JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume 13 Nomor 02 Bulan September Tahun 2024 E-ISSN 2622 8688

DOI : 10.59672/emasains.v13i2.4048

ESTIMATOR TRUNCATED SPLINE DAN SIFAT LINIERNYA DALAM REGRESI NONPARAMETRIK MULTIVARIABEL

Ni Putu Ayu Mirah Mariati [a,*], I Wayan Sudiarsa[b], Gusti Ayu Made Arna Putri[c]

[a,c]Universitas Mahasaraswati Denpasar, Indonesia
[b]Universitas PGRI Mahadewa Indonesia, Denpasar, Indonesia
*email: ayumirahmariati@unmas.ac.id

Abstrak. Analisis regresi digunakan untuk menyelidiki pola hubungan antara variabel dependen dan variabel independen. Hal ini dapat dilakukan dengan dua pendekatan, yaitu pendekatan parametrik dan nonparametrik. Pendekatan parametrik mengasumsikan bentuk model mengikuti suatu pola tertentu. Namun, jika tidak ada informasi tentang bentuk fungsi regresi, maka pendekatan yang digunakan adalah pendekatan regresi nonparametrik. Ada beberapa pendekatan untuk menaksir kurva regresi nonparametrik, salah satunya adalah truncated spline. Keuntungan dari truncated spline adalah dapat menggambarkan perubahan pola perilaku fungsi pada subinterval tertentu. Estimator spline sangat bergantung pada titik knot dan sifat estimatornya adalah linier.

Kata Kunci: Spline, Nonparametric Regression.

PENDAHULUAN

Terdapat tiga pendekatan analisis regresi dalam menaksir kurva regresi, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik (Eubank, 1999). Pendekatan model regresi parametrik memiliki sifat-sifat yang baik dari perspektif Statistika Inferensial seperti sederhana, mudah diinterpretasikan, ekonomis, tidak bias, merupakan estimator linier, efisien, konsisten, dan Best Linear Unbiased Estimator (BLUE). Meskipun sangat reliabel dari perspektif inferensial, pendekatan ini mensyaratkan terpenuhinya asumsi-asumsi yang diperlukan. Tidak semua permasalahan pola hubungan dapat didekati dengan regresi parametrik karena tidak semua permasalahan memiliki informasi mengenai bentuk hubungan atau kurva regresi antara variabel respon dengan variabel prediktor. Jika dipaksakan dengan pendekatan regresi parametrik, maka akan memberikan simpulan yang menyesatkan. Oleh karena itu, terdapat pilihan untuk menggunakan pendekatan nonparametrik yang tidak mensyaratkan informasi mengenai pola hubungan antar variabel. Model regresi nonparametrik yang bentuk kurva regresinya diasumsikan tidak diketahui. Selain itu diasumsikan mulus dimana kurva berada pada ruang fungsi seperti ruang Hilbert, ruang Sobolev, ruang fungsi kontinu dan lain-lain (Hardle,1994).

Perbedaan antara parametrik dan nonparametrik adalah pada pendekatan parametrik data cenderung dipaksa mengikuti pola tertentu, sedangkan pada pendekatan nonparametrik data diberi kebebasan untuk mencari pola kurva regresinya sendiri sehingga sangat fleksibel dan objektif. Beberapa model regresi nonparametrik yang sering digunakan adalah Histogram, Kernel, kNN, Spline, Neural Network (NN), Local Polynominal, Orthogonal Series, Fourier Series, Wavelet, MARS, dan lain-lain (Wahba, 1985). Semua model memiliki kelebihan dan kekurangan serta motivasinya sendiri dalam memodelkan pola. Dari beberapa model regresi

JEms
Emasains
Jurnal Edukasi Matematika dan Sains
Volume 13 Nomor 02 Bulan September Tahun 2024
P-ISSN 2302-2124
E-ISSN 2622 8688
DOI : 10.59672/emasains.v13i2.4048

nonparametrik, yang paling populer adalah Spline. Penelitian secara teoritis juga banyak berkembang, diantaranya (Dette *et al.*, 2018)Spline merupakan potongan polinomial yang memiliki sifat tersegmentasi dan kontinu (terpotong). Regresi spline terpotong digunakan karena kelebihannya adalah model ini dapat menemukan estimasi datanya sendiri ke mana pun pola data bergerak (Hidayat, *et al*., 2019) Dengan titik-titik simpul ini, spline dapat memberikan fleksibilitas yang lebih besar daripada polinomial, sehingga memungkinkan untuk beradaptasi secara efektif dengan karakteristik lokal. Alasan lain untuk memilih regresi spline terpotong adalah karena bersifat objektif dan optimasinya menggunakan metode kuadrat terkecil sehingga secara matematis mudah, sederhana, dan baik dalam membantu inferensi statistik dalam artikelnya menyarankan untuk menggunakan regresi spline polinomial jika plot data tidak jelas, simpangan baku besar, dan untuk penyederhanaan. Pada regresi spline nonparametrik data cross section dan satu respon, bahwa jika nilai parameter penghalusan sangat kecil akan memberikan estimator kurva regresi yang sangat kasar. Sebaliknya, jika nilai parameter penghalusan sangat besar, akan dihasilkan estimator kurva regresi nonparametrik yang sangat halus (Gu, C, 2013). Oleh karena itu, dalam penduga spline untuk data cross section, perlu dipilih parameter penghalusan yang optimal agar diperoleh penduga yang paling tepat untuk data tersebut. Dalam penelitian ini, kurva regresi nonparametrik akan diestimasi menggunakan pendekatan spline terpotong dan sifat-sifat liniernya.

METODE PENELITIAN

Mendapatkan estimasi kurva regresi nonparametrik Spline nonparametrik multivariabel langkah-langkahnya adalah sebagai berikut:

a. Membentuk model dengan Nonparametrik Spline Multivariabel

$$y_i = \sum_{j=1}^{p} f_j\left(x_{ji}\right) + \varepsilon_i \; ; \; i = 1, 2, ..., n$$

b. Menghampiri Kurva Regresi Spline *Truncated* derajat m dengan r titik knot

$$f_j\left(x_{ji}\right) = \sum_{v=1}^{m} \beta_{vj} x_{ji}^{v} + \sum_{k=1}^{r} \beta_{j(m+k)} \left(x_{ji} - K_{jk}\right)_{+}^{m}$$

c. Membentuk optimasi *Least Square*

d. Menyajikan persamaan *Least Square* dalam bentuk matriks

$$\underset{\underset{\sim}{\beta} \in R^{p(m+r)}}{Min} \left\{ n^{-1} \left\| \underset{\sim}{y} - X \underset{\sim}{\beta} \right\| \right\}$$

e. Menyelesaikan optimasi (d)dengan derivatif parsial.

f. Mencari sifat linier estimatornya.

HASIL DAN PEMBAHASAN

JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume 13 Nomor 02 Bulan September Tahun 2024 E-ISSN 2622 8688
DOI : 10.59672/emasains.v13i2.4048

Estimator Spline Truncated

Dalam bagian ini dibahas tentang estimasi Spline multivariabel. Spline merupakan jumlahan dari fungsi polinomial dengan suatu fungsi (*truncated*).
Diberikan model regresi nonparametrik multivariabel:

$$y_i = \mu\left(x_{1i}, x_{2i}, ..., x_{pi}\right) + \varepsilon_i$$

Selanjutnya, kurva regresi f_j dihampiri dengan fungsi Spline multivariabel maka dapat ditulis menjadi:

$$\mu\left(x_{1i}, x_{2i}, ..., x_{pi}\right) = \sum_{j=1}^{p} f_j\left(x_{ji}\right) + \varepsilon_i; i = 1, 2, ..., n$$

dengan:

$$f_j\left(x_{ji}\right) = \sum_{v=1}^{m} \beta_{vj} x_{ji}^{v} + \sum_{k=1}^{r} \beta_{j(k+m)} \left(x_{ji} - K_{jk}\right)_{+}^{m}$$

dimana $j = 1, 2, ..., p$; $v = 1, 2, ..., m$; dan $k = 1, 2, ..., r$ titik-titik knot yang memperlihatkan perubahan pola perilaku dari fungsi tersebut pada sub-sub interval yang berbeda. Dengan demikian model regresi dapat ditulis menjadi:

$$\begin{aligned}\sum_{j=1}^{p} f_j\left(x_{ji}\right) &= \sum_{j=1}^{p}\left(\beta_{1j} x_{ji}^{1} + ... + \beta_{mj} x_{ji}^{m} + \beta_{j(1+m)}\left(x_{ji} - K_{j1}\right)_{+}^{m} + ... + \beta_{j(r+m)}\left(x_{ji} - K_{jr}\right)_{+}^{m}\right) \\ &= \left(\beta_{11} x_{1i}^{1} + ... + \beta_{m1} x_{1i}^{m} + \beta_{1(1+m)}\left(x_{1i} - K_{11}\right)_{+}^{m} + ... + \beta_{1(r+m)}\left(x_{1i} - K_{1r}\right)_{+}^{m}\right) + ... + \\ &\quad \left(\beta_{1p} x_{pi}^{1} + ... + \beta_{mp} x_{pi}^{m} + \beta_{p(1+m)}\left(x_{pi} - K_{p1}\right)_{+}^{m} + ... + \beta_{p(r+m)}\left(x_{pi} - K_{pr}\right)_{+}^{m}\right) \end{aligned} \quad (1)$$

Apabila persamaan (1) disajikan dalam bentuk matrik, maka didapat:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11}^{1} & \cdots & x_{11}^{m} & \left(x_{11} - K_{11}\right)_{+}^{m} & \cdots & \left(x_{11} - K_{1r}\right)_{+}^{m} \\ x_{12}^{1} & \cdots & x_{12}^{m} & \left(x_{12} - K_{11}\right)_{+}^{m} & \cdots & \left(x_{12} - K_{1r}\right)_{+}^{m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{1n}^{1} & \cdots & x_{1n}^{m} & \left(x_{1n} - K_{11}\right)_{+}^{m} & \cdots & \left(x_{1n} - K_{1r}\right)_{+}^{m} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{m1} \\ \beta_{1(1+m)} \\ \vdots \\ \beta_{1(r+m)} \end{bmatrix} + ... +$$

$$\begin{bmatrix} x_{p1}^{1} & \cdots & x_{p1}^{m} & \left(x_{p1} - K_{p1}\right)_{+}^{m} & \cdots & \left(x_{p1} - K_{pr}\right)_{+}^{m} \\ x_{p2}^{1} & \cdots & x_{p2}^{m} & \left(x_{p2} - K_{p1}\right)_{+}^{m} & \cdots & \left(x_{p2} - K_{pr}\right)_{+}^{m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{pn}^{1} & \cdots & x_{pn}^{m} & \left(x_{pn} - K_{p1}\right)_{+}^{m} & \cdots & \left(x_{pn} - K_{pr}\right)_{+}^{m} \end{bmatrix} \begin{bmatrix} \beta_{1p} \\ \vdots \\ \beta_{mp} \\ \beta_{p(1+m)} \\ \vdots \\ \beta_{p(r+m)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2)$$

Persamaan (2) dapat ditulis menjadi:

$$\underset{\sim}{y} = X\left(K\right)\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

dengan,

$$\underset{\sim}{y} = \left[y_1, ..., y_n\right]$$

JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume 13 Nomor 02 Bulan September Tahun 2024 E-ISSN 2622 8688
DOI : 10.59672/emasains.v13i2.4048

$K = \left[K_{11},...,K_{1r} \vdots \cdots \vdots K_{p1},...,K_{pr} \right]$

$X(K) = \left[A_1 \ \vdots \ \cdots \ \vdots \ A_p \right]$

$\underset{\sim}{\beta} = \left(\underset{\sim}{\beta}_1',...,\underset{\sim}{\beta}_p' \right)'$

$\underset{\sim}{\beta}_1 = \left(\beta_{11},...,\beta_{m1},\beta_{1(1+m)},...,\beta_{1(r+m)} \right)',...,\underset{\sim}{\beta}_p = \left(\beta_{1p},...,\beta_{mp},\beta_{p(1+m)},...,\beta_{p(r+m)} \right)'$

dan $\underset{\sim}{\varepsilon} = (\underset{\sim}{\varepsilon}_1,...,\underset{\sim}{\varepsilon}_n)'$

Estimator parameter $\underset{\sim}{\beta}$ didapat dari menyelesaikan optimasi:

$$\underset{\underset{\sim}{\beta} \varepsilon R^{p(m+r)}}{Min} \left\{ n^{-1} \left\| \underset{\sim}{y} - X \underset{\sim}{\beta} \right\|^2 \right\} = \left\{ \left(\underset{\sim}{y} - X(K) \underset{\sim}{\beta} \right)' \left(\underset{\sim}{y} - X(K) \underset{\sim}{\beta} \right) \right\}$$

Untuk menyelesaikan optimasi dengan menggunakan derivatif parsial, misalkan:

$$\square \left(\underset{\sim}{\beta} \right) = \left\{ \left(\underset{\sim}{y} - X(K) \underset{\sim}{\beta} \right)' \left(\underset{\sim}{y} - X(K) \underset{\sim}{\beta} \right) \right\} \quad (3)$$

Persamaan (3) diturunkan terhadap $\underset{\sim}{\beta}$:

$$\frac{\partial \square \left(\underset{\sim}{\beta} \right)}{\partial \underset{\sim}{\beta}} = -2X'(K) \underset{\sim}{y} \quad +2X'(K) X(K) \underset{\sim}{\beta} \quad (4)$$

Setelah persamaan (4) diturunkan hasilnya disamakan dengan 0, maka diperoleh persamaan:

$$0 = -2X'(K) \underset{\sim}{y} + 2X'(K) X(K) \underset{\sim}{\hat{\beta}} \quad (5)$$

Hasil yang diperoleh dari persamaaan (5) adalah

$X'(K) X(K) \underset{\sim}{\hat{\beta}} = X'(K) \underset{\sim}{y}$

Sehingga estimator $\underset{\sim}{\hat{\beta}}$ diberikan oleh:

$$\underset{\sim}{\hat{\beta}} = \left(X'(K) X(K) \right)^{-1} X'(K) \underset{\sim}{y}$$

dengan $\underset{\sim}{\hat{\beta}} = \left(\underset{\sim}{\hat{\beta}}_1',...,\underset{\sim}{\hat{\beta}}_p' \right)'$.

Estimator kurva regresi $\underset{\sim}{\hat{f}}(x)$ diperoleh dari:

$$\begin{aligned} \underset{\sim}{\hat{f}}(x) &= \left[\underset{\sim}{\hat{f}}(x_1)', \underset{\sim}{\hat{f}}(x_2)',..., \underset{\sim}{\hat{f}}(x_p)' \right]' \\ &= X(K) \underset{\sim}{\hat{\beta}} \\ &= \left(X'(K) X(K) \right)^{-1} X'(K) \underset{\sim}{y} \end{aligned}$$

Sehingga diperoleh,

$\underset{\sim}{\hat{f}}(x) = \mathrm{B}(K) \underset{\sim}{y}$.

JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume 13 Nomor 02 Bulan September Tahun 2024 E-ISSN 2622 8688
DOI : 10.59672/emasains.v13i2.4048

SIFAT LINIER ESTIMATOR

Model regresi nonparametrik spline dapat ditulis dalam bentuk matriks berikut. $\underset{\sim}{y} = X(K)\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ dengan $K = \left[K_{11},...,K_{1r} \vdots \cdots \vdots K_{p1},...,K_{pr}\right]$. Jika dituliskan $\underset{\sim}{f} = X(K)\underset{\sim}{\beta}$ dengan $X(K)$ merupakan matriks fungsi K, maka diperoleh: $\underset{\sim}{y} = \underset{\sim}{f} + \underset{\sim}{\varepsilon}$.

$$\begin{aligned}\underset{\sim}{\hat{f}} &= X(K)\underset{\sim}{\hat{\beta}} \\ &= \left(X'(K)X(K)\right)^{-1}X'(K)\underset{\sim}{y} \\ &= \mathrm{B}(K)\underset{\sim}{y}\end{aligned}$$

Berdasarkan persamaan di atas, terlihat bahwa estimator spline $\underset{\sim}{\hat{f}}$ merupakan estimator yang linier. Kelinieran ini dapat memberikan kemudahan bagi peneliti dalam membentuk statistik inferensi untuk pendekatan spline.

SIMPULAN DAN SARAN

Simpulan

Jika diketahui model regresi nonparametrik: $y_i = \mu\left(x_{1i}, x_{2i},..., x_{pi}\right) + \varepsilon_i$ kurva regresi f_j dihampiri dengan fungsi Spline multivariabel maka dapat ditulis menjadi: $\mu\left(x_{1i}, x_{2i},..., x_{pi}\right) = \sum_{j=1}^{p} f_j\left(x_{ji}\right) + \varepsilon_i; i = 1,2,...,n$ sehingga diperoleh, $\underset{\sim}{\hat{f}}(x) = \mathrm{B}(K)\underset{\sim}{y}$. Sifat estimator yang diperoleh adalah linier.

Saran

Dapat dikembangkan secara teoritis dalam mencari sifat estimator yang lainnya dan diaplikasikan pada kasus riil.

DAFTAR PUSTAKA

Dette, H., Möllenhoff, K., Volgushev, S., and Bretz, F. (2018). Equivalence of Regression Curves. *Journal of the American Statistical Association.* 113 (522), 711–729.

Draper, N. R and Smith, H. (1998). *Applied Regression Analysis* . USA: John Wiley&Sons.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Deker.

Fitriana, D.m Budiantara, I.N.m and Ratnasari, V. (2017). Semiparametric Spline Truncated Regression on Modelling AHH in Indonesia. *Proceedings The 3rd International Seminar on Science and Technology*, 2 , 26–31.

Fithriasari, K., Hariastuti, I., and Wening, K. S. (2020). Handling Imbalance Data in Classification Model with Nominal Predictors, *International Journal Of Computing Science And Applied Mathematics.* 6(1), 33-37.

Gu, C. (2013). *Smoothing Spline ANOVA Models*. New York: Springer

Hardle, W. K. (1994). *Applied Nonparametric Regression*. New York: Cambridge University Press.

JEms Emasains
Jurnal Edukasi Matematika dan Sains P-ISSN 2302-2124
Volume 13 Nomor 02 Bulan September Tahun 2024 E-ISSN 2622 8688
DOI : 10.59672/emasains.v13i2.4048

Hidayat, R., Budiantara, I.N., Otok, B.W., and Ratnasari, V. (2019). A reproducing kernel hilbert space approach and smoothing parameters selection in spline-kernel regression, *Journal of Theoretical and Applied Information Technology*, 97 (2), 465–475.

Liu, X., and Preve, D. (2016). Measure of location-based estimators in simple linear regression, *Journal of Statistical Computation and Simulation*. 86(9), 1771–1784.

Mozumder, S. I., Rutherford, M., and Lambert, P. (2017). Direct Likelihood Inference On The Cause-Specific Cumulative Incidence Function: A Flexible Parametric Regression Modelling Approach. *Statistics in Medicine*. 37, 1–16.

Takezawa, K. (2006). *Introduction to Nonparametric Regression* . USA: John Wiley&Sons.

Wahba, G. (1985). A Comparison of GCV and GML for Choosing The Smoothing Parameterin The Generalized Spline Smoothing Problem, *The Annals of Statistics*. 13 (4), 1378-1402.

Wahba, G. (1990). *Spline Models For Observasion Data*. Pensylvania: SIAM.

Wang, X., Shen, J., and Ruppert, D. (2011). On the asymptotics of penalized spline smoothing, *Electronic Journal of Statistics*. 5, 1–17.

Wang, Y. (2011). Smoothing Splines. Methods and Applications. New York,: Chapman & Hall CRC

Wahba, G. (1990). *Spline Models for Observational Data*. Pensylvania: SIAM.