REGRESI POLINOMIAL LOKAL

NURMA NUGRAHA

0305010432

UNIVERSITAS INDONESIA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

DEPARTEMEN MATEMATIKA

DEPOK

2009

REGRESI POLINOMIAL LOKAL

Skripsi diajukan sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Oleh:

NURMA NUGRAHA

0305010432

DEPOK

2009

SKRIPSI : REGRESI POLINOMIAL LOKAL

NAMA : NURMA NUGRAHA

NPM : 0305010432

SKRIPSI INI TELAH DIPERIKSA

DEPOK, 2 DESEMBER 2009

DRA. YEKTI WIDYANINGSIH, M.Si	SARINI ABDULLAH, S.Si., M.Stats
PEMBIMBING I	PEMBIMBING II

Tanggal lulus Ujian Sidang Sarjana : 15 Desember 2009

Penguji I : Dra. Yekti Widyaningsih, M.Si

Penguji II : Dra. Siti Nurrohmah, M.Si

Penguji III : Dra. Nora Hariadi, M.Si

KATA PENGANTAR

Alhamdulillahi rabbil ‘aalamiin. Segala puji dan syukur hanya kepada ALLAH SWT, Yang Maha Pengasih, yang membuat penulis dapat menyelesaikan tugas akhir ini. Shalawat dan salam penulis sampaikan kepada suri tauladan kita, manusia biasa dengan akhlak luar biasa, Rasulullah SAW.

Terselesaikannya skripsi ini tidak terlepas dari bantuan, bimbingan, dorongan, dan doa yang tulus dari banyak pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Orang tua penulis. Mama dan Ayah yang selalu berdoa untuk penulis disetiap sujudnya. Terima kasih atas doa, kasih sayang, semangat, pengorbanan yang tak pernah putus dan membuat penulis mampu menyelesaikan tugas akhir ini.
2. Ibu Yekti Widyaningsih selaku Pembimbing I penulis, yang telah banyak meluangkan waktunya disela kesibukkan ibu menempuh studi S3, untuk memberikan bimbingan, saran, pengarahan dan kemudahan lainnya dengan sangat sabar sehingga skripsi ini dapat diselesaikan dengan baik.
3. Ibu Sarini Abdullah selaku Pembimbing II penulis, yang juga telah banyak meluangkan waktu untuk memberikan bimbingan, saran, pengarahan dan kemudahan lainnya dengan sabar dan ikhlas. Terima kasih bu.

4. Ibu Siti Nurrohmah selaku pembimbing akademis yang telah memberikan nasihat dan bimbingannya.
5. Seluruh dosen Departemen Matematika atas segala ilmu yang penulis peroleh selama menjadi mahasiswa Matematika UI.
6. Seluruh karyawan Departemen Matematika, baik TU maupun Perpustakaan Matematika yang telah banyak memberikan bantuannya demi kelancaran penyusunan skripsi penulis.
7. Kakak dan adik penulis, Ce Diah, Mas Eko, Ika, I'am, yang telah banyak memberikan dukungan, bantuan, dan doanya.
8. Kakek, nenek, dan seluruh keluarga besar penulis yang banyak memberikan dukungan dan doa.
9. Untuk sahabat-sahabat terbaik penulis, Atikah Ashriyani, Kiki, Karlina, Desti, yang telah memberikan semangat dan do'a.
10. Teman-teman yang sama-sama berjuang untuk menyelesaikan skripsi semester ini, Mya, Icha, Rani, Akmal, Mba Avi, Mba Intan, Miranti, Yuni, Anggi, Aris, Udin, Nasib, Inul, Rahanti. Alhamdulillah, sesungguhnya bersama kesulitan ada kemudahan.
11. Teman-teman satu bimbingan, Hairu, Putri, Nafia, dan Dian, terus berjuang ! Allah akan memberikan yang terbaik untuk kita .
12. Teman-teman Musholla Izzatul Islam FMIPA UI 2007, 2008, 2009, rekan-rekan tentor di Primagama Kalisari, teman-teman Salam UI X2 , teman-teman Musholla An-Nur, teman-teman IKRM, Syukron Jazakumullah atas

do’a, motivasi dan pengertiannya, perjalanan hidup ini tak akan indah tanpa ukhuwah dari kalian.

13. Teman-teman angkatan 2004, 2005, 2006, 2007, 2008, 2009.
14. Semua pihak yang telah membantu penulis dengan dukungan dan doanya.

Mohon maaf jika pada skripsi ini terdapat kesalahan dan kekurangan. Semoga skripsi ini dapat berguna bagi siapa saja yang mengkajinya, serta dapat dikembangkan dan disempurnakan agar lebih bermanfaat untuk kepentingan orang banyak.

Bukankah Kami telah melapangkan untukmu dadamu ?
dan Kami telah menghilangkan daripadamu bebanmu,
yang memberatkan punggungmu ?
Dan Kami tinggikan bagimu sebutan (nama) mu,
Karena sesungguhnya sesudah kesulitan itu ada kemudahan,
sesungguhnya sesudah kesulitan itu ada kemudahan.
Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain.
dan hanya kepada Tuhanmulah hendaknya kamu berharap.

Penulis

2009

ABSTRAK

Analisis regresi merupakan suatu metode statistik untuk menyelidiki dan memodelkan hubungan antara satu variabel respon Y dengan satu atau lebih variabel prediktor X. Hubungan antara variabel prediktor X dan variabel respon Y secara umum dapat dimodelkan dengan sebuah fungsi regresi. Menentukan fungsi taksiran regresi dapat dilakukan secara parametrik dan nonparametrik. Dalam tugas akhir ini fungsi regresi ditaksir secara nonparametrik dengan metode regresi polinomial lokal. Regresi polinomial lokal adalah suatu metode regresi nonparametrik, dengan fungsi regresi ditaksir menggunakan bentuk polinomial. Jika pada regresi polinomial biasa persamaan regresi di-*fit* untuk seluruh wilayah data maka dalam regresi polinomial lokal persamaan regresi di-*fit* sepotong-sepotong. Kemulusan kurva dari taksiran regresi ini tergantung pada pemilihan parameter pemulus atau *bandwidth*, sehingga diperlukan pemilihan *bandwidth* yang optimal, yaitu *bandwidth* yang meminimumkan GCV. Dalam aplikasi metode regresi polinomial lokal dibandingkan dengan metode Nadaraya-Watson. Hasil yang diperoleh adalah metode regresi polinomial lokal akan baik menaksir data yang nilainya menyimpang jauh dibandingkan nilai data yang lain, sedangkan metode Nadaraya-Watson akan baik menaksir pada data yang berkumpul.

Kata kunci: *bandwidth*; regresi nonparametrik; regresi polinomial lokal; CV; GCV; Nadaraya-Watson; fungsi *kernel*.

xiii + 76 hlm.; lamp

Bibliografi: 6 (1990-2006)

DAFTAR ISI

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

DAFTAR NOTASI

X	Variabel prediktor
Y	Variabel respon
i	Pengamatan ke-
X_i	Variabel prediktor pengamatan $ke-i$
Y_i	Variabel respon pengamatan $ke-i$
x	Variabel prediktor yang nilainya tidak terdapat pada data awal namun akan digunakan untuk menaksir Y
$m(x)$	Suatu fungsi regresi
$\hat{m}(x)$	Taksiran fungsi regresi
ε_i	*Random error*, suatu variabel acak yang menggambarkan variasi Y di sekitar $m(x)$
n	Banyaknya pengamatan
$x*$	Nilai tetap yang ditentukan
$m(x, x*)$	Persamaan regresi pada polinomial lokal untuk nilai prediktor x yang dekat dengan $x*$,
p	Derajat polinomial
h	*Bandwidth*, yaitu konstanta positif yang menentukan lebarnya lingkungan di sekitar $x*$

$w(\cdot)$	Fungsi *kernel* yang dijadikan sebagai fungsi pembobot
$\hat{m}(x^*, x^*)$	Taksiran $m(x, x^*)$ pada saat $x = x^*$
$[\mathbf{X}]_{ij}$	Entri baris ke i kolom ke j pada matriks $\mathbf{X}$
$[\mathbf{a}]_j$	Baris ke j pada matriks $\mathbf{a}$
$[\mathbf{H}]_{kj}$	Entri baris ke k kolom ke j pada matriks $\mathbf{H}$

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Analisis regresi merupakan suatu metode statistik untuk menyelidiki dan memodelkan hubungan antara satu variabel respon Y dengan satu atau lebih variabel prediktor X. Misalnya, diberikan himpunan data $\{(X_i, Y_i)\}$, i = 1, …, n. Secara umum hubungan antara Y dan X dapat ditulis sebagai :

$$Y_i = m(X_i) + \varepsilon_i \qquad (1.1)$$

dengan $m(x)$ adalah suatu fungsi regresi, dan ε_i adalah suatu variabel acak yang menggambarkan variasi Y di sekitar $m(x)$. Permasalahan dalam analisis regresi adalah menentukan fungsi dugaan $(\hat{m}(x))$ yang mewakili keterkaitan antara X dan Y pada data yang diberikan.

Penentuan fungsi dugaan regresi dapat dilakukan secara parametrik dan nonparametrik. Penaksiran fungsi dugaan yang paling umum dan sering kali digunakan adalah penaksiran fungsi dugaan secara parametrik. Pada penaksiran fungsi secara parametrik biasanya fungsi regresi diasumsikan merupakan suatu fungsi yang diketahui bentuknya. Fungsi tersebut digambarkan oleh sejumlah hingga parameter yang harus ditaksir. Dalam regresi parametrik terdapat beberapa asumsi mengenai model, sehingga diperlukan pengecekan akan terpenuhinya asumsi tersebut. Namun, apabila tidak ada referensi bentuk kurva tertentu maka digunakan penaksiran fungsi regresi secara nonparametrik yang lebih fleksibel. Ada beberapa metode pendugaan dalam regresi nonparametrik, antara lain penduga *kernel*, penduga *spline*, penduga deret *orthogonal*, dan analisis *wavelet*.

Tugas akhir ini membahas penaksiran fungsi regresi secara nonparametrik menggunakan metode regresi polinomial lokal.

1.2 PERUMUSAN MASALAH

Permasalahan yang akan dibahas dalam tulisan ini adalah :
Bagaimana mencari taksiran fungsi $m(x)$ pada regresi nonparametrik dengan menggunakan metode regresi polinomial lokal ?

1.3 TUJUAN PENULISAN

Menjelaskan metode taksiran regresi polinomial lokal, mengaplikasikannya pada data, serta membandingkan metode regresi polinomial lokal dengan metode Nadaraya-Watson.

1.4 PEMBATASAN MASALAH

Regresi polinomial lokal yang akan dibahas dibatasi untuk polinomial berderajat 2, menggunakan fungsi *kernel Gaussian*, dan satu variabel prediktor (X).

1.5 SISTEMATIKA PENULISAN

Penulisan pada tugas akhir ini dibagi menjadi 5 bab, yaitu :

Bab I : Pendahuluan

Membahas latar belakang penulisan, perumusan masalah, tujuan penulisan, pembatasan masalah, dan sistematika penulisan.

Bab II : Landasan Teori

Membahas teori dasar yang digunakan pada pembahasan regresi polinomial lokal.

Bab III : Regresi Polinomial Lokal

Membahas dan menjelaskan Regresi Polinomial Lokal.

Bab IV : Penerapan Regresi Polinomial Lokal pada Data

Membahas penerapan dari Regresi Polinomial Lokal pada data.

Bab V : Kesimpulan dan Saran

BAB II

LANDASAN TEORI

Untuk memahami metode regresi polinomial lokal, perlu dibahas terlebih dahulu teori-teori yang mendasarinya, yaitu metode regresi, hampiran Taylor, *bandwidth*, matriks *Hat*, fungsi *kernel*, dan regresi nonparametrik dengan metode estimasi Nadaraya-Watson.

2.1 METODE REGRESI

Metode regresi adalah suatu metode statistik untuk menyelidiki dan memodelkan hubungan antara variabel respon Y dan variabel prediktor X. Misalnya diberikan himpunan data $\{(X_i, Y_i)\}$, i = 1, …, n. Secara umum hubungan antara Y dan X dapat ditulis sebagai berikut :

$$Y_i = m(X_i) + \varepsilon_i \qquad (2.1)$$

dengan $m(x)$ adalah suatu fungsi regresi yang belum diketahui dan ingin ditaksir, dan ε_i adalah suatu variabel acak yang menggambarkan variasi Y di sekitar $m(x)$ (Hardle, 1990).

Penaksiran fungsi regresi dapat dilakukan dengan dua cara, yaitu secara parametrik dan nonparametrik. Pada regresi parametrik digunakan bentuk fungsi parametrik tertentu sebagai $m(x)$. $m(x)$ digambarkan oleh sejumlah hingga parameter yang harus ditaksir. Dalam regresi parametrik terdapat beberapa asumsi mengenai model, sehingga diperlukan pengecekan akan terpenuhinya asumsi tersebut. Contoh bentuk model regresi parametrik dengan satu variabel prediktor :

- Model regresi linier sederhana :

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{2.2}$$

  model dengan satu variabel prediktor yang hubungannya dengan variabel respon Y digambarkan oleh sebuah garis lurus.

- Model regresi polinomial order 2 (model kuadratik) :

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \tag{2.3}$$

  kurva regresi digambarkan oleh kurva lengkung kuadratik.

- Model polinomial order ke-k :

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_k X^k + \varepsilon . \tag{2.4}$$

Pada model-model regresi tersebut parameter regresi biasanya ditaksir dengan menggunakan metode *least square*. Metode *least square* merupakan salah satu metode yang paling banyak digunakan untuk menduga parameter-parameter regresi. Biasanya penduga *least square* ini diperoleh dengan meminimumkan jumlah kuadrat residual.

$$SSR = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \qquad (2.5)$$

Pada regresi nonparametrik, fungsi regresi ($m(x)$) ditaksir tanpa referensi bentuk kurva tertentu. Cara ini lebih fleksibel karena tidak memerlukan informasi apa-apa tentang fungsi regresinya, dan $m(x)$ akan mengikuti bentuk data.

2.2 HAMPIRAN TAYLOR

Teorema Taylor

Misalnya f adalah fungsi yang memiliki turunan-turunan keberapapun pada suatu selang $(a-r, a+r)$, dengan a adalah titik yang berada dekat dengan x , dan r adalah besar lingkungan di sekitar a. Deret Taylor

$$f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots \qquad (2.6)$$

merepresentasikan fungsi f pada selang $(a-r, a+r)$ jika dan hanya jika

$$\lim_{n\to\infty} R_n(x) = 0$$

dimana $R_n(x)$ adalah suku sisa dalam rumus Taylor,

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \qquad (2.7)$$

dan c adalah titik pada $(a-r, a+r)$.

Deret Taylor ini digunakan untuk mencari nilai hampiran dari sebuah fungsi. Misalnya pada deret Taylor orde ke-1, hampiran yang dihasilkan disebut hampiran linier terhadap f di sekitar a :

$$P_1(x) = f(a) + f'(a)(x-a) \tag{2.8}$$

Hampiran linier $P_1(x)$ akan berhasil dengan baik ketika x berada dekat a, tetapi kurang begitu bagus jika x tidak berada dekat a dan penjumlahan menuju suku-suku dengan orde yang lebih tinggi dalam deret Taylor biasanya akan memberikan hampiran yang lebih baik. Jadi, polinomial orde dua :

$$P_2(x) = f(a) + f'(a)(x-a) + \frac{f''(x)}{2!}(x-a)^2 \tag{2.9}$$

yang tersusun atas tiga suku pertama dari deret Taylor untuk f, menghasilkan hampiran yang lebih baik untuk f dibandingkan dengan hampiran linier $P_1(x)$. Polinomial Taylor orde n adalah :

$$P_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n . \tag{2.10}$$

2.3 *BANDWIDTH*

Bandwith yang dinotasikan dengan h adalah suatu konstanta positif untuk menentukan kemulusan dari kurva taksiran regresi. Dengan memilih *bandwidth* yang mendekati 0 maka taksiran yang didapatkan tidak mulus bahkan hanya menonjolkan plot datanya saja. Jika nilai *bandwidth* semakin besar maka taksiran yang akan didapat semakin mulus tetapi dengan konsekuensi biasnya semakin besar. Oleh karena itu, diperlukan pemilihan *bandwidth* yang optimal, yaitu *bandwidth* yang menghasilkan kurva regresi yang mulus dan meminimumkan nilai biasnya. CV dan GCV adalah ukuran yang akan digunakan untuk menentukan *bandwidth* yang optimal.

Cross-Validation (CV) dinyatakan dengan formula sebagai berikut :

$$CV[\hat{m}(x)] = \sum_{i=1}^{n} \frac{(Y_i - \hat{m}(X_i))^2}{n.(1-[\mathbf{H}]_{ii})^2} \tag{2.11}$$

Meskipun *cross-validation* merupakan metode yang efektif untuk menentukan bandwidth yang optimal, namun *cross-validation* juga mempunyai kelemahan. Karenanya, dikembangkan metode *Generalized Cross-Validation* (GCV). GCV dinyatakan dengan formula berikut :

$$GCV = \frac{\sum_{i=1}^{n} (Y_i - \hat{m}(X_i))^2}{n.\left(1 - \frac{\sum_{i=1}^{n} [\mathbf{H}]_{ii}}{n}\right)^2} \tag{2.12}$$

dengan $\sum_{i=1}^{n} [\mathbf{H}]_{ii}$ adalah *trace* dari matriks **H** yang sangat berperan penting dalam persamaan ini.

2.4 MATRIKS *HAT*

Matriks *hat* berperan penting dalam regresi secara umum, untuk pendugaan yang menggunakan penduga linier. Hubungan matriks *hat* dalam regresi secara umum adalah :

$$\hat{\mathbf{Y}} = \mathbf{HY} \tag{2.13}$$

dengan $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ matriks yang berukuran n x n *(lampiran 1)*.

Dalam regresi polinomial lokal, entri dari matriks *hat* adalah

$[\mathbf{H}]_{kj} = \mathbf{e}_1{}^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{e}_j{}' \cdot w\left(\dfrac{X_j - X_k}{h}\right)$ *(lampiran 2)*, dengan elemen diagonal

$[\mathbf{H}]_{kk} = [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t]_{11}.w(0)$ *(lampiran 3).*

2.5 FUNGSI *KERNEL*

Fungsi kernel dinotasikan dengan $K(u)$, merupakan suatu fungsi yang pada pemanfaatannya diberlakukan pada setiap titik pada data. Fungsi ini mempunyai tiga sifat, yaitu :

- $K(u) \geq 0$, untuk semua u,
- $\int_{-\infty}^{\infty} K(u) = 1$,

- $K(-u) = K(u)$, untuk semua u ($K(u)$ bersifat simetris).

Tabel 1 menunjukkan beberapa fungsi kernel :

Tabel 1. Beberapa Fungsi Kernel

Kernel	**K(u)**
Uniform	$\frac{1}{2} I(\lvert u\rvert \leq 1)$
Triangle	$(1-\lvert u\rvert) I(\lvert u\rvert \leq 1)$
Epanechnikov	$\frac{3}{4}(1-u^2) I(\lvert u\rvert \leq 1)$
Quartic	$\frac{15}{16}(1-u^2)^2 I(\lvert u\rvert \leq 1)$
Triweight	$\frac{35}{32}(1-u^2)^3 I(\lvert u\rvert \leq 1)$
Gaussian	$\exp(-\frac{1}{2}u^2)$
Cosinus	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right) I(\lvert u\rvert \leq 1)$

2.6 METODE NADARAYA-WATSON

Salah satu teknik regresi nonparametrik untuk menaksir fungsi regresi $(m(x))$ adalah dengan menggunakan penduga *Nadaraya-Watson.* Penduga ini diperoleh dengan menggunakan metode penaksiran fungsi densitas kernel. Bentuk persamaan regresi dengan metode Nadaraya-Watson adalah :

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{ki=1}^{n} K\left(\frac{x - X_k}{h}\right)} \tag{2.14}$$

$\hat{m}(x)$ = fungsi taksiran regresi

x = variabel prediktor yang nilainya tidak teramati namun akan digunakan untuk menaksir Y

X_i = variabel prediktor pada data ke-i

Y_i = variabel respon pada data ke-i

$K(\cdot)$ = fungsi kernel

n = ukuran sampel/banyak pengamatan

h = lebar *bandwidth*

Cara lain untuk memperoleh penduga *Nadaraya-Watson* ($\hat{m}(x)$), yaitu dengan meminimumkan kuadrat kesalahan *(square error)* yang diberi bobot, yaitu sebagai berikut :

$$E_{constant}(x) = \sum_{i=1}^{n} (Y_i - a_0)^2 \, w\left(\frac{x - X_i}{h}\right) \tag{2.15}$$

$w\left(\frac{x - X_i}{h}\right)$ adalah pembobotnya.

Dengan menurunkan $E_{constant}(x)$ terhadap a_0, maka akan diperoleh :

$$\hat{m}(x) = a_0 = \frac{\sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right)} \tag{2.16}$$

(lampiran 4)

Pada persamaan tersebut ditetapkan bahwa $w\left(\frac{x - X_i}{h}\right)$ sebagai $K\left(\frac{x - X_i}{h}\right)$.

Maka persamaan ini menghasilkan persamaan yang sama dengan persamaan (2.14). Dari cara ini, penduga *Nadaraya-Watson* menjadi sebuah metode untuk mendugai $m(x)$ dengan menggunakan $w\left(\frac{x - X_i}{h}\right)$ sebagai fungsi pembobot. Metode ini disebut sebagai metode pengepasan konstanta secara lokal *(method of fitting constants locally)* karena $m(x)$ terdiri dari konstanta-konstanta lokal. Oleh karena itu teknik ini disebut pendugaan konstanta lokal.

BAB III

REGRESI POLINOMIAL LOKAL

Data yang berbentuk pasangan $\{(X_i, Y_i)\}$, i = 1, ..., n, dalam ilmu statistika umumnya dapat dinyatakan dengan persamaan :

$$Y_i = m(X_i) + \varepsilon_i \tag{3.1}$$

$m(x)$ adalah suatu fungsi regresi yang menggambarkan keadaan data dan ingin ditaksir, dan ε_i adalah suatu variabel random yang menggambarkan variasi Y disekitar $m(x)$ (Hardle, 1990).

Salah satu cara menaksir fungsi regresi ini adalah dengan me-*fit* plot kurva dari data dengan model regresi polinomial dan menaksir parameter-parameter regresinya dengan metode *least square*. Namun metode ini tidak dapat begitu saja digunakan untuk pemulusan. Hal ini disebabkan oleh fakta bahwa hanya satu polinomial di-*fit* untuk seluruh wilayah data dengan kondisi data yang kompleks. Mungkin persamaan tersebut cocok untuk wilayah data tertentu namun tidak cocok untuk wilayah data yang lain, atau kurva regresi tersebut tidak sesuai dengan data. Pertimbangan-pertimbangan ini mengantarkan kita pada konsep pembagian data ke dalam beberapa wilayah dan mem-*fit* kurva untuk masing-masing wilayah. Selanjutnya, hasil dari masing-masing wilayah tersebut dihilangkan sehingga didapatkan dugaan fungsi regresi untuk seluruh wilayah nilai data. Konsep ini disebut *piecewise* *fitting* dari persamaan regresi. Namun banyak masalah baru yang muncul ketika mencoba *piecewise fitting*. Masalah tersebut berhubungan dengan penentuan banyaknya dan posisi dari *knot* (simpul) untuk *piecewise fitting.* Jika fungsi regresi yang digunakan untuk *piecewise fitting* cukup kompleks (dengan banyak koefisien regresi), ini akan menimbulkan kesulitan. Oleh karena itu regresi lokal dikembangkan untuk membahas pemodelan tersebut.

3.1 REGRESI POLINOMIAL LOKAL

Regresi polinomial lokal adalah suatu metode regresi nonparametrik , dimana fungsi regresi $m(x)$ ditaksir menggunakan bentuk polinomial. Pada regresi polinomial biasa persamaan regresi di-*fit* untuk seluruh wilayah data sedangkan dalam regresi polinomial lokal persamaan regresi di-*fit* hanya dalam lingkungan x^* saja.

Persamaan regresi pada polinomial lokal orde ke- p $(m(x, x^*))$, untuk nilai prediktor (x) yang dekat dengan x^*, didefinisikan sebagai berikut :

$$m(x, x^*) = a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(x - x^*)^j \qquad (3.2)$$

Persamaan (3.2) bersesuaian dengan sebuah hampiran yang menggunakan p suku pertama perluasan deret Taylor dari $m(x)$ di sekitar x^*. $\{a_0(x^*), a_1(x^*), a_2(x^*), \cdots, a_p(x^*)\}$ diperoleh dengan meminimumkan :

$$
\begin{aligned}
E_{lokal}(x^*) &= \sum_{i=1}^{n}\left(w\left(\frac{X_i - x^*}{h}\right)\left(m(X_i, x^*) - Y_i\right)^2 \right) \\
&= \sum_{i=1}^{n}\left(w\left(\frac{X_i - x^*}{h}\right)\left(a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(X_i - x^*)^j - Y_i \right)^2 \right)
\end{aligned}
\qquad (3.3)
$$

dengan :

x : Variabel prediktor yang nilainya tidak teramati namun akan digunakan untuk menaksir Y

x^* : nilai tetap yang ditentukan

h : *bandwidth*, yaitu konstanta positif yang menentukan lebarnya lingkungan di sekitar x^*

Y_i : variabel respon pengamatan $ke-i$

X_i : variabel prediktor pengamatan $ke-i$

$w(\cdot)$: fungsi kernel yang dijadikan sebagai fungsi pembobot.

Fungsi kernel biasanya adalah sebuah fungsi yang bernilai 0 atau bernilai positif dan w akan bernilai maksimum ketika $X_i - x^* = 0$ *(lampiran 5*), serta akan menurun ketika $|X_i - x^*|$ membesar *(lampiran 6)*. Ini berarti data yang berada dalam lingkungan x^* akan memberikan kontribusi yang lebih besar. $m(x, x^*)$ akan memberikan hampiran yang baik ketika x dekat dengan x^*. Sehingga taksiran $m(x, x^*)$ pada saat $x = x^*$ adalah $\hat{m}(x^*, x^*) = \hat{a}_0(x^*)$ dan $\{\hat{a}_0(x^*), \hat{a}_1(x^*), \hat{a}_2(x^*), \cdots, \hat{a}_p(x^*)\}$ didapat dengan cara yang sama seperti

$\{a_0(x^*), a_1(x^*), a_2(x^*), \cdots, a_p(x^*)\}$, yaitu dengan meminimumkan persamaan (3.3).

Fungsi kernel $w\left(\frac{X_i - x^*}{h}\right)$ yang akan digunakan adalah fungsi kernel Gaussian, yaitu :

$$w\left(\frac{X_i - x^*}{h}\right) = \exp\left(-\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right) \tag{3.4}$$

Fungsi kernel lain yang bisa digunakan sebagai pembobot adalah fungsi pembobot *bisquare* atau *biweight*, *triweight*, dan fungsi pembobot *tricube*, secara berurutan sebagai berikut :

$$w\left(\frac{X_i - x^*}{h}\right) = \begin{cases} \left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2 & jika\left(\frac{X_i - x^*}{h}\right)^2 \le 1 \\ 0 & jika\left(\frac{X_i - x^*}{h}\right)^2 > 1 \end{cases} \tag{3.5}$$

$$w\left(\frac{X_i - x^*}{h}\right) = \begin{cases} \left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^3 & jika\left(\frac{X_i - x^*}{h}\right)^2 \le 1 \\ 0 & jika\left(\frac{X_i - x^*}{h}\right)^2 > 1 \end{cases} \tag{3.6}$$

$$w\left(\frac{X_i - x^*}{h}\right) = \begin{cases} \left(1 - \left(\frac{|X_i - x^*|}{h}\right)^3\right)^3 & jika\left(\frac{|X_i - x^*|}{h}\right)^3 \le 1 \\ 0 & jika\left(\frac{|X_i - x^*|}{h}\right)^3 > 1 \end{cases} \tag{3.7}$$

Untuk mempermudah proses pe-minimum-an persamaan (3.3), maka persamaan tersebut dapat ditulis dalam notasi matriks sebagai berikut :

Matriks $\mathbf{X}$ adalah matriks berukuran n x (p+1) dan $\mathbf{X}$ di sini merupakan fungsi dari x^* :

$$\mathbf{X} = \begin{bmatrix} 1 & (X_1 - x^*) & (X_1 - x^*)^2 & \cdots & (X_1 - x^*)^p \\ 1 & (X_2 - x^*) & (X_2 - x^*)^2 & \cdots & (X_2 - x^*)^p \\ 1 & (X_3 - x^*) & (X_3 - x^*)^2 & \cdots & (X_3 - x^*)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x^*) & (X_n - x^*)^2 & \cdots & (X_n - x^*)^p \end{bmatrix} \tag{3.8}$$

matriks $\mathbf{y}$ adalah matriks kolom berukuran n x 1 : :

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \tag{3.9}$$

matriks diagonal $\mathbf{W}$ dengan ukuran n x n :

$$\mathbf{W} = \begin{bmatrix} w\left(\frac{X_1 - x^*}{h}\right) & 0 & \cdots & 0 \\ 0 & w\left(\frac{X_2 - x^*}{h}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w\left(\frac{X_n - x^*}{h}\right) \end{bmatrix} \tag{3.10}$$

serta matriks $\mathbf{a}$, berukuran (p+1) x 1 :

$$\mathbf{a} = \begin{bmatrix} a_0(x^*) \\ \vdots \\ a_p(x^*) \end{bmatrix} \tag{3.11}$$

Sehingga persamaan (3.3) dapat ditulis ulang dalam bentuk matriks sebagai berikut *(lampiran 7)* :

$$E_{lokal}(x^*) = (\mathbf{Xa} - \mathbf{y})^t \, \mathbf{W}(\mathbf{Xa} - \mathbf{y}) \tag{3.12}$$

Dengan meminimumkan $E_{lokal}(x^*)$ maka akan didapat :

$$\hat{\mathbf{a}} = [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W} \mathbf{y} \tag{3.13}$$

Hasil tersebut berasal dari proses perhitungan sebagai berikut :

Sebelum dilakukan pe-minimum-an, $E_{lokal}(x^*)$ diuraikan terlebih dahulu

$$E_{lokal}(x^*) = (\mathbf{Xa} - \mathbf{y})^t \, \mathbf{W}(\mathbf{Xa} - \mathbf{y})$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{p+1} [\mathbf{X}]_{ij} [\mathbf{a}]_j - \mathbf{Y}_i \right) [\mathbf{W}]_{ii} \left(\sum_{k=1}^{p+1} [\mathbf{X}]_{ik} [\mathbf{a}]_k - \mathbf{Y}_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} [\mathbf{X}]_{ij} [\mathbf{a}]_j [\mathbf{W}]_{ii} [\mathbf{X}]_{ik} [\mathbf{a}]_k - \sum_{i=1}^{n} \sum_{j=1}^{p+1} [\mathbf{X}]_{ij} [\mathbf{a}]_j [\mathbf{W}]_{ii} \mathbf{Y}_i -$$

$$\sum_{i=1}^{n} \sum_{k=1}^{p+1} \mathbf{Y}_i [\mathbf{W}]_{ii} [\mathbf{X}]_{ik} [\mathbf{a}]_k + \sum_{i=1}^{n} \mathbf{Y}_i [\mathbf{W}]_{ii} \mathbf{Y}_i$$

(3.14)

Selanjutnya, persamaan (3.14) diturunkan secara parsial terhadap $[\mathbf{a}]_l$,

dengan $1 \le l \le p+1$, sebagai berikut :

$$\frac{\partial E_{lokal}(x^*)}{\partial [\mathbf{a}]_l} = \sum_{i=1}^{n} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} [\mathbf{X}]_{ij} \frac{\partial [\mathbf{a}]_j}{\partial [\mathbf{a}]_l} [\mathbf{W}]_{ii} [\mathbf{X}]_{ik} [\mathbf{a}]_k + \sum_{i=1}^{n} \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} [\mathbf{X}]_{ij} [\mathbf{a}]_j [\mathbf{W}]_{ii} [\mathbf{X}]_{ik} \frac{\partial [\mathbf{a}]_k}{\partial [\mathbf{a}]_l}$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{p+1} [\mathbf{X}]_{ij} \frac{\partial [\mathbf{a}]_j}{\partial [\mathbf{a}]_l} [\mathbf{W}]_{ii} \mathrm{Y}_i - \sum_{i=1}^{n} \sum_{k=1}^{p+1} \mathrm{Y}_i [\mathbf{W}]_{ii} [\mathbf{X}]_{ik} \frac{\partial [\mathbf{a}]_k}{\partial [\mathbf{a}]_l}$$

dengan $\frac{\partial [\mathbf{a}]_i}{\partial [\mathbf{a}]_j} \quad \begin{cases} 1, \text{ jika i = j} \\ 0, \text{ jika i} \neq \text{j} \end{cases}$

maka :

$$\frac{\partial E_{lokal}(x^*)}{\partial a_l} = \sum_{i=1}^{n}\sum_{k=1}^{p+1}[\mathbf{X}]_{il}[\mathbf{W}]_{ii}[\mathbf{X}]_{ik}[\mathbf{a}]_k + \sum_{i=1}^{n}\sum_{j=1}^{p+1}[\mathbf{X}]_{ij}[\mathbf{a}]_j[\mathbf{W}]_{ii}[\mathbf{X}]_{il}$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{p+1}[\mathbf{X}]_{il}[\mathbf{W}]_{ii}\mathbf{Y}_i - \sum_{i=1}^{n}\sum_{k=1}^{p+1}\mathbf{Y}_i[\mathbf{W}]_{ii}[\mathbf{W}]_{il}$$

$$= 2\sum_{i=1}^{n}\sum_{j=1}^{p+1}[\mathbf{X}^t]_{li}[\mathbf{W}]_{ii}[\mathbf{X}]_{ij}[\mathbf{a}]_j - 2\sum_{i=1}^{n}[\mathbf{X}^t]_{li}[\mathbf{W}]_{ii}\mathbf{Y}_i \qquad (3.15)$$
$$= 0$$

$$2\sum_{i=1}^{n}\sum_{j=1}^{p+1}[\mathbf{X}^t]_{li}[\mathbf{W}]_{ii}[\mathbf{X}]_{ij}[\mathbf{a}]_j = 2\sum_{i=1}^{n}[\mathbf{X}^t]_{li}[\mathbf{W}]_{ii}\mathbf{Y}_i$$

$$\mathbf{X}^t\mathbf{W}\mathbf{X}\mathbf{a} = \mathbf{X}^t\mathbf{W}\mathbf{y}$$

$$[\mathbf{X}^t\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}^t\mathbf{W}\mathbf{X}\mathbf{a} = [\mathbf{X}^t\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y}$$

$$\mathbf{I}\mathbf{a} = [\mathbf{X}^t\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y}$$

$$\mathbf{a} = [\mathbf{X}^t\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y}$$

Maka $\hat{m}(x^*, x^*) = \hat{a}_0(x^*)$ yang merupakan taksiran $m(x, x^*)$ pada saat $x = x^*$ menjadi :

$$\hat{m}(x^*, x^*) = \begin{bmatrix} a_0(x^*) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_1^{\,t} \cdot \mathbf{a} = \mathbf{e}_1^{\,t}[\mathbf{X}^t\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y} \qquad (3.16)$$

dengan $\mathbf{e}_1$ adalah vektor kolom (1, 0, 0 …, 0)^t yang berukuran (p+1) x 1.

Selanjutnya, didefinisikan vektor kolom

$\mathbf{q}(x^*) = (q_1(x^*), q_2(x^*), \cdots, q_n(x^*))^t$ sebagai pembobot, yaitu :

$$\mathbf{q}(x^*)^t = \mathbf{e}_1{}^t [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W} \tag{3.17}$$

Dengan menyubstitusi $\mathbf{q}(x^*)^t$ ke dalam $\hat{m}(x^*, x^*)$ dapat ditulis :

$$\hat{m}(x^*, x^*) = \mathbf{q}(x^*)^t \mathbf{y} = \sum_{i=1}^{n} q_i(x^*) Y_i \tag{3.18}$$

Akan ditunjukkan bahwa :

$$\sum_{i=1}^{n} q_i(x^*) = 1 \tag{3.19}$$

Dengan menggunakan persamaan (3.17), dan kedua sisinya dikalikan dengan $\mathbf{X}$ dari sebelah kanan, maka diperoleh :

$$\begin{aligned} \mathbf{q}(x^*)^t \mathbf{X} &= \mathbf{e}_1{}^t [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W} \mathbf{X} \\ &= \mathbf{e}_1{}^t \end{aligned}$$

Maka

$$\mathbf{q}(x^*)^t \mathbf{X} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

$$\begin{bmatrix} q_1(x^*) & \cdots & q_n(x^*) \end{bmatrix} \begin{bmatrix} 1 & (X_1 - x^*) & (X_1 - x^*)^2 & \cdots & (X_1 - x^*)^p \\ 1 & (X_2 - x^*) & (X_2 - x^*)^2 & \cdots & (X_2 - x^*)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x^*) & (X_n - x^*)^2 & \cdots & (X_n - x^*)^p \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

$$\begin{bmatrix} q_1(x^*) + \cdots + q_n(x^*) & q_1(x^*)(X_1 - x^*) + \cdots + q_n(x^*)(X_n - x^*) & \cdots & q_1(x^*)(X_1 - x^*)^p + \cdots + q_n(X_n - x^*)^p \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

$$\begin{bmatrix} \sum_{i=1}^{n} q_i(x^*) & \sum_{i=1}^{n} q_1(x^*)(X_i - x^*) & \cdots & \sum_{i=1}^{n} q_i(x^*)(X_i - x^*)^p \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

dengan menyamakan tiap entri matriks yang bersesuaian maka diperoleh :

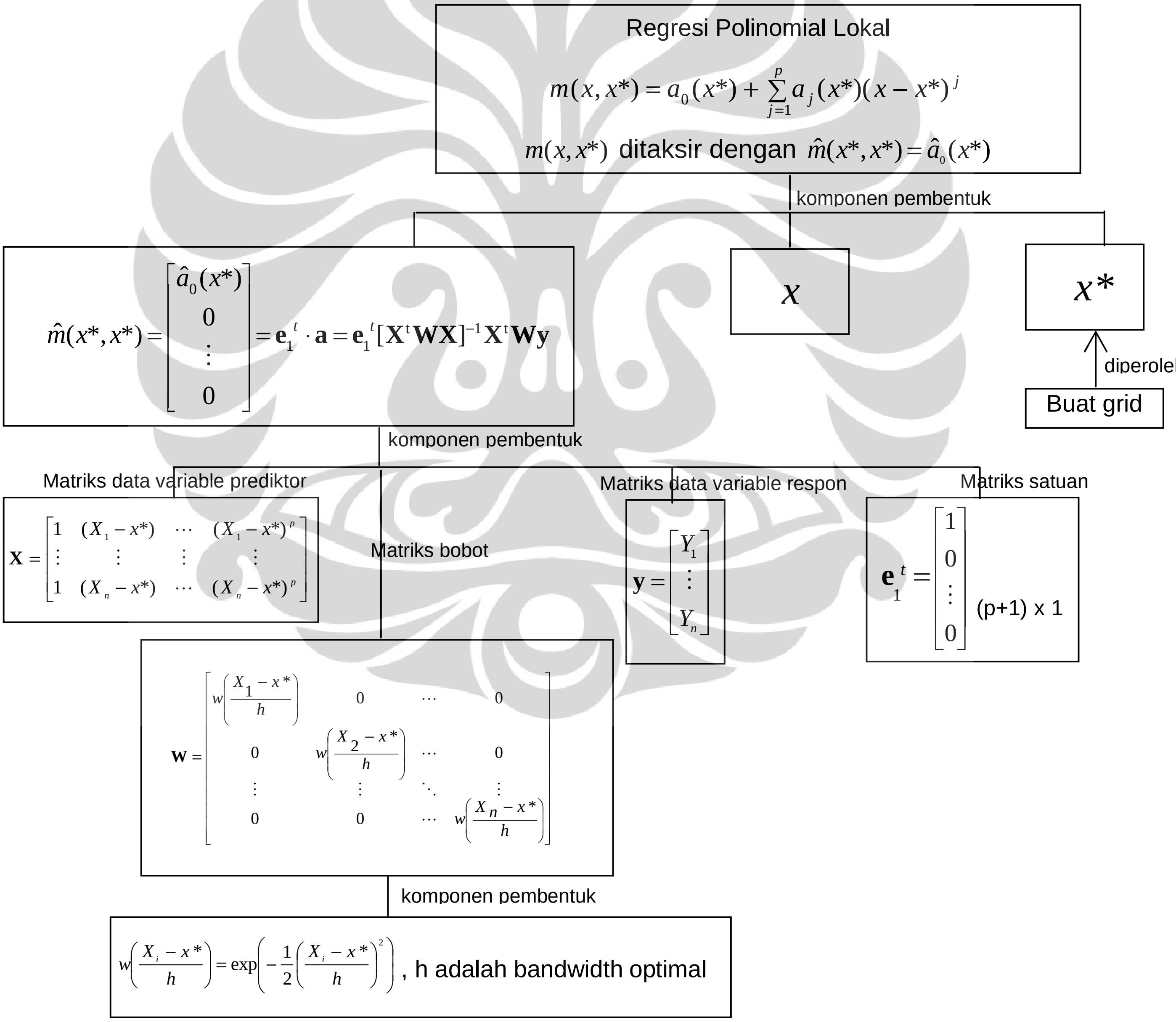
$$\sum_{i=1}^{n} q_i(x^*) = 1 \qquad (3.20)$$

Persamaan (3.18) dan (3.20) mengindikasikan bahwa sebuah dugaan yang diberikan oleh regresi polynomial lokal adalah pembobotan rata-rata (*weighted average)* dari $\{Y_i\}$.

3.2 LANGKAH KERJA REGRESI POLINOMIAL LOKAL

Proses perhitungan taksiran regresi polinomial lokal secara ringkas dijelaskan dengan bagan berikut :

LANGKAH PEMBENTUKAN $\hat{m}(x, x^*)$



Sedangkan prosedur pemilihan *bandwidth* yang optimal dengan CV atau GCV adalah sebagai berikut :

- Prosedur Pemilihan Bandwidth dengan CV

$$CV[\hat{m}(x)] = \sum_{i=1}^{n} \frac{(Y_i - \hat{m}(X_i))^2}{n.(1-[\mathbf{H}]_{ii})^2}$$

Tentukan barisan *bandwidth* $\{h_1, h_2, \cdots, h_{10}\}$ yang nilainya berada disekitar $\hat{h}_0$,

$\hat{h}_0 \approx 1.06\hat{\sigma}n^{-1/5}$ *(Rule of Thumb)*, dengan $\hat{\sigma}$ adalah standar deviasi dari Y dan n adalah banyak pengamatan (Hardle, 1990).

h_1

Bentuk $m(x, x^*) = a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(x - x^*)^j$ sebanyak n kali, dengan $x = x^* = X_i$

dengan $[\mathbf{H}]_{ii} = [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}]_{11} \cdot w(0)$

Hitung $CV[\hat{m}(x)] = \sum_{i=1}^{n} \frac{(Y_i - \hat{m}(X_i))^2}{n.(1-[\mathbf{H}]_{ii})^2}$

•
•
•

h_{10}

Bentuk $m(x, x^*) = a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(x - x^*)^j$ sebanyak n kali, dengan $x = x^* = X_i$

dengan $[\mathbf{H}]_{ii} = [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}]_{11} \cdot w(0)$

Hitung $CV[\hat{m}(x)] = \sum_{i=1}^{n} \frac{(Y_i - \hat{m}(X_i))^2}{n.(1-[\mathbf{H}]_{ii})^2}$

CV_j minimum → h pada CV_j optimal

- Prosedur Pemilihan Bandwidth dengan GCV

$$GCV = \frac{\sum_{i=1}^{n}(Y_i - \hat{m}(X_i))^2}{n \cdot \left(1 - \frac{\sum_{i=1}^{n}[\mathbf{H}]_{ii}}{n}\right)^2}$$

Tentukan barisan *bandwidth* $\{h_1, h_2, \cdots, h_{10}\}$ yang nilainya berada disekitar $\hat{h}_0$, $\hat{h}_0 \approx 1.06\hat{\sigma}n^{-1/5}$ *(Rule of Thumb)*, dengan $\hat{\sigma}$ adalah standar deviasi dari Y dan n adalah banyak pengamatan (Hardle, 1990).

h_1

Bentuk $m(x, x^*) = a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(x - x^*)^j$ sebanyak n kali, dengan $x = x^* = X_i$

Dengan $[\mathbf{H}]_{ii} = [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}]_{11} \cdot w(0)$

Hitung

$$GCV = \frac{\sum_{i=1}^{n}(Y_i - \hat{m}(X_i))^2}{n \cdot \left(1 - \frac{\sum_{i=1}^{n}[\mathbf{H}]_{ii}}{n}\right)^2}$$

•
•
•

h_{10}

Bentuk $m(x, x^*) = a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(x - x^*)^j$ sebanyak n kali, dengan $x = x^* = X_i$

dengan $[\mathbf{H}]_{ii} = [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}]_{11} \cdot w(0)$

Hitung

$$GCV = \frac{\sum_{i=1}^{n}(Y_i - \hat{m}(X_i))^2}{n \cdot \left(1 - \frac{\sum_{i=1}^{n}[\mathbf{H}]_{ii}}{n}\right)^2}$$

GCV_j minimum → h pada GCV_j optimal

3.3 BIAS DAN VARIANSI TAKSIRAN REGRESI POLINOMIAL LOKAL

Akan dicari bias dari taksiran $\hat{m}(x^*, x^*)$.

Bias dari suatu taksiran adalah :

$$Bias(\hat{t}) = E(\hat{t}) - t \tag{3.21}$$

Sehingga bias dari taksiran $\hat{m}(x^*, x^*)$ adalah :

$$Bias[\hat{m}(x^*, x^*)] = E[\hat{m}(x^*, x^*)] - m(x^*, x^*) \tag{3.22}$$

Dengan menggunakan persamaan (3.16) :

$$\hat{m}(x^*, x^*) = \mathbf{e}_1{}^t [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W} \mathbf{y}$$

dengan $Y_i = m(X_i) + \varepsilon_i$ dan $E[\varepsilon_i] = 0, (1 \le i \le n)$, maka $E[\hat{m}(x^*, x^*)]$:

$$E[\hat{m}(x^*, x^*)] = E[\mathbf{e}_1{}^t [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}(\mathbf{m} + \varepsilon)]$$

$$= \mathbf{e}_1{}^t [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W} \mathbf{m} \tag{3.23}$$

dengan $\mathbf{m} = (m((X_1), m(X_2), \cdots, m(X_n))^t$, vektor kolom (dengan n elemen) yang menyatakan nilai dari $m(x)$ yang terletak pada $\{X_i\}$ dan $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^t$.

Perluasan Taylor dari $m(x)$ dalam lingkungan x^* adalah :

$$m(x) = m(x^*) + m^{(1)}(x^*)(x - x^*) + \frac{m^{(2)}(x^*)}{2!}(x - x^*)^2 + \cdots + \frac{m^{(p)}(x^*)}{p!}(x - x^*)^p + r(x) \tag{3.24}$$

Dengan $m^{(p)}(x)$ adalah turunan ke- p dari $m(x)$ terhadap x dan $p!$ adalah faktorial dari p. $r(x)$ adalah sisa yaitu :

$$r(x) = m(x) - m(x^*) + m^{(1)}(x^*)(x - x^*) + \frac{m^{(2)}(x^*)}{2!}(x - x^*)^2 + \cdots + \frac{m^{(p)}(x^*)}{p!}(x - x^*)^p \tag{3.25}$$

Sehingga persamaan (3.23) dapat diubah menjadi :

$$E[\hat{m}(x^*, x^*)] = \mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}\mathbf{r}$$

$$= m(x^*) + \mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}\mathbf{r} \qquad (3.26)$$

dengan $\boldsymbol{\beta} = \left(m(x^*), m^{(1)}(x^*), \dfrac{m^{(2)}(x^*)}{2}, \cdots, \dfrac{m^{(p)}(x^*)}{p!} \right)^t$, dan

$$\mathbf{r} = \left(r(X_1), r(X_2), \cdots, r(X_n) \right)^t .$$

Sehingga didapat :

$$Bias[\hat{m}(x^*, x^*)] = E[\hat{m}(x^*, x^*)] - m(x^*, x^*)$$

$$= m(x^*, x^*) + \mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}\mathbf{r} - m(x^*, x^*)$$

$$Bias[\hat{m}(x^*, x^*)] = \mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}\mathbf{r} \qquad (3.27)$$

Dari persamaan (3.27) terlihat bahwa $Bias[\hat{m}(x^*, x^*)]$ adalah $\hat{r}(x^*)$ yaitu residual dari regresi polinomial lokal.

Selanjutnya akan dicari estimasi variansi dari $\hat{m}(x^*, x^*)$ yang diperoleh dengan regresi polinomial lokal.

Variansi dari suatu taksiran adalah :

$$Var(\hat{t}) = E[\hat{t} - E(\hat{t})]^2 \qquad (3.28)$$

Sehingga variansi dari $\hat{m}(x^*, x^*)$ adalah sebagai berikut :

$$Var[\hat{m}(x^*, x^*)] = E[(\hat{m}(x^*, x^*) - E[\hat{m}(x^*, x^*)])^2]$$

$$= E\left(\left(\sum_{i=1}^{n} q_i(x^*)\left(m(X_i) + \varepsilon_i\right) - \sum_{i=1}^{n} q_i(x^*) m(X_i)\right)\right.$$

$$\left(\sum_{j=1}^{n} q_j(x^*)\big(m(X_j)+\varepsilon_j\big)-\sum_{j=1}^{n} q_j(x^*)m(X_j)\right)\Bigg)$$

$$= E\left(\left(\sum_{i=1}^{n} q_i(x^*)m(X_i)+\sum_{i=1}^{n} q_i(x^*)\varepsilon_i-\sum_{i=1}^{n} q_i(x^*)m(X_i)\right)\right.$$

$$\left.\left(\sum_{j=1}^{n} q_j(x^*)m(X_j)+\sum_{j=1}^{n} q_j(x^*)\varepsilon_j-\sum_{j=1}^{n} q_j(x^*)m(X_j)\right)\right)$$

$$= E\left(\left(\sum_{i=1}^{n} q_i(x^*)\varepsilon_i\right)\left(\sum_{j=1}^{n} q_j(x^*)\varepsilon_j\right)\right)$$

$$= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} q_i(x^*)q_j(x^*)\varepsilon_i\varepsilon_j\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} q_i(x^*)q_j(x^*)\delta_{ij}\sigma^2$$

$$= \sigma^2 \cdot \sum_{i=1}^{n} q_i(x^*)^2 \tag{3.29}$$

dengan :

$$E[\varepsilon_i\varepsilon_j] = \sigma^2 \cdot \delta_{ij}$$

δ_{ij} didefinisikan sebagai :

$$\delta_{ij} = \begin{cases} 1 & \text{jika i = j} \\ 0 & \text{jika i} \neq \text{j} \end{cases}$$

dan σ^2 adalah variansi dari ε_i.

Maka variansi dari $\hat{m}(X_k, X_k)$ adalah :

$$Var[\hat{m}(X_k, X_k)] = \sigma^2 \sum_{i=1}^{n} q_i(X_k)^2$$

$$= \sigma^2 \sum_{i=1}^{n} \left([H]_{ki}\right)^2 \qquad \textit{(lampiran 8)}$$

Selanjutnya akan dibuktikan $Var[\hat{m}(X_k, X_k)] \leq \sigma^2$:

$$E[(\hat{m}(X_k) - E[\hat{m}(X_k)])^2]$$

$$= Var[\hat{m}(X_k)]$$

$$= \sigma^2 \cdot \sum_{i=1}^{n} ([\mathbf{H}]_{ki})^2$$

$$= \sigma^2 \sum_{i=1}^{n} \mathrm{e}_1^{\,t} \left((\mathrm{X}^t \mathrm{W} \mathrm{X})^{-1} \mathrm{X}^t\right) \mathrm{e}'_i \cdot w\left(\frac{X_i - X_k}{h}\right)$$

$$\mathrm{e}_1^{\,t} \left((\mathrm{X}^t \mathrm{W} \mathrm{X})^{-1} \mathrm{X}^t\right) \mathrm{e}'_i \cdot w\left(\frac{X_i - X_k}{h}\right)$$

$$= \sigma^2 \sum_{i=1}^{n} \mathrm{e}_1^{\,t} \left((\mathrm{X}^t \mathrm{W} \mathrm{X})^{-1} \mathrm{X}^t\right) \mathrm{e}'_i \, (\mathrm{e}'_i)^t \left(\mathbf{X}(\mathrm{X}^t \mathrm{W} \mathrm{X})^{-1}\right) \mathrm{e}_1 \cdot w\left(\frac{X_i - X_k}{h}\right)^2$$

$$\leq \sigma^2 \sum_{i=1}^{n} \mathrm{e}_1^{\,t} \left((\mathrm{X}^t \mathrm{W} \mathrm{X})^{-1} \mathrm{X}^t\right) \mathrm{e}'_i \, (\mathrm{e}'_i)^t \left(\mathrm{X}(\mathrm{X}^t \mathrm{W} \mathrm{X})^{-1}\right) \mathrm{e}_1 \cdot w\left(\frac{X_i - X_k}{h}\right) \cdot w(0)$$

$$= \sigma^2 \mathbf{e}_1^t \left((\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{X} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1}\right) \mathbf{e}_1 \cdot w(0)$$

$$= \sigma^2 \mathbf{e}_1^t \left((\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1}\right) \mathbf{e}_1 \cdot w(0)$$

$$= \sigma^2 [(\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1}]_{11} \cdot w(0)$$

$$= \sigma^2 [\mathbf{H}]_{kk} \tag{3.30}$$

dengan $\mathbf{W} = \sum_{j=1}^{n} e_j'(e_j')^t \cdot w\left(\frac{X_j - X_k}{h}\right)$.

Maka dari pertidaksamaan (3.23) didapat :

$$\sum_{i=1}^{n} ([\mathbf{H}]_{ki})^2 \leq [\mathbf{H}]_{kk} \tag{3.31}$$

karena $([\mathbf{H}]_{kk})^2 \leq \sum_{i=1}^{n} ([\mathbf{H}]_{ki})^2$, maka diperoleh :

$$([\mathbf{H}]_{kk})^2 \leq \sum_{i=1}^{n} ([\mathbf{H}]_{ki})^2 \leq [\mathbf{H}]_{kk}$$

sehingga :

$$0 \leq ([\mathbf{H}]_{kk})^2 \leq [\mathbf{H}]_{kk} \tag{3.32}$$

Oleh karena itu , didapat hubungan :

$$0 \leq [\mathbf{H}]_{kk} \leq 1 \tag{3.33}$$

Dengan menyubstitusi persamaan (3.33) ke persamaan (3.30) dihasilkan :

$$Var[\hat{m}(X_k)] \leq \sigma^2 \tag{3.34}$$

Hasil tersebut telah membuktikan bahwa variansi dari estimasi pemulusan dengan regresi polinomial lokal lebih kecil daripada variansi error.

BAB IV

APLIKASI REGRESI POLINOMIAL LOKAL

Untuk melengkapi pembahasan metode regresi polinomial lokal, bab ini membahas contoh data yang dapat diselesaikan dengan metode regresi polinomial lokal.

4.1 PENDAHULUAN

Untuk mendapatkan taksiran regresi dengan menggunakan metode regresi polinomial lokal, dilakukan langkah-langkah sebagai berikut :

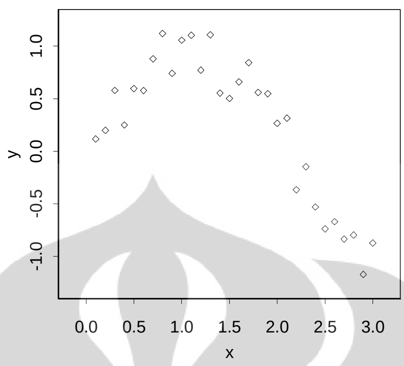
1. Penyedian data.
2. Plot data.
3. Mencari *bandwidth* yang optimal (*bandwidth* yang meminimumkan GCV).
4. Melakukan penaksiran regresi polinomial lokal.

Data yang digunakan adalah data Test31 (*Introduction to Nonparametric Regression, Kunio Takezawa, 2006*) *(lampiran 9)* dan data demam berdarah tahun 2008 Kota Depok *(lampiran 10)*.

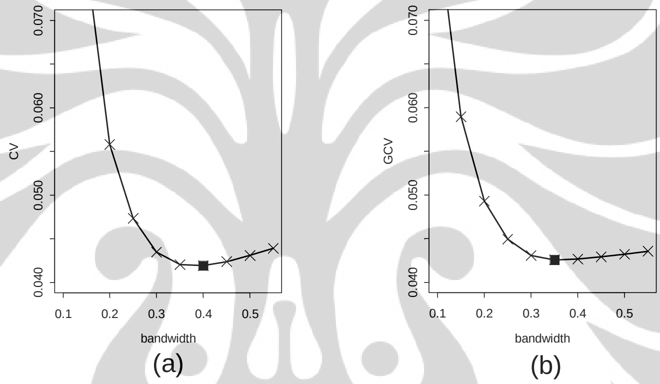
4.2 APLIKASI REGRESI POLINOMIAL LOKAL PADA DATA TEST31

Data Test31 merupakan data dengan 2 variabel yaitu X sebagai variabel prediktor dan Y sebagai variabel respon, dan terdiri dari 30 pengamatan.

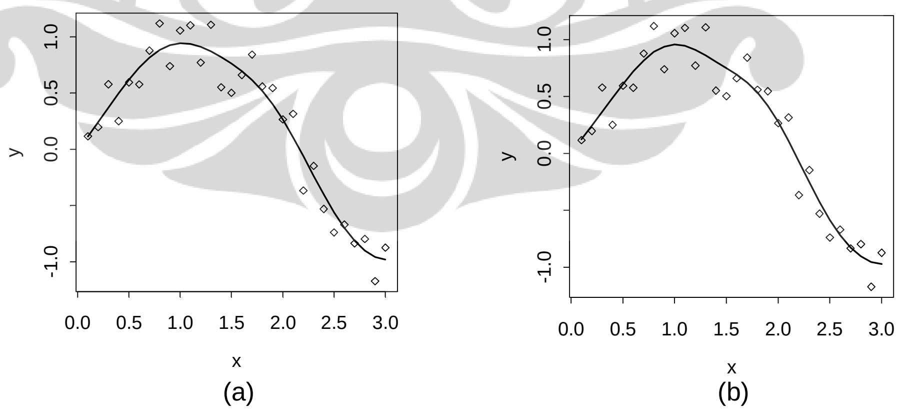
Gambar 4.1 memperlihatkan plot data Test31 yang sebarannya cenderung dapat diikuti oleh suatu kurva. Pada plot ini akan dilakukan pendugaan kurva menggunakan metode regresi polinomial lokal dengan *bandwidth* optimal yang ditunjukkan pada Gambar 4.2. *Bandwidth* optimal yang meminimumkan CV adalah h = 0.4 dan *bandwidth* optimal yang meminimumkan GCV adalah h = 0.35. Selanjutnya Gambar 4.3 memperlihatkan kurva dugaan menggunakan metode regresi polinomial lokal derajat 2 (p = 2) dengan fungsi kernel Gaussian. Gambar 4.3 (a) menunjukkan $\hat{m}(x)$ dengan h = 0.4, yaitu *bandwidth* yang meminimumkan CV. Gambar 4.3 (b) menunjukkan $\hat{m}(x)$ dengan h = 0.35, yaitu bandwidth yang meminimumkan GCV. Pada kurva dengan h = 0.35, lekukan kurva terlihat lebih jelas dibanding kurva dengan h = 0.4.



Gambar 4.1 Plot data Test31



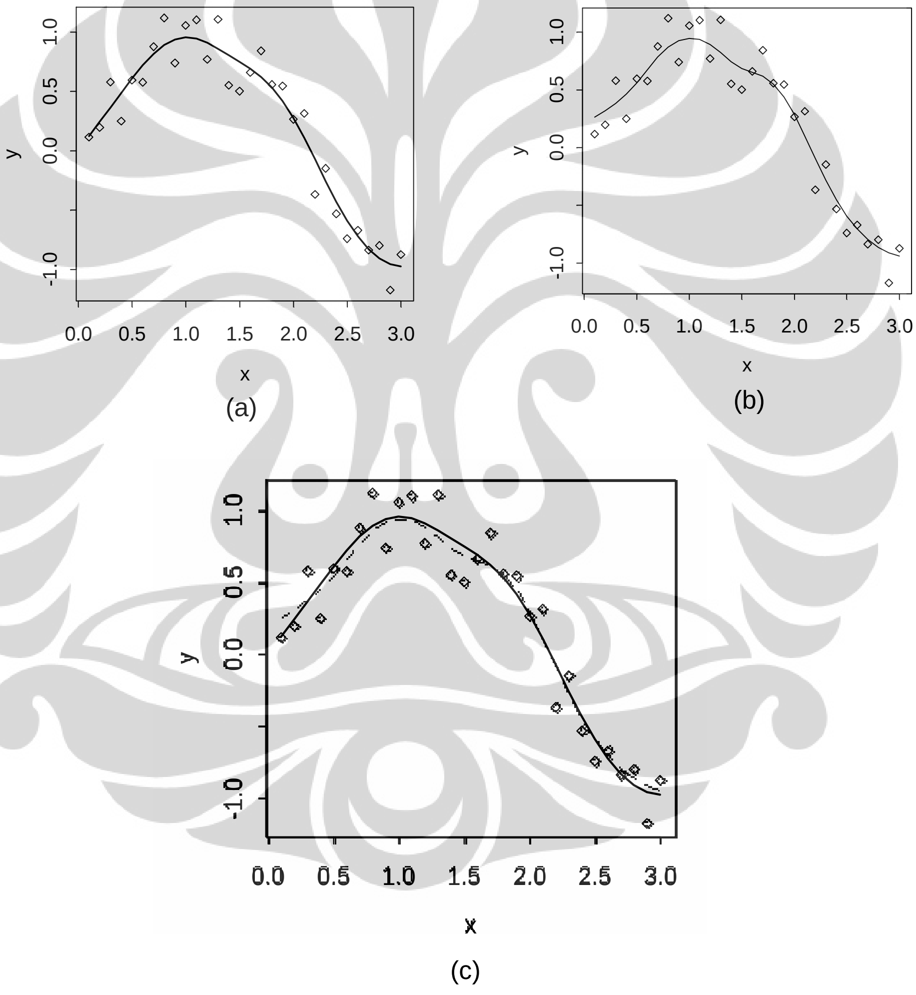
Gambar 4.2 Plot (a) CV vs h, (b) GCV vs h



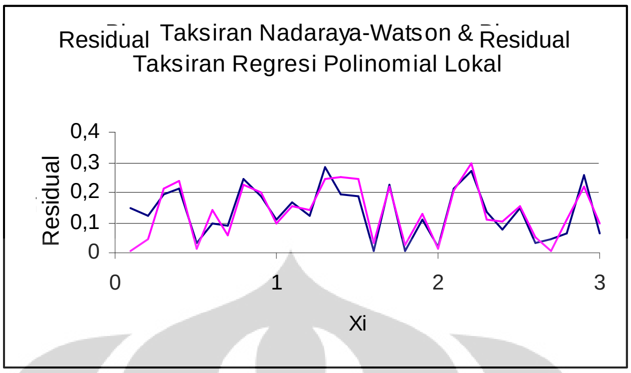
Gambar 4.3 Plot $\hat{m}(x)$ dengan metode Regresi Polinomial Lokal (a) h = 0.4, (b) h = 0.35

4.3 PERBANDINGAN METODE REGRESI POLINOMIAL LOKAL DENGAN METODE NADARAYA-WATSON PADA DATA TEST31

Untuk melihat kelebihan dan kekurangan dari metode regresi polinomial lokal, maka dilakukan pembandingan dengan metode Nadaraya Watson.



Gambar 4.4 (a) plot $\hat{m}(x)$ dengan metode Regresi Polinomial Lokal, (b) plot $\hat{m}(x)$ dengan metode Nadaraya Watson, (c) plot gabungan $\hat{m}(x)$ dengan metode Regresi Polinomial Lokal (padat), $\hat{m}(x)$ dengan metode Nadaraya Watson (putus-putus).



Gambar 4.5 Residual $\hat{m}(x)$ metode Regresi Polinomial Lokal (pink) dan bias $\hat{m}(x)$ metode Nadaraya-Watson (biru)

Gambar 4.4 (a) menunjukkan kurva dugaan menggunakan metode regresi polinomial lokal berderajat 2 (p = 2) dengan fungsi *kernel Gaussian* dan h = 0.35. Gambar 4.4 (b) menunjukkan kurva dugaan dengan metode Nadaraya-Watson dengan fungsi *kernel Gaussian* dan h = 0.18 (h optimal untuk Nadaraya-Watson). Gambar 4.4 (c) menunjukkan gabungan dari kedua kurva tersebut, dari Gambar 4.4 (c), terlihat bahwa hasil dugaan kurva antara metode regresi polinomial lokal dan Nadaraya-Watson tidak jauh berbeda, hanya saja pada ujung-ujung kurva regresi yang menggunakan metode regresi polinomial lokal terlihat lebih dekat dengan data dibanding kurva regresi yang menggunakan metode Nadaraya-Watson.

Gambar 4.5 memperlihatkan residual dari regresi polinomial lokal dan residual Nadaraya-Watson. Dari gambar ini terlihat mutlak residual regresi polinomial lokal lebih kecil dari mutlak residual Nadaraya-Watson pada saat ujung data saja, selebihnya tidak jauh berbeda.

Tabel 4.1 adalah perhitungan residual $\hat{m}(X_i, X_i)$ metode regresi polinomial lokal dan Nadaraya-Watson *(lampiran 11)* :

Tabel 4.1 Residual $\hat{m}(X_i, X_i)$ dari Metode Regresi Polinomial Lokal dan Nadaraya-Watson Data Test31

I	X_i	Y_i	$\hat{m}(x)$		Residual	
			Nadaraya-Watson	**Regresi Polinomial Lokal (p = 2)**	**Nadaraya-Watson**	**Regresi Polinomial Lokal**
1	**0.1**	**0.1151**	**0.26318520**	**0,124217**	**0,148085**	**0,009117**
2	**0.2**	**0.19762**	**0.31751713**	**0,244089**	**0,119897**	**0,046469**
3	0.3	0.578587	0.38292012	0,365024	-0,19567	-0,21357
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	2.8	-0.79768	-0.86413664	-0,90657	-0,06646	-0,10889
29	2.9	-1.17407	-0.91339296	-0,9564	0,260677	0,217671
30	3	-0.87411	-0.94115323	-0,97399	-0,06704	-0,09988

Dari Tabel 4.1 terlihat bahwa residual dari taksiran yang menggunakan metode regresi polinomial lokal pada data pertama dan data kedua sangat kecil dibandingkan dengan residual dari taksiran yang menggunakan metode Nadaraya-Watson, tetapi untuk data-data lainnya residual taksiran dari kedua metode tidak jauh berbeda.

4.4 APLIKASI REGRESI POLINOMIAL LOKAL PADA DATA DEMAM BERDARAH DI KOTA DEPOK TAHUN 2008

4.4.1 Latar Belakang Masalah

Kota Depok adalah sebuah kota di Provinsi Jawa Barat yang terletak tepat di selatan Jakarta, yakni antara Jakarta dan Bogor. Sejak tanggal 20 April 1999, Depok ditetapkan menjadi kotamadya yang terpisah dari Kabupaten Bogor. Kota Depok terdiri atas 6 kecamatan dan dibagi menjadi 63 kelurahan.

Depok merupakan kota penyangga Jakarta, karena Depok merupakan kota yang berada di sekitar Jakarta yang menyokong Jakarta untuk menyediakan kebutuhan bahan pangan dan tenaga kerja. Pada saat menjadi kota administratif pada tahun 1982, penduduknya berjumlah sekitar 240.000 jiwa, dan ketika menjadi kotamadya pada tahun 1999 penduduknya meningkat menjadi sekitar 1,2 juta jiwa, sedangkan pada tahun 2008 jumlah penduduknya telah mencapai 1.304.643 jiwa (sumber : Dinas Kependudukan Kota Depok tahun 2008) dengan kepadatan penduduk 47,02 jiwa/km^2.

Kota Depok merupakan salah satu kota yang setiap tahunnya terkena wabah penyakit Demam Berdarah Dengue (DBD). DBD merupakan salah satu penyakit menular yang berbasis lingkungan. Artinya, kejadian dan penularannya dipengaruhi oleh berbagai faktor lingkungan. Lingkungan yang padat penduduk dan sanitasi yang buruk dapat meningkatkan perkembangbiakan nyamuk demam berdarah, sehingga meningkatkan jumlah korban demam berdarah.

4.4.2 Permasalahan

Ingin dicari kurva dugaan dengan menggunakan regresi polinomial lokal berderajat 2 (p=2), dengan variabel prediktor adalah kepadatan penduduk di tiap kelurahan di Kota Depok dan variabel respon adalah banyaknya kasus DBD di tiap kelurahan tersebut pada tahun 2008.

4.4.3 Data

Data yang digunakan sebagai berikut :

1. Data kepadatan penduduk Kota Depok per kelurahan yang bersumber dari "Kota Depok Dalam Angka Tahun 2008", oleh Badan Pusat Statistik Kota Depok.
2. Data rekapitulasi kasus DBD di Kota Depok menurut kelurahan tahun 2008, yang bersumber dari Dinas Kesehatan Kota Depok.

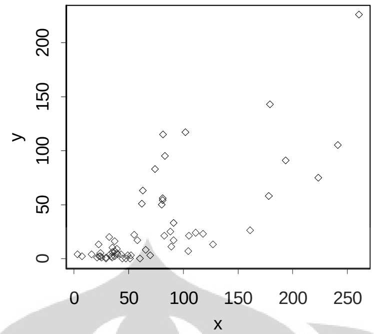
Dari kedua sumber data diperoleh angka kepadatan penduduk dan jumlah kasus DBD per kelurahan (desa) dengan 63 kelurahan. Sehingga dalam kasus ini terdapat 63 pengamatan

4.4.4 Tujuan

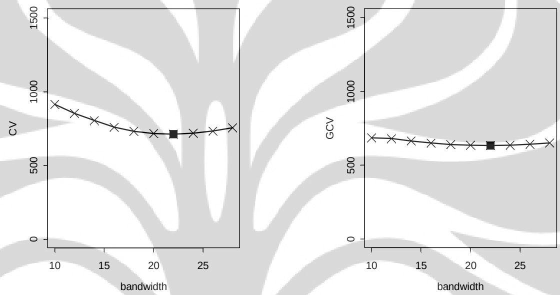
Tujuan dari aplikasi ini adalah mencari kurva dugaan jumlah kasus DBD berdasarkan angka kepadatan penduduk di Kota Depok menggunakan regresi polinomial lokal berderajat 2 (p=2), dengan variabel prediktor adalah kepadatan penduduk di setiap kelurahan di Kota Depok dan variabel target adalah banyaknya kasus DBD di setiap kelurahan tersebut pada tahun 2008.

4.4.5 Pengolahan dan Analisis Data

Berikut ini adalah hasil pengolahan dan analisis data mengenai keterkaitan antara banyaknya kasus DBD dan kepadatan penduduk di Kota Depok tahun 2008.

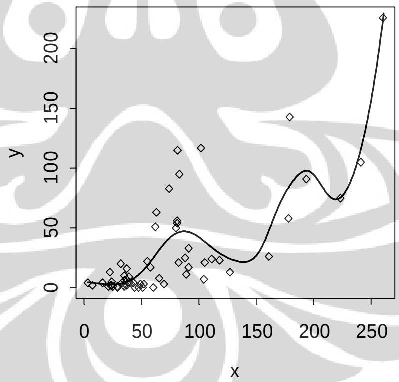


Gambar 4.6 Plot Data DBD 2008



(a) (b)

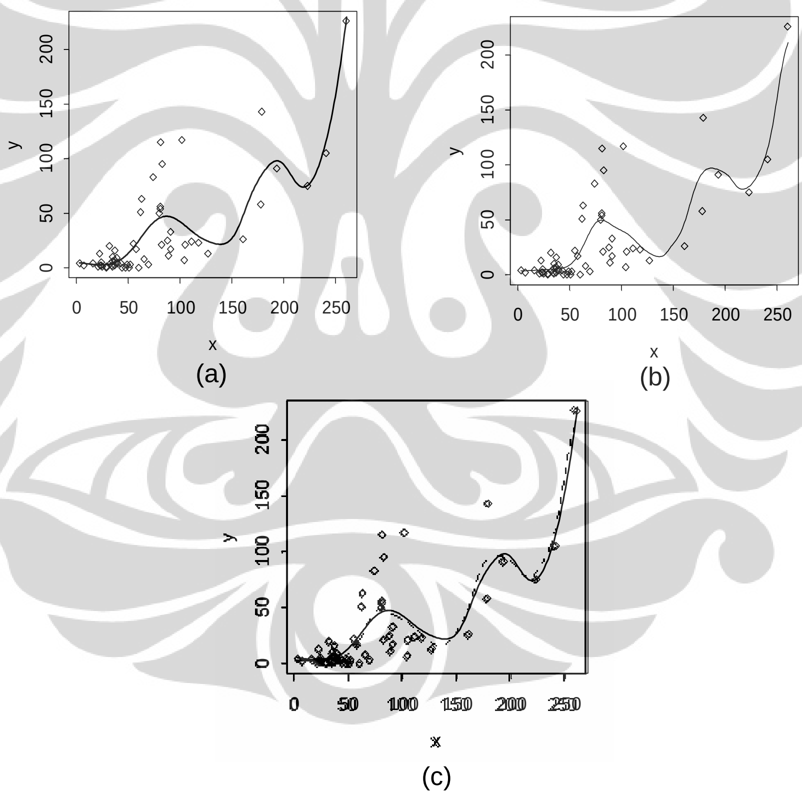
Gambar 4.7 Plot (a) CV vs h, (b) GCV vs h



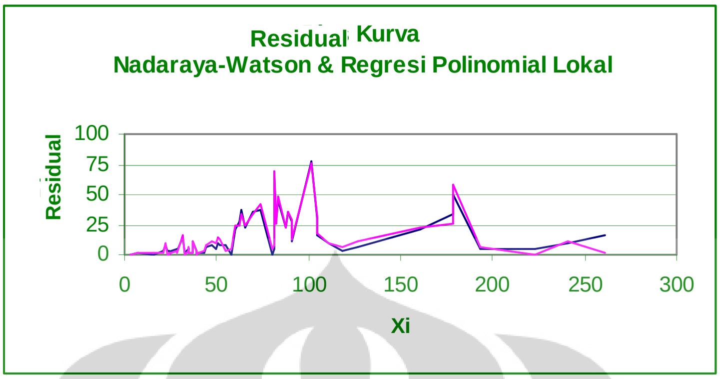
Gambar 4.8 Plot $\hat{m}(x)$ dengan metode Regresi Polinomial Lokal

Gambar 4.6 menunjukkan plot antara banyaknya kasus DBD (variabel respon) dengan kepadatan penduduk (variabel prediktor) di 63 kelurahan di Kota Depok. Pada Gambar 4.7 (a), *bandwidth* optimal yang meminimumkan CV adalah h = 22 dan pada Gambar 4.7 (b), *bandwidth* optimal yang meminimumkan GCV adalah h = 22 juga. Selanjutnya Gambar 4.8 menunjukkan kurva dugaan menggunakan metode regresi polinomial lokal derajat 2 (p = 2) dengan fungsi *kernel Gaussian*, dan h = 22. Dengan kurva regresi ini kita dapat menaksir banyaknya kasus DBD pada suatu daerah (di Kota Depok) bila diketahui besar kepadatan penduduknya.

Untuk melihat kelebihan dan kekurangan dari metode regresi polinomial lokal, maka pada data DBD ini dilakukan pembandingan dengan metode Nadaraya-Watson.



Gambar 4.9 (a) plot $\hat{m}(x)$ dengan metode Regresi Polinomial Lokal, (b) plot $\hat{m}(x)$ dengan metode Nadaraya Watson, (c) plot gabungan $\hat{m}(x)$ dengan metode Regresi Polinomial Lokal (padat), $\hat{m}(x)$ dengan metode Nadaraya Watson (putus-putus).



Gambar 4.10 Residual $\hat{m}(x)$ metode Regresi Polinomial Lokal (pink) dan Bias $\hat{m}(x)$ dengan metode Nadaraya-Watson (biru).

Gambar 4.9 (a) menunjukkan kurva dugaan menggunakan metode regresi polinomial lokal berderajat 2 (p = 2) dengan fungsi kernel Gaussian dan h = 22. Gambar 4.9 (b) menunjukkan kurva dugaan dengan metode Nadaraya-Watson dengan fungsi kernel Gaussian dan h = 10 (h optimal untuk metode Nadaraya-Watson). Gambar 4.9 (c) menunjukkan gabungan dari kedua kurva tersebut, dari gambar tersebut terlihat hasil dugaan kurva antara metode regresi polinomial lokal dengan Nadaraya-Watson tidak jauh berbeda. Pada ujung kurva regresi dengan metode regresi polinomial lokal terlihat lebih dekat dengan data dibanding kurva regresi dengan metode Nadaraya-Watson.

Gambar 4.10 memperlihatkan residual dari regresi polinomial lokal dan residual Nadaraya-Watson, dari gambar ini terlihat mutlak residual regresi polinomial lokal lebih kecil dari mutlak residual Nadaraya-Watson pada bagian ujung data saja, selebihnya tidak jauh berbeda.

Tabel 4.2 adalah perhitungan residual $\hat{m}(X_i, X_i)$ metode regresi polinomial lokal dan Nadaraya-Watson *(lampiran 12)* :

Tabel 4.1 Residual $\hat{m}(X_i, X_i)$ dari Metode Regresi Polinomial Lokal dan Nadaraya-Watson Data DBD 2008

i	**Xi**	**Yi**	$\hat{m}(X_i, X_i)$		**Residual**	
			Nadaraya-Watson	**Regresi Polinomial Lokal (p=2)**	**Nadaraya-Watson**	**Regresi Polinomial Lokal (p=2)**
1	3.2	4	3.473327	4.526208	-0.52667	0.526208
2	7.28	2	3.628954	4.050214	1.628954	2.050214
3	16	4	4.078972	3.068864	0.078972	-0.93114
⋮	⋮	⋮	⋮	⋮	⋮	⋮
61	223.32	75	80.35957	75.3176	5.35957	0.3176
62	241.09	105	114.2572	115.6965	9.2572	10.6965
63	**260.39**	**226**	**209.6146**	**224.7358**	**-16.3854**	**-1.2642**

4.5 PERBANDINGAN METODE REGRESI POLINOMIAL LOKAL DENGAN METODE NADARAYA-WATSON

Berikut adalah tabel yang memperlihatkan secara ringkas perbedaan taksiran regresi dengan metode polinomial lokal dan metode Nadaraya-Watson :

Tabel 4.3 Perbandingan Metode Regresi Polinomial Lokal dengan Metode Nadaraya-Watson

No		Nadaraya-Watson	Regresi Polinomial Local
1	Bentuk fungsi penaksir	$\hat{m}(x)=\dfrac{\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)Y_i}{\sum_{k=1}^{n}K\left(\frac{x-X_k}{h}\right)}$	$m(x,x^*)=a_0(x^*)+\sum_{j=1}^{p}a_j(x^*)(x-x^*)^j$ dan $m(x,x^*)$ ditaksir dengan $\hat{m}(x^*,x^*)=\hat{a}_0(x^*)$
2	Variable prediktor yang menaksir Y	x langsung	Menggunakan x^* yang berada dekat x
3	Langkah penaksiran	$\hat{m}(x)\xrightarrow{\text{penaksir}}Y$	$\hat{m}(x^*,x^*)\xrightarrow{\text{penaksir}}m(x,x^*)\xrightarrow{\text{penaksir}}Y$
4	Bentuk kurva	Mengikuti sebaran data	Mengikuti sebaran data

Pada Contoh Aplikasi

Data		Nadaraya-Watson	Regresi Polinomial Lokal
Test31	*Sum Square Error*	0.74867	0.773546
	Bentuk Kurva	Baik pada data yang mengumpul	Baik di ujung data, selebihnya tidak jauh berbeda
Data DBD 2008	*Sum Square Error*	28500.75	29983.52
	Bentuk Kurva	Baik pada data yang mengumpul	Baik di ujung data, selebihnya tidak jauh berbeda

Berdasarkan dua contoh ini, metode Nadaraya-Watson lebih baik dibandingkan regresi polinomial lokal. Hal ini disebabkan oleh metode

Nadaraya-Watson menggunakan variabel prediktor (x) langsung untuk menaksir Y . Sedangkan metode regresi polinomial lokal menggunakan $x*$ yang dekat dengan variabel prediktor (x) untuk menaksir Y . Serta, pada metode Nadaraya-Watson variabel respon Y ditaksir oleh $\hat{m}(x)$, sedangakan pada metode regresi polinomial lokal variabel respon Y ditaksir oleh $m(x, x*)$ dan $m(x, x*)$ ditaksir lagi oleh $\hat{m}(x*, x*)$. Namun jika interval $x*$ dibuat sekecil mungkin sedemikian sehingga $x*$ tepat berada pada setiap x yang akan ditaksir nilai Y nya, maka taksiran yang dihasilkan oleh metode regresi polinomial lokal akan lebih baik dari taksiran yang dihasilkan oleh metode Nadaraya-Watson. Dan metode regresi polinomial tetap lebih baik daripada metode Nadaraya-Watson pada ujung-ujung data seperti ditunjukkan pada Tabel 4.1 dan 4.2.

BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

1. Data yang tidak berpola dan menyebar tidak beraturan, tetap dapat dicari kurva regresinya, salah satu caranya dengan metode regresi polinomial lokal.
2. Regresi polinomial lokal adalah suatu metode regresi nonparametrik, dimana fungsi regresi ditaksir menggunakan bentuk polinomial. Jika pada regresi polinomial biasa persamaan regresi di-*fit* untuk seluruh wilayah data maka dalam regresi polinomial lokal persamaan regresi di-*fit* sepotong-sepotong.
3. Metode regresi polinomial lokal baik untuk menaksir data yang nilainya menyimpang jauh dibandingkan nilai data yang lain.
4. Untuk mendapatkan taksiran kurva yang baik diperlukan lebar *bandwidth* yang optimal, yaitu *bandwidth* yang meminimumkan GCV.
5. Dalam aplikasi yang telah dibahas dalam skripsi ini , metode Nadaraya-Watson dan metode regresi polinomial lokal menghasilkan kurva taksiran regresi yang tidak jauh berbeda, hanya saja metode

regresi polinomial lokal baik untuk menaksir pengamatan yang nilainya menyimpang jauh dibandingkan nilai pengamatan yang lain.

5.2 SARAN

Dalam aplikasi metode regresi polinomial sebaiknya digunakan untuk data yang menyebar secara acak dan pada data yang nilai ujungnya menyimpang jauh dibandingkan nilai pengamatan yang lain. Untuk mendapatkan taksiran yang lebih teliti, penulis menyarankan untuk membuat jarak antar x^* sekecil mungkin.

Pada tugas akhir ini penulis menyadari keterbatasan metode yang dibahas hanya untuk polinomial orde 2 dan menggunakan fungsi *kernel Gaussian* saja, padahal pemilihan fungsi Kernel juga penting dalam mencari taksiran kurva regresi. Untuk itu metode ini dapat dikembangkan untuk polinomial orde 3 dan pemilihan fungsi *kernel*.

DAFTAR PUSTAKA

Badan Pusat Statistik Kota Depok . *Kota Depok Dalam Angka Tahun 2008*.

Dinas Kesehatan Kota Depok . *Data rekapitulasi kasus DBD di Kota Depok menurut kelurahan tahun 2008*.

Edwin J Purcell, Dale Varberg, Steven E. Rigdon. 2003. *Kalkulus*. Prentice Hall, Inc.

Hardle, Wolfgang. 1990. *Smoothing Techniques With Implementation in S.* New York: Springer Verlag.

Montgomery, Douglas. 1991. *Introduction To Linear Regression Analysis*. New York: John Wiley & Sons, Inc.

Takezawa, Konio. 2006. *Introduction to Nonparametric Regression.* New Jersey: John Wiley & Sons, Inc.

LAMPIRAN 1

Pembentukan Matriks Hat

Akan dibuktikan :

$$\hat{\mathbf{Y}} = \mathbf{HY} \tag{2.13}$$

dengan $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ matriks yang berukuran n x n.

Bukti :

Pada model regresi linier berganda dengan k variabel prediktor :

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \qquad\qquad i = 1,2,\cdots,n \end{aligned} \tag{1}$$

Akan dicari taksiran dari parameter-parameter regresi $(\beta_0, \beta_1, \beta_2, \cdots, \beta_k)$ dengan menggunakan metode *least square*. Bentuk fungsi *least square :*

$$\begin{aligned} S(\beta_0, \beta_1, \cdots, \beta_k) &= \sum_{i=1}^{n} \varepsilon_i^{\,2} \\ &= \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \end{aligned} \tag{2}$$

Fungsi S diminimumkan , dan estimator *least square* $(\beta_0, \beta_1, \beta_2, \cdots, \beta_k)$ harus memenuhi :

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k} = -2\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) = 0 \tag{3}$$

dan

$$\frac{\partial S}{\partial \beta_j}\bigg|_{\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_k} = -2\sum_{i=1}^{n}\left(y_i - \hat\beta_0 - \sum_{j=1}^{k}\hat\beta_j x_{ij}\right)x_{ij} = 0 \qquad j = 1,2,\cdots,k \quad (4)$$

dari persamaan (3) dan (4), diperoleh persamaan :

$$n\hat\beta_0 + \hat\beta_1\sum_{i=1}^{n}x_{i1} + \hat\beta_2\sum_{i=1}^{n}x_{i2} + \cdots + \hat\beta_k\sum_{i=1}^{n}x_{ik} = \sum_{i=1}^{n}y_i$$

$$\hat\beta_0\sum_{i=1}^{n}x_{i1} + \hat\beta_1\sum_{i=1}^{n}x_{i1}^{\ 2} + \hat\beta_2\sum_{i=1}^{n}x_{i1}x_{i2} + \cdots + \hat\beta_k\sum_{i=1}^{n}x_{i1}x_{ik} = \sum_{i=1}^{n}x_{i1}y_i$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\hat\beta_0\sum_{i=1}^{n}x_{ik} + \hat\beta_1\sum_{i=1}^{n}x_{ik}x_{i1} + \hat\beta_2\sum_{i=1}^{n}x_{ik}x_{i2} + \cdots + \hat\beta_k\sum_{i=1}^{n}x_{ik}^{\ 2} = \sum_{i=1}^{n}x_{ik}y_i \qquad (5)$$

Maka didapat persamaan normal sebanyak $p = k+1$, dengan parameter regresi yang tidak diketahui sebanyak $p = k+1$ juga.

Agar lebih sederhana persamaan ini dapat dinyatakan dalam bentuk matriks. Model dalam persamaan (1) dapat ditulis dengan notasi matriks sebagai :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

dengan $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, vektor dari observasi yang berukuran n x 1,

$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$, matriks dari variabel prediktor berukuran n x p ,

$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$, vektor koefisien regresi yang berukuran p x 1, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, vektor dari random error yang berukuran n x 1.

Maka vektor *least square* :

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^{\,2} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \qquad (6)$$

karena $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ adalah matriks 1 x 1 atau skalar, maka transpos $(\boldsymbol{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ adalah skalar yang sama. Maka estimator *laest square* harus memenuhi :

$$\frac{\partial S}{\partial \beta}\Big|_{\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'y \qquad (7)$$

asalkan $(X'X)^{-1}$ ada. $(X'X)^{-1}$ akan selalu ada jika variabel prediktor-prediktornya saling bebas linier, yaitu, jika tidak ada kolom dari matriks **X** yang merupakankombinasi linier dari kolom lainnya.

Maka persamaan (7) bila ditulis dalam bentuk matriks :

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^{\ 2} & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik}x_{i1} & \sum_{i=1}^{n} x_{ik}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik}^{\ 2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}y_i \end{bmatrix}$$

Perkalian matriks ini identik dengan persamaan (5).

Model regresi yang cocok yang bersesuaian dengan variabel prediktor $x' = [1, x_1, x_2, \cdots, x_k]$ adalah :

$$\hat{y} = x'\hat{\beta}$$

$$= \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_j$$

Vektor dari $\hat{y}_i$ yang bersesuaian dengan nilai yang diobservasi y_i adalah :

$$\hat{y} = X\hat{\beta}$$

$$= X(X'X)^{-1}X'y$$

$$= Hy \qquad (8)$$

dengan $H = X(X'X)^{-1}X'$ matriks nxn yang biasa disebut matriks Hat, karena memetakan vektor dari nilai yang diobservasi kedalam vektor nilai yang di-*fit*.

LAMPIRAN 2

Elemen Matriks Hat dalam Regresi Polinomial Lokal

Sebelumnya akan dibuktikan :

$$[\mathbf{H}]_{kj} = q_j(X_k)$$

Bukti :

$$m(x, x^*) = a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(x - x^*)^j$$

dengan $\mathbf{a} = \begin{bmatrix} a_0(x^*) \\ \vdots \\ a_p(x^*) \end{bmatrix} = [\mathbf{X}^t\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y}$

ketika x = x*, maka $\hat{m}(x, x^*) = \hat{m}(x^*, x^*) = a_0(x^*)$

$$= \mathbf{e}_1{}^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y}$$

dengan $\mathbf{e}_1$ adalah vektor kolom (1, 0, 0, …, 0)^t yang berukuran (p+1) x 1.

Selanjutnya didefinisikan $\mathbf{q}(x^*)^t = \mathbf{e}_1{}^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}$ (dimana $q(x^*)^t = [q_1(x^*) \quad q_2(x^*) \quad \cdots \quad q_n(x^*)]$ matriks berukuran 1 x n)

Maka, $\hat{m}(x^*, x^*) = q(x^*)^t\, y$

$$= [q_1(x^*) \quad q_2(x^*) \quad \cdots \quad q_n(x^*)] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= [q_1(x^*)Y_1 + q_2(x^*)Y_2 + \cdots + q_n(x^*)Y_n]$$

$$= \left[\sum_{j=1}^{n} q_j(x^*) Y_j \right]$$

$$\hat{m}(X_k, X_k) = \hat{Y}_k = \sum_{j=1}^{n} q_j(X_k) Y_j$$

$$= q(X_k)^t y$$

Dalam regresi secara umum : $\hat{y} = Hy$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1n} \\ H_{21} & H_{22} & \cdots & H_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n1} & H_{n2} & \cdots & H_{nn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

jika disamakan $\hat{Y}_k = q(X_k)^t y$

maka terbukti $q(X_k)^t$ adalah elemen baris ke-k pada matriks Hat, dan $q_j(X_k)$ adalah elemen baris ke-k dan kolom ke-j pada matriks Hat.

Selanjutnya akan dicari elemen baris ke-k dan kolom ke-j dari matriks hat.

$$[\mathbf{H}]_{kj} = q_j(X_k)$$

$$= [\mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}]_j$$

$$= [(\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}]_{1j}$$

$$= \sum_{i=1}^{n} [(\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}]_{1i} [\mathbf{W}]_{ij}$$

$$= [(\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t]_{1j} [\mathbf{W}]_{jj}$$

$$= [(\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t]_{1j} \cdot w\left(\frac{X_j - X_k}{h}\right)$$

$$= \mathbf{e}_1{}^t (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{e}_j{}' \cdot w\left(\frac{X_j - X_k}{h}\right)$$

Dengan $\mathbf{e}_j{}'$ adalah vektor kolom dengan n elemen dan hanya elemen ke-j yang bernilai 1 yang lainnya bernilai 0.

LAMPIRAN 3

Elemen Diagonal Matriks Hat

Akan dicari elemen diagonal dari matriks Hat.

Pada lampiran 5 telah dibuktikan bahwa $q_j(X_k)$ adalah elemen baris ke-k dan kolom ke-j pada matriks Hat, maka $q_k(X_k)$ adalah elemen diagonal dari matriks Hat :

$$
\begin{aligned}
[\mathbf{H}]_{kk} &= q_k(X_k) \\
&= \mathbf{e}_1^{t}(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{e}'_k . w(0) \\
&= [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t]_{1k} . w(0) \\
&= \sum_{i=1}^{p+1}[(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}]_{1i}[\mathbf{X}^t]_{ik} . w(0) \\
&= \sum_{i=1}^{p+1}[(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}]_{1i}[\mathbf{e}'_i]_i . w(0) \\
&= [(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t]_{11} . w(0)
\end{aligned}
$$

LAMPIRAN 4

Estimator Nadaraya-Watson

Cara lain untuk memperoleh estimator *Nadaraya-Watson* ($\hat{m}(x)$), yaitu dengan meminimumkan kuadrat kesalahan (error square) yang diberi bobot :

$$E_{constant}(x) = \sum_{i=1}^{n} (Y_i - a_0)^2 w\left(\frac{x - X_i}{h}\right) \tag{2.15}$$

$w\left(\frac{x - X_i}{h}\right)$ adalah pembobotnya.

Akan diminimumkan persamaan (2.15), dengan cara menurunkan persamaan (2.15) terhadap a_0, $\frac{\partial E_{constant}(x)}{\partial a_0} = 0$. Kemudian dicari solusi untuk persamaan tersebut dengan langkah sebagai berikut :

$$E_{constant}(x) = \sum_{i=1}^{n} (Y_i - a_0)^2 w\left(\frac{x - X_i}{h}\right)$$

$$= \sum_{i=1}^{n} (Y_i^2 - 2Y_i a_0 + a_0^2) w\left(\frac{x - X_i}{h}\right)$$

$$= \sum_{i=1}^{n} Y_i^2 w\left(\frac{x - X_i}{h}\right) - 2a_0 \sum_{i=1}^{n} Y_i w\left(\frac{x - X_i}{h}\right) + a_0^2 \sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right)$$

$$\frac{\partial E_{constant}(x)}{\partial a_0} = 0$$

$$\frac{\partial E_{constant}(x)}{\partial a_0} = \frac{\partial\left(\sum_{i=1}^{n} Y_i^{\,2} w\left(\frac{x-X_i}{h}\right) - 2a_0 \sum_{i=1}^{n} Y_i w\left(\frac{x-X_i}{h}\right) + a_0^{\,2} \sum_{i=1}^{n} w\left(\frac{x-X_i}{h}\right)\right)}{\partial a_0}$$

$$-2\sum_{i=1}^{n} Y_i w\left(\frac{x-X_i}{h}\right) + 2a_0 \sum_{i=1}^{n} w\left(\frac{x-X_i}{h}\right) = 0$$

$$2a_0 \sum_{i=1}^{n} w\left(\frac{x-X_i}{h}\right) = 2\sum_{i=1}^{n} Y_i w\left(\frac{x-X_i}{h}\right)$$

$$a_0 \sum_{i=1}^{n} w\left(\frac{x-X_i}{h}\right) = \sum_{i=1}^{n} Y_i w\left(\frac{x-X_i}{h}\right)$$

$$a_0 = \frac{\sum_{i=1}^{n} Y_i w\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} w\left(\frac{x-X_i}{h}\right)}$$ *Terbukti.*

LAMPIRAN 5

Pembuktian w akan Bernilai Maksimum ketika $X_i - x^* = 0$

- Pada fungsi kernel gaussian :

$$w\left(\frac{X_i - x^*}{h}\right) = \exp\left(-\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)$$

Akan dibuktikan nilai w akan maksimum ketika $X_i - x^* = 0$

Bukti :

$$\exp\left(-\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right) = \frac{1}{\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)}$$

$\dfrac{1}{\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)}$ akan bernilai maksimum jika $\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right) = 1$

$$\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right) = e^0$$

$$\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2 = 0$$

$$\left(\frac{X_i - x^*}{h}\right)^2 = 0$$

$$\frac{X_i - x^*}{h} = 0$$

$$X_i - x^* = 0$$

Terbukti bahwa nilai w maksimum ketika $X_i - x^* = 0$

- Pada fungsi pembobot *bisquare* atau *biweight* :

$$w\left(\frac{X_i - x^*}{h}\right) = \begin{cases} \left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2 & jika\left(\frac{X_i - x^*}{h}\right)^2 \leq 1 \\ 0 & jika\left(\frac{X_i - x^*}{h}\right)^2 > 1 \end{cases}$$

dari bentuk tersebut terlihat bahwa w akan bernilai 0 atau $\left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2$, dan w akan bernilai maksimum ketika nilai $\left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2$ maksimum.

Akan dibuktikan $\left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2$ maksimum ketika $X_i - x^* = 0$.

Bukti :

w bernilai $\left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2$ jika $\left(\frac{X_i - x^*}{h}\right)^2 \leq 1$

Artinya :

$$\left(\frac{X_i - x^*}{h}\right)^2 \leq 1$$

$$\left(\frac{X_i - x^*}{h}\right)^2 - 1 \leq 0$$ bentuk kuadrat sempurna

$$\left(\left(\frac{X_i - x^*}{h}\right) - 1\right)\left(\left(\frac{X_i - x^*}{h}\right) + 1\right) \le 0$$

Maka w akan bernilai $\left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2$ jika $-1 \le \left(\frac{X_i - x^*}{h}\right) \le 1$

sehingga $0 \le \left(\frac{X_i - x^*}{h}\right)^2 \le 1.$ Dari sini terlihat bahwa $\left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2$ bernilai maksimum ketika $\left(\frac{X_i - x^*}{h}\right)^2 = 0$

$$\left(\frac{X_i - x^*}{h}\right)^2 = 0$$

$$\frac{X_i - x^*}{h} = 0$$

$$X_i - x^* = 0$$

Terbukti bahwa nilai w maksimum ketika $X_i - x^* = 0$

- Pada fungsi pembobot *triweight* dan fungsi pembobot *tricube* dapat dibuktikan dengan cara yang sama dengan fungsi pembobot *bisquare.*

LAMPIRAN 6

Pembuktian w akan Menurun ketika $|X_i - x^*|$ Membesar

- Pada fungsi kernel gaussian :

$$w\left(\frac{X_i - x^*}{h}\right) = \exp\left(-\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)$$

Akan dibuktikan nilai w akan menurun ketika $|X_i - x^*|$ menbesar

Bukti :

$$\exp\left(-\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right) = \frac{1}{\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)}$$

$\dfrac{1}{\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)}$ akan menurun ketika nilai $\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)$ semakin besar, nilai $\exp\left(\frac{1}{2}\left(\frac{X_i - x^*}{h}\right)^2\right)$ akan membesar ketika $\left(\frac{X_i - x^*}{h}\right)^2$ membesar, dan $\left(\frac{X_i - x^*}{h}\right)^2$ akan membesar ketika $|X_i - x^*|$ membesar. Maka terbukti w akan menurun ketika $|X_i - x^*|$ membesar.

- Pada fungsi pembobot *bisquare* atau *biweight* :

$$w\left(\frac{X_i - x^*}{h}\right) = \begin{cases} \left(1 - \left(\frac{X_i - x^*}{h}\right)^2\right)^2 & jika\left(\frac{X_i - x^*}{h}\right)^2 \leq 1 \\ 0 & jika\left(\frac{X_i - x^*}{h}\right)^2 > 1 \end{cases}$$

dari bentuk tersebut terlihat bahwa w akan bernilai 0 atau $\left(1-\left(\frac{X_i - x^*}{h}\right)^2\right)^2$.

Pada saat w bernilai 0, besar nilai w konstan (nilai w tidak akan terpengaruh dengan perubahan $|X_i - x^*|$, selama $\left(\frac{X_i - x^*}{h}\right)^2 > 1$)

Sedangkan saat w bernilai $\left(1-\left(\frac{X_i - x^*}{h}\right)^2\right)^2$, nilai w akan terpengaruh dengan perubahan $|X_i - x^*|$. Akan dibuktikan nilai w akan menurun ketika $|X_i - x^*|$ membesar.

Bukti :

w bernilai $\left(1-\left(\frac{X_i - x^*}{h}\right)^2\right)^2$ jika $\left(\frac{X_i - x^*}{h}\right)^2 \leq 1$

Artinya, w bernilai $\left(1-\left(\frac{X_i - x^*}{h}\right)^2\right)^2$ jika $-1 \leq \left(\frac{X_i - x^*}{h}\right) \leq 1$

Dari bentuk $\left(1-\left(\frac{X_i - x^*}{h}\right)^2\right)^2$ terlihat bahwa nilai $\left(1-\left(\frac{X_i - x^*}{h}\right)^2\right)^2$ akan menurun jika $\left(\frac{X_i - x^*}{h}\right)^2$ semakin besar, $\left(\frac{X_i - x^*}{h}\right)^2$ membesar ketika $|X_i - x^*|$ membesar. Maka terbukti w akan menurun ketika $|X_i - x^*|$ membesar.

- Pada fungsi pembobot *triweight* dan fungsi pembobot *tricube* dapat dibuktikan dengan cara yang sama dengan fungsi pembobot *bisquare.*

LAMPIRAN 7

Pembuktian Persamaan $E_{lokal}(x^*)$ dalam Bentuk Matriks

Akan dibuktikan

$$E_{lokal}(x^*) = (\mathbf{Xa} - \mathbf{y})^t \mathbf{W}(\mathbf{Xa} - \mathbf{y})$$

$$E_{lokal}(x^*) = \sum_{i=1}^{n} \left(w\left(\frac{X_i - x^*}{h}\right) \left(a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(X_i - x^*)^j - Y_i \right)^2 \right)$$

$$(\mathbf{Xa} - \mathbf{y})^t \mathbf{W}(\mathbf{Xa} - \mathbf{y}) = \sum_{i=1}^{n} \left(w\left(\frac{X_i - x^*}{h}\right) \left(a_0(x^*) + \sum_{j=1}^{n} a_j(x^*)(X_i - x^*)^j - Y_i \right)^2 \right)$$

Bukti :

$$E_{lokal}(x^*) = (\mathbf{Xa} - \mathbf{y})^t \mathbf{W}(\mathbf{Xa} - \mathbf{y})$$

$$\mathbf{Xa} - \mathbf{y} = \begin{bmatrix} 1 & (X_1 - x^*) & (X_1 - x^*)^2 & \cdots & (X_1 - x^*)^p \\ 1 & (X_2 - x^*) & (X_2 - x^*)^2 & \cdots & (X_2 - x^*)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x^*) & (X_n - x^*)^2 & \cdots & (X_n - x^*)^p \end{bmatrix} \begin{bmatrix} a_0(x^*) \\ a_1(x^*) \\ a_2(x^*) \\ \vdots \\ a_p(x^*) \end{bmatrix} - \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \begin{bmatrix} a_0(x^*) + a_1(x^*)(X_1 - x^*) + a_2(x^*)(X_1 - x^*)^2 + \cdots + a_p(x^*)(X_1 - x^*)^p \\ a_0(x^*) + a_1(x^*)(X_2 - x^*) + a_2(x^*)(X_2 - x^*)^2 + \cdots + a_p(x^*)(X_2 - x^*)^p \\ \vdots \\ a_0(x^*) + a_1(x^*)(X_n - x^*) + a_2(x^*)(X_n - x^*)^2 + \cdots + a_p(x^*)(X_n - x^*)^p \end{bmatrix} - \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j \\ \sum_{j=0}^{p} a_j(x^*)(X_2 - x^*)^j \\ \vdots \\ \sum_{j=0}^{p} a_j(x^*)(X_n - x^*)^j \end{bmatrix} - \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \\ \sum_{j=0}^{p} a_j(x^*)(X_2 - x^*)^j - Y_2 \\ \vdots \\ \sum_{j=0}^{p} a_j(x^*)(X_n - x^*)^j - Y_n \end{bmatrix}$$ matriks berukuran n x 1

Maka $(\mathbf{Xa} - \mathbf{y})^t$ berukuran 1 x n :

$$(\mathbf{Xa} - \mathbf{y})^t = \begin{bmatrix} \sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 & \sum_{j=0}^{p} a_j(x^*)(X_2 - x^*)^j - Y_2 & \cdots & \sum_{j=0}^{p} a_j(x^*)(X_n - x^*)^j - Y_n \end{bmatrix}$$

Maka $(\mathbf{Xa} - \mathbf{y})^t \mathbf{W}$

$$= \begin{bmatrix} \sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 & \cdots & \sum_{j=0}^{p} a_j(x^*)(X_n - x^*)^j - Y_n \end{bmatrix} \begin{bmatrix} w\left(\frac{X_1 - x^*}{h}\right) & 0 & \cdots & 0 \\ 0 & w\left(\frac{X_2 - x^*}{h}\right) & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & w\left(\frac{X_n - x^*}{h}\right) \end{bmatrix}$$

$$= \left[\left(\sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \right) w\left(\frac{X_1 - x^*}{h}\right) \quad \cdots \quad \left(\sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \right) w\left(\frac{X_1 - x^*}{h}\right) \right]$$

Maka

$$E_{lokal}(x^*) = (\mathbf{Xa} - \mathbf{y})^t \mathbf{W} (\mathbf{Xa} - \mathbf{y}) :$$

$$= \left[\left(\sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \right) w\left(\frac{X_1 - x^*}{h} \right) \quad \cdots \quad \left(\sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \right) w\left(\frac{X_1 - x^*}{h} \right) \right]$$

$$\begin{bmatrix} \sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \\ \sum_{j=0}^{p} a_j(x^*)(X_2 - x^*)^j - Y_2 \\ \vdots \\ \sum_{j=0}^{p} a_j(x^*)(X_n - x^*)^j - Y_n \end{bmatrix}$$

$$= \left[\left(\sum_{j=0}^{p} a_j(x^*)(X_1 - x^*)^j - Y_1 \right)^2 w\left(\frac{X_1 - x^*}{h} \right) + \cdots + \left(\sum_{j=0}^{p} a_j(x^*)(X_n - x^*)^j - Y_n \right)^2 w\left(\frac{X_n - x^*}{h} \right) \right]$$

$$= \left[\sum_{i=1}^{n} \left(w\left(\frac{X_i - x^*}{h} \right) \left(\sum_{j=0}^{p} a_j(x^*)(X_i - x^*)^j - Y_i \right)^2 \right) \right]$$

$$= \left[\sum_{i=1}^{n} \left(w\left(\frac{X_i - x^*}{h} \right) \left(a_0(x^*) + \sum_{j=1}^{p} a_j(x^*)(X_i - x^*)^j - Y_i \right)^2 \right) \right]$$

terbukti.

LAMPIRAN 8

Variansi dari $\hat{m}(X_K)$

Akan dibuktikan

$$Var\left[\hat{m}(X_k)\right] = \sigma^2 \cdot \sum_{i=1}^{n} ([\mathbf{H}]_{ki})^2$$

Bukti :

$$E[(\hat{m}(X_k) - E[\hat{m}(X_k)])^2]$$

$$= E\left[\left(\sum_{i=1}^{n} q_i(X_k)(m(X_i)+\varepsilon_i) - \sum_{j=1}^{n} q_i(X_k)m(X_i)\right)\left(\sum_{i=1}^{n} q_j(X_k)(m(X_j)+\varepsilon_j) - \sum_{j=1}^{n} q_j(X_k)m(X_j)\right)\right]$$

$$= E\left[\left(\sum_{j=1}^{n}\sum q_i(X_k)q_j(X_k)\varepsilon_j\varepsilon_k\right)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} q_i(X_k)q_j(X_k)\delta_{ij}\sigma^2$$

$$= \sigma^2 \sum_{i=1}^{n} \left(q_i(X_k)\right)^2$$

$$= \sigma^2 \cdot \sum_{i=1}^{n} ([H]_{ki})^2$$

LAMPIRAN 9

Data Test31

i	X_i	Y_i
1	0.1	0.1151
2	0.2	0.19762
3	0.3	0.578587
4	0.4	0.249493
5	0.5	0.595004
6	0.6	0.576665
7	0.7	0.878052
8	0.8	1.119478
9	0.9	0.740204
10	1	1.056001
11	1.1	1.102345
12	1.2	0.771004
13	1.3	1.107321
14	1.4	0.55236
15	1.5	0.502116
16	1.6	0.659377
17	1.7	0.841634
18	1.8	0.55896
19	1.9	0.545494
20	2	0.264694
21	2.1	0.314588
22	2.2	-0.36701

23	2.3	-0.14781
24	2.4	-0.5318
25	2.5	-0.74024
26	2.6	-0.67046
27	2.7	-0.83693
28	2.8	-0.79768
29	2.9	-1.17407
30	3	-0.87411

LAMPIRAN 10

Data Kepadatan Penduduk dan Rekapitulasi Kasus DBD di Kota Depok per Kelurahan Tahun 2008

i	Kelurahan	Kepadatan Penduduk (X_i)	Banyaknya Kasus DBD Y_i
1	Limo	3,2	4
2	Gandul	7,28	2
3	Tapos	16	4
4	Jatimulya	20,92	1
5	Cipayung	22,35	13
6	Leuwinanggung	23	2
7	Bedahan	24	2
8	Curug	24	5
9	Pasir Putih	25	1
10	Cinangka	29	0
11	Serua	29	0
12	Duren Seribu	29	1
13	Harjamukti	32	20
14	Grogol	33,1	4
15	Tirtajaya	33,15	4
16	Pengasinan	34	2
17	Kalibaru	34,84	1
18	Pondok Cina	35	10
19	Kalimulya	35,09	6

20	Sawangan	37	7
21	Cimpaeun	37	2
22	Cilodong	37,16	16
23	Kukusan	39	9
24	Sawangan Baru	39	5
25	Meruyung	39,71	4
26	Pondok Petir	43	4
27	Bojongsari Baru	44	0
28	Kedaung	44	0
29	Pangkalanjati Baru	47,42	0
30	Krukut	49,12	3
31	Bojongsari	51	0
32	Duren Mekar	52	3
33	Cilangkap	55	22
34	Cisalak	57,7	17
35	Cipayung Jaya	60,23	0
36	Sukmajaya	61,9	51
37	Tanah Baru	63	63
38	Rangkapan Jaya Baru	65,44	8
39	Pangkalan Jati Lama	69,61	3
40	Depok	74,13	83
41	Sukamaju Baru	80	50
42	Mekarsari	81	54
43	Jatijajar	81	56
44	Pancoran Mas	81,33	115
45	Mampang	82,37	21
46	Sukatani	83	95
47	Curug	88	5
48	Pasir gunung Selatan	89	11

49	Ratu Jaya	90,97	17
50	Rangkapan Jaya	91,05	33
51	Sukamaju	101,84	117
52	Pondok Jaya	104,19	7
53	Beji Timur	105	21
54	Cinere	111,18	24
55	Cisalak Pasar	118	23
56	Pondok Terong	127,05	13
57	Kemiri Muka	161	26
58	Tugu	178	58
59	Beji	179	143
60	Depok Jaya	193,5	91
61	Bakti Jaya	223,32	75
62	Abadi Jaya	241,09	105
63	Mekar Jaya	260,39	226

LAMPIRAN 11

Residual dari $\hat{m}(X_i, X_i)$ dengan Metode Regresi Polinomial Lokal dan Nadaraya-Watson pada Data Test31

i	X_i	Y_i	$\hat{m}(X_i, X_i)$		Residual	
			Nadaraya-Watson	Regresi Polinomial Lokal (p = 2)	Nadaraya-Watson	Regresi Polinomial Lokal
1	0.1	0.1151	0.26318520	0,124217	0,148085	0,009117
2	0.2	0.19762	0.31751713	0,244089	0,119897	0,046469
3	0.3	0.578587	0.38292012	0,365024	-0,19567	-0,21357
4	0.4	0.249493	0.46246776	0,487554	0,212978	0,238064
5	0.5	0.595004	0.56093866	0,60823	-0,03406	0,01323
6	0.6	0.576665	0.67482406	0,720844	0,098164	0,144184
7	0.7	0.878052	0.78583556	0,817972	-0,09221	-0,06008
8	0.8	1.119478	0.87224668	0,892553	-0,24723	-0,22693
9	0.9	0.740204	0.92523336	0,939338	0,185033	0,199138
10	1	1.056001	0.94702077	0,956111	-0,10898	-0,09989
11	1.1	1.102345	0.93762445	0,944532	-0,16472	-0,15781
12	1.2	0.771004	0.89405811	0,91022	0,123058	0,13922
13	1.3	1.107321	0.82225698	0,861493	-0,28506	-0,24583
14	1.4	0.55236	0.74420023	0,806594	0,19184	0,254234
15	1.5	0.502116	0.68604124	0,750203	0,183921	0,248083
16	1.6	0.659377	0.65230340	0,691028	-0,00708	0,031648
17	1.7	0.841634	0.61779249	0,622097	-0,22384	-0,21953
18	1.8	0.55896	0.55118352	0,533968	-0,00778	-0,02499

19	1.9	0.545494	0.43880448	0,419258	-0,10669	-0,12623
20	2	0.264694	0.28453097	0,27615	0,019841	0,01146
21	2.1	0.314588	0.10086925	0,109309	-0,21372	-0,20528
22	2.2	-0.36701	-0.09462874	-0,0717	0,272381	0,295311
23	2.3	-0.14781	-0.28368387	-0,25518	-0,13587	-0,10737
24	2.4	-0.5318	-0.45260429	-0,42995	0,079196	0,10185
25	2.5	-0.74024	-0.59353411	-0,58708	0,146706	0,153156
26	2.6	-0.67046	-0.70567483	-0,72067	-0,03522	-0,05021
27	2.7	-0.83693	-0.79449773	-0,8276	0,042432	0,009326
28	2.8	-0.79768	-0.86413664	-0,90657	-0,06646	-0,10889
29	2.9	-1.17407	-0.91339296	-0,9564	0,260677	0,217671
30	3	-0.87411	-0.94115323	-0,97399	-0,06704	-0,09988

LAMPIRAN 12

Residual dari $\hat{m}(X_i, X_i)$ dengan Metode Regresi Polinomial Lokal dan Nadaraya-Watson pada Data DBD 2008

i	Xi	Yi	$\hat{m}(X_i, X_i)$		Residual	
			Nadaraya-Watson	Regresi Polinomial Lokal (p=2)	Nadaraya-Watson	Regresi Polinomial Lokal (p=2)
1	3.2	4	3.473327	4.526208	-0.52667	0.526208
2	7.28	2	3.628954	4.050214	1.628954	2.050214
3	16	4	4.078972	3.068864	0.078972	-0.93114
4	20.92	1	4.362816	2.708768	3.362816	1.708768
5	22.35	13	4.44671	2.656139	-8.55329	-10.3439
6	23	2	4.484687	2.64204	2.484687	0.64204
7	24	2	4.542634	2.633527	2.542634	0.633527
8	24	5	4.542634	2.633527	-0.45737	-2.36647
9	25	1	4.599683	2.64226	3.599683	1.64226
10	29	0	4.811924	2.879943	4.811924	2.879943
11	29	0	4.811924	2.879943	4.811924	2.879943
12	29	1	4.811924	2.879943	3.811924	1.879943
13	32	20	4.94963	3.31595	-15.0504	-16.6841
14	33.1	4	4.996896	3.541402	0.996896	-0.4586
15	33.15	4	4.999036	3.552555	0.999036	-0.44745
16	34	2	5.035593	3.754642	3.035593	1.754642
17	34.84	1	5.07276	3.978248	4.07276	2.978248
18	35	10	5.080045	4.023619	-4.91996	-5.97638

19	35.09	6	5.084183	4.049538	-0.91582	-1.95046
20	37	7	5.182005	4.669554	-1.818	-2.33045
21	37	2	5.182005	4.669554	3.182005	2.669554
22	37.16	16	5.19142	4.727762	-10.8086	-11.2722
23	39	9	5.322033	5.470796	-3.67797	-3.5292
24	39	5	5.322033	5.470796	0.322033	0.470796
25	39.71	4	5.387106	5.795043	1.387106	1.795043
26	43	4	5.867837	7.588766	1.867837	3.588766
27	44	0	6.097003	8.233196	6.097003	8.233196
28	44	0	6.097003	8.233196	6.097003	8.233196
29	47.42	0	7.326313	10.80197	7.326313	10.80197
30	49.12	3	8.263974	12.29041	5.263974	9.29041
31	51	0	9.605053	14.09616	9.605053	14.09616
32	52	3	10.45792	15.12225	7.45792	12.12225
33	55	22	13.61676	18.44949	-8.38324	-3.55051
34	57.7	17	17.21447	21.71128	0.21447	4.71128
35	60.23	0	21.19134	24.92426	21.19134	24.92426
36	61.9	51	24.10827	27.08781	-26.8917	-23.9122
37	63	63	26.14068	28.51645	-36.8593	-34.4836
38	65.44	8	30.88352	31.65095	22.88352	23.65095
39	69.61	3	39.04501	36.68693	36.04501	33.68693
40	74.13	83	46.0099	41.29492	-36.9901	-41.7051
41	80	50	49.59868	45.39192	-0.40132	-4.60808
42	81	54	49.62305	45.85053	-4.37695	-8.14947
43	81	56	49.62305	45.85053	-6.37695	-10.1495
44	81.33	115	49.60061	45.98618	-65.3994	-69.0138
45	82.37	21	49.44025	46.36279	28.44025	25.36279
46	83	95	49.28289	46.55344	-45.7171	-48.4466
47	88	25	46.93991	47.08721	21.93991	22.08721

48	89	11	46.33972	46.99314	35.33972	35.99314
49	90.97	17	45.13179	46.62435	28.13179	29.62435
50	91.05	33	45.0829	46.60439	12.0829	13.60439
51	101.84	117	39.436	40.97196	-77.564	-76.028
52	104.19	7	38.14451	39.19439	31.14451	32.19439
53	105	21	37.64891	38.5604	16.64891	17.5604
54	111.18	24	32.89614	33.6745	8.89614	9.6745
55	118	23	26.55712	28.88167	3.55712	5.88167
56	127.05	13	19.69667	24.19042	6.69667	11.19042
57	161	26	47.48962	47.80818	21.48962	21.80818
58	178	58	92.3495	83.80095	34.3495	25.80095
59	179	143	93.47799	85.27404	-49.522	-57.726
60	193.5	91	95.63611	97.89794	4.63611	6.89794
61	223.32	75	80.35957	75.3176	5.35957	0.3176
62	241.09	105	114.2572	115.6965	9.2572	10.6965
63	260.39	226	209.6146	224.7358	-16.3854	-1.2642