Stat Mod 2 <u>Local Poly. Regr. A</u> (p5) 2/18/15 Maurice Diesendruck

Show: <u>if</u> unnormalized weights are $w_i \equiv w(x, x_i) = \frac{1}{h} K\left(\frac{x_i - x}{h}\right)$

<u>then</u> the solution is exactly the kernel-regression estimator.

Start with $E[Y|x] = m(x)$ defining a functional relationship between input X, and output Y.

$$\left.\begin{aligned} m(x) &= \int y\, f(y|x)\, dy = \int y \cdot \frac{f(x,y)}{f(x)}\, dy \end{aligned}\right\} \text{Definition of } E[\cdot] \text{ and conditional distr.}$$

$$= \frac{\int y \cdot f(x,y)\, dy}{f(x)} = \frac{\int y \cdot f(x,y)\, dy}{\int f(x,y)\, dy}$$

Use $\hat{f}(x,y)$ in place of $f(x,y)$:

Next page

1

$$f(x) = a = \underset{\mathbb{R}}{\arg\min} \sum_{i=1}^{n} w_i (y_i - a)^2$$

If unnormalized weights are $w_i \equiv w(x, x_i) = \frac{1}{h} K\left(\frac{x_i - x}{h}\right)$

then normalized weights are $\frac{w_i}{\sum w_i} = \frac{\cancel{\frac{1}{h}} K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} \cancel{\frac{1}{h}} K\left(\frac{x_i - x}{h}\right)} = \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)}$

The kernel-regression estimator is therefore:

$$\boxed{\sum_{i=1}^{n} \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)} \cdot y_i}$$

$$\underset{a}{\arg\min} \sum \frac{1}{h} K\left(\frac{x_i - x}{h}\right) (y_i - a)^2$$

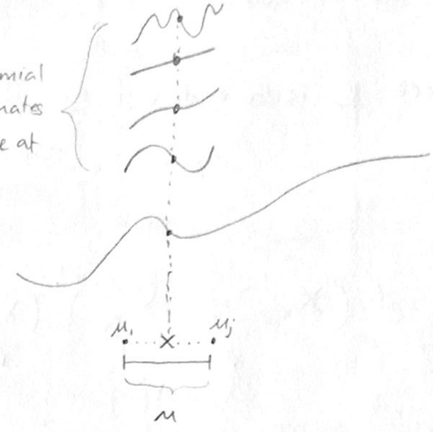
$$\frac{d}{da}\left(\sum \frac{1}{h} K\left(\frac{x_i - x}{h}\right) (y_i - a)^2 \right) = 0 \Rightarrow \sum_{i=1}^{n} \frac{d}{da}\left[\frac{1}{h} K\left(\frac{x_i - x}{h}\right) (y_i - a)^2 \right] = 0$$

$$\sum_{i=1}^{n} \left[\frac{1}{h} K\left(\frac{x_i - x}{h}\right) \cdot 2(y_i - a) \right] = 0 \Rightarrow \sum_{i=1}^{n} \left[\frac{K\left(\frac{x_i - x}{h}\right)}{h} \cdot 2 \cdot y_i - \frac{K\left(\frac{x_i - x}{h}\right)}{h} \cdot 2a \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{2}{h} K\left(\frac{x_i - x}{h}\right) y_i - \sum_{i=1}^{n} \frac{2}{h} K\left(\frac{x_i - x}{h}\right) \cdot a = 0 \Rightarrow \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) y_i = \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) a$$

$$\Rightarrow \boxed{a^* = \frac{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)} = \frac{\sum w_i y_i}{\sum w_i}}$$

Which polynomial best approximates the true value at x, $f(x)$?



Given $(x, y) = \left((x_1, y_1), \ldots, (x_n, y_n)\right)$; for each x_i:

Estimate local behavior using polynomial $g_x(\underline{\mu}, \underline{a}) = a_0 + \sum_{k=1}^{D} a_k (\mu - x)^k$

for all points $\underline{\mu} = \mu_1, \ldots, \mu_j$ in a neighborhood of x, and for weight coefficients $\underline{a} = a_0, \ldots, a_D$.

Want to estimate $\underline{a}$ in $g_x(\mu; a) = a_0 + a_1(\mu_1 - x) + a_2(\mu_2 - x)^2 + \ldots + (\mu_j - x)^j$

Such that $\sum_{i=1}^{n} w_i \left(y_i - g_x(\mu; a)\right)^2$ is minimized. (*)

Including the w_i definition:

$$\hat{f}(x_0) = \underset{a \in \mathbb{R}^{D+1}}{\arg\min} \sum_{i=1}^{n} \left(K\left(\frac{x_i - x_0}{h}\right) \left[y_i - a_0 - a_1(x_i - x_0) - a_2(x_i - x_0)^2 - \ldots - a_D(x_i - x_0)^D \right]^2 \right) \cdot \frac{1}{\sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h}\right)}$$

This is a weighted least squares form, with weights equal to kernel functions, $K\left(\frac{x_i - x}{h}\right)$.

Set up something that resembles $Y = XA$.

$$\underset{n \times (D+1)}{X_{x_0}} = \begin{bmatrix} 1 & (x_1 - x_0) & (x_1 - x_0)^2 & \cdots & (x_1 - x_0)^D \\ 1 & (x_2 - x_0) & (x_2 - x_0)^2 & & (x_2 - x_0)^D \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & (x_n - x_0) & (x_n - x_0)^2 & \cdots & (x_n - x_0)^D \end{bmatrix}, \quad \underset{(D+1) \times 1}{A} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_D \end{bmatrix}, \quad \underset{n \times 1}{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\underset{n \times n}{W_{x_0}} = \begin{bmatrix} K\left(\frac{x_1 - x_0}{h}\right) & 0 & \cdots & 0 \\ 0 & K\left(\frac{x_2 - x_0}{h}\right) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & K\left(\frac{x_n - x_0}{h}\right) \end{bmatrix}$$

By WLS, (*) is minimized with $\hat{A} = \left(X_{x_0}' W_{x_0} X_{x_0}\right)^{-1} \left(X_{x_0}' W_{x_0} Y\right)$

as long as $\left(X_{x_0}' W_{x_0} X_{x_0}\right)^{-1}$ is non-singular.

Putting the new estimator to work, $\hat{f}(x_0) = \underline{e}'(X_{x_0}' W_{x_0} X_{x_0})^{-1}(X_{x_0}' W_{x_0} Y)$

where $\underline{e}$ is a matrix of size $(D+1)\times 1$, and takes on values $\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\\\vdots\\0\end{bmatrix}, \dots, \begin{bmatrix}1\\\vdots\\1\end{bmatrix}$

$$e'\,(X_{x_0}' W_{x_0} X_{x_0})^{-1}(X_{x_0}' W_{x_0})\cdot Y$$

$$\Downarrow$$

$$\begin{bmatrix}1\times(D+1)\end{bmatrix}\begin{bmatrix}(D+1)\times(D+1)\end{bmatrix}\begin{bmatrix}(D+1)\times n\end{bmatrix}\cdot Y$$

$$\Downarrow$$

$$\begin{bmatrix}1\times(D+1)\end{bmatrix}\begin{bmatrix}(D+1)\times n\end{bmatrix}\cdot Y$$

for example: $\underbrace{\underbrace{[1,1,0,\dots,0]}_{1\times(D+1)}\underbrace{\left[\;\tilde{H}\;\right]}_{(D+1)\times n}}_{1\times n}\cdot Y$

where two 1's selects only the first two terms of the polynomial estimate, i.e. the local linear estimate

---

estimate: $g_x(u;a) = a_0 + \sum_{i=1}^{n} a_i(u-x)^k$, $\hat{A} = \underbrace{(X_x' W X_x)^{-1} X_x' W}_{Q'}\, Y$

$\hat{g}_x = [Q_1' Y]_i$ $\qquad = Q'Y$

where Q_1 is the first row of Q

StatMod 2 Local Poly Regr. C (p6) 2/24/15

$$R_x = \begin{bmatrix} 1 & (x_1-x) \\ \vdots & \vdots \\ 1 & (x_n-x) \end{bmatrix} \quad A = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad W = \begin{bmatrix} K\left(\frac{x_1-x}{h}\right) & & 0 \\ & \ddots & \\ 0 & & K\left(\frac{x_n-x}{h}\right) \end{bmatrix}$$

$$\hat{Y} = \hat{A} = (R_x' W R_x)^{-1} R_x' W Y \quad , \quad \text{let } \lambda_i = K\left(\frac{x_i-x}{h}\right), \ p_i = (x_i - x)$$

$$= \left[\underbrace{\begin{bmatrix} 1 & \cdots & 1 \\ p_1 & \cdots & p_n \end{bmatrix}}_{2\times n} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} 1 & p_1 \\ \vdots & \vdots \\ 1 & p_n \end{bmatrix}}_{n\times 2} \right]^{-1} \underbrace{\begin{bmatrix} 1 & \cdots & 1 \\ p_1 & \cdots & p_n \end{bmatrix}}_{2\times n} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1}$$

$$= \left[\begin{bmatrix} \lambda_1 & \cdots & \lambda_n \\ \lambda_1 p_1 & \cdots & \lambda_1 p_n \end{bmatrix} \begin{bmatrix} 1 & p_1 \\ \vdots & \vdots \\ 1 & p_n \end{bmatrix} \right]^{-1} \underbrace{\begin{bmatrix} \lambda_1 & \cdots & \lambda_n \\ \lambda_1 p_1 & \cdots & \lambda_1 p_n \end{bmatrix}}_{2\times n} \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1}$$

$$= \begin{bmatrix} \Sigma \lambda_i & \Sigma \lambda_i p_i \\ \Sigma \lambda_i p_i & \Sigma \lambda_i p_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \lambda_i y_i \\ \Sigma \lambda_i p_i y_i \end{bmatrix}$$

$$= \frac{1}{(\Sigma \lambda_i)(\Sigma \lambda_i p_i^2) - (\Sigma \lambda_i p_i)^2} \begin{bmatrix} \Sigma \lambda_i p_i^2 & -\Sigma \lambda_i p_i \\ -\Sigma \lambda_i p_i & \Sigma \lambda_i \end{bmatrix} \begin{bmatrix} \Sigma \lambda_i y_i \\ \Sigma \lambda_i p_i y_i \end{bmatrix}$$

$$= \frac{1}{(\Sigma \lambda_i)\left[\Sigma \lambda_i p_i^2 - (p_i)(\Sigma \lambda_i p_i)\right]} \begin{bmatrix} \downarrow \end{bmatrix} \begin{bmatrix} \downarrow \end{bmatrix}$$

$$= \frac{1}{\Sigma \lambda_i \left[S_2 - (x_i - x) S_1 \right]} \begin{bmatrix} S_2 & -S_1 \\ -S_1 & S_0 \end{bmatrix} \begin{bmatrix} \Sigma \lambda_i y_i \\ \Sigma \lambda_i p_i y_i \end{bmatrix} = \frac{1}{\Sigma w_i} \begin{bmatrix} S_2 \Sigma \lambda_i y_i - S_1 \Sigma \lambda_i p_i y_i \\ -S_1 \Sigma \lambda_i y_i + S_0 \Sigma \lambda_i p_i y_i \end{bmatrix}$$

$$= \frac{1}{\Sigma w_i} \begin{bmatrix} \Sigma \lambda_i (S_2 - S_1 \cdot p_i)\, y_i \\ -\Sigma \lambda_i (S_1 - S_0 \cdot p_i)\, y_i \end{bmatrix} = \frac{1}{\Sigma w_i} \begin{bmatrix} \Sigma w_i y_i \\ \end{bmatrix}$$

Thus, at point x, only the constant $a_0 = [\hat{A}]_1$ is used: $\hat{f}(x) = a_0 = \dfrac{\Sigma w_i y_i}{\Sigma w_i}$

where $\Sigma w_i = \sum_{i=1}^{n} \left[\mathbf{K}\left(\frac{x_i - x}{h}\right) \left[\left(\sum_{j=1}^{n} \mathbf{K}\left(\frac{x_j-x}{h}\right)(x_j-x)^2 \right) - (x_i - x)\left(\sum_{k=1}^{n} \mathbf{K}\left(\frac{x_k-x}{h}\right)(x_k-x) \right) \right] \right]$

Supposing residuals have constant variance σ^2, derive mean and variance of sampling distribution for local polynomial estimate $\hat{f}(x)$ at some arbitrary x_0.

Recall: $Y = \overset{\nearrow f(x)}{XA} + e$ where $E[e_i] = 0$, $Var[e_i] = \sigma^2$ $\Rightarrow$ $E[Y] = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$, $var(Y) = \sigma^2$

$\hat{A} = (X'WX)^{-1}X'WY$, $e'\hat{A}$ is the first term of $\hat{A}$.

$$E[\hat{A}] = E\left[(X'WX)^{-1}X'WY\right] = E\left[(X'WX)^{-1}X'W(XA+e)\right]$$

$$= E\left[(X'WX)^{-1}X'WX \cdot A\right] + E\left[(X'WX)^{-1}X'We\right] = (X'WX)^{-1}XWX \cdot E[A] + (X'WX)^{-1}XW \cdot E[e]$$

$$= A + 0 = A$$

$$= \boxed{(X'WX)^{-1}X'W \cdot E[Y]}$$

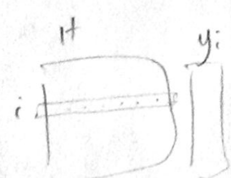
$$Var(\hat{A}) = \boxed{(X'WX)^{-1}X'W \cdot var(Y) \cdot W'X(X'WX)^{-1}}$$

$$E[\hat{f}(x_0)] = E[e'A] = e' \cdot E[A] = e' \cdot (X_x'WX)^{-1}X_x'W \cdot E[Y] = \boxed{\Sigma\, w_i \cdot f(x_i)}$$

where $E[Y] = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$

$$Var\left(\hat{f}(x_0)\right) = e'(X_x'WX_x)^{-1}X_x'W \cdot var(Y) \cdot W'X_x(X_x'WX_x)^{-1'}e$$

$$= \boxed{\Sigma\, w_i^2 \cdot \sigma^2}$$



$\underline{\text{If}}$ $x \sim \begin{matrix} E[x] = \mu \\ Var(x) = \Sigma \end{matrix}$, $\underline{\text{then}}$ for $Q = Q'$, $E[x'Qx] = tr(Q\Sigma) + \mu' Q \mu$.

Write vector of residuals as $r = y - \hat{y} = y - Hy$, where H is smoothing matrix.

Compute $E[\hat{\sigma}^2]$ where $\hat{\sigma}^2 = \dfrac{\|r\|_2^2}{n - (2tr(H) - tr(H'H))}$

$\boxed{\text{Note:}}$ $\|r\|_2 = (r_1^2 + r_2^2 + \dots + r_n^2)^{\frac{1}{2}}$

$\|r\|_2^2 = (r_1^2 + r_2^2 + \dots + r_n^2) = SSE$

$\|r\|_2^2 = (y - Hy)'(y - Hy)$

$= \sum_{i=1}^{n} (y_i - [Hy]_i)^2$

$$\hat{\sigma}^2 = \frac{(y - Hy)'(y - Hy)}{n - (2tr(H) - tr(H'H))}$$

$$= \frac{[(I-H)y]'[(I-H)y]}{n - (2tr(H) - tr(H'H))}$$

$$= \frac{y'(I-H)'(I-H)y}{n - (2tr(H) - tr(H'H))}, \quad \text{let } Q = (I-H)'(I-H) = (I-H)$$

$$= \frac{y'Qy}{n - (2tr(H) - tr(H'H))} \Rightarrow E[\hat{\sigma}^2] = E\left[\frac{y'Qy}{n - (2tr(H) - tr(H'H))}\right]$$

where denominator is a constant, so comes out of $E[\cdot]$.

$$\Rightarrow E[\hat{\sigma}^2] = \frac{1}{n - (2tr(H) - tr(H'H))} \cdot \left[tr(Q\Sigma) + (Hy)'QHy\right]$$

because $E[y] = Hy$, $var(y) = \Sigma$

$$= \frac{tr((I-H)'(I-H)\Sigma) + (Hy)'(I-H)'(I-H)Hy}{n - [2tr(H) - tr(H'H)]}$$

$\to 0$, $\because Hy \in C(x)$, $(I-H)Hy = 0$

$$= \frac{tr((I-H)\Sigma) + (Hy)'(I-H)Hy}{n - [2tr(H) - tr(H'H)]} = \frac{\Sigma[tr(I-H)] + \cancel{y'H'(I-H)Hy}}{n - [2tr(H) - \underbrace{tr(H'H)}_{tr(H),\ \because H'H = H}]} = \frac{\Sigma[r(I-H)]}{n - [2 \cdot r(H) - r(H)]}$$

$\hookrightarrow = r(H)$

$$= \frac{\Sigma[r(I-H)]}{n - r[H]} = \frac{\Sigma[n - r[H]]}{n - r[H]} = \Sigma$$

?

This an unbiased estimator $\forall n$.