $f(x_0) > 0$.

- v is continuous at x_0.

- m is twice continuously differentiable at x_0.

- K is a symmetric pdf supported on $[-1, 1]$.

Local constant estimator: For $x_0 \in [h, 1-h]$,

$$
\begin{aligned}
& E(\widehat{m}_{\mathrm{NW}}(x_0)|X_1, \cdots, X_n) - m(x_0) \\
&= \frac{h^2}{2}\left\{\frac{m''(x_0)f(x_0) + 2m'(x_0)f'(x_0)}{f(x_0)}\right\}\mu_2(K) \\
&\quad + o(h^2) + O_P(n^{-1/2}h^{1/2}),
\end{aligned}
$$

$$
\begin{aligned}
& Var(\widehat{m}_{\mathrm{NW}}(x_0)|X_1, \cdots, X_n) \\
&= \frac{1}{nh}\frac{v(x_0)}{f(x_0)}R(K) + o_P(n^{-1}h^{-1}).
\end{aligned}
$$

$$
\begin{aligned}
&\mathsf{AMSE}(\widehat{m}_{\mathsf{NW}}(x_0)|X_1, \cdots, X_n) \\
&= \frac{h^4}{4}\left\{\frac{m''(x_0)f(x_0) + 2m'(x_0)f'(x_0)}{f(x_0)}\right\}^2 \mu_2(K)^2 \\
&\qquad + \frac{1}{nh}\frac{v(x_0)}{f(x_0)}R(K).
\end{aligned}
$$

Local linear estimator:

$$E(\widehat{m}_{\mathsf{LL}}(x_0)|X_1, \cdots, X_n) - m(x_0)$$
$$= \frac{h^2}{2} m''(x_0)\mu_2(K) + o(h^2) + O_P(n^{-1/2}h^{1/2}),$$

$$Var(\widehat{m}_{\mathsf{LL}}(x_0)|X_1, \cdots, X_n)$$
$$= \frac{1}{nh}\frac{v(x_0)}{f(x_0)} R(K) + o_P(n^{-1}h^{-1}),$$

$$\mathsf{AMSE}(\widehat{m}_{\mathsf{LL}}(x_0)|X_1, \cdots, X_n)$$
$$= \frac{h^4}{4} m''(x_0)^2 \mu_2(K)^2 + \frac{1}{nh}\frac{v(x_0)}{f(x_0)} R(K).$$

Bandwidth Selection

$$h_{\text{opt}} \sim n^{-1/5}$$

Cross-validation:

- Prediction error: $E\{Y_{\text{new}} - \widehat{m}(x_{\text{new}})\}^2$

- $CV(h) := \sum_{i=1}^{n} \left\{Y_i - \widehat{m}_{(-i)}(X_i)\right\}^2$

- h_{CV} = the minimizer of $CV(h)$.

Plug-in method:

To plug-in the estimators of unknown functionals in the expression of the optimal bandwidth.