Jurnal Barekeng
Vol. 7 No. 2 Hal 9 –12 (2013)

Barekeng
jurnal ilmu matematika dan terapan

REGRESI FUZZY

Fuzzy Regression

DORTEUS LODEWYIK RAHAKBAUW
Staf Jurusan Matematika FakultasMIPA UniversitasPattimura
Jl. Ir. M. Putuhena, Kampus Unpatti, Poka-Ambon
E-mail: lodewyik@gmail.com

ABSTRAK

Dalam statistik, Analisis regresi memiliki bahasan ketergantungan satu variabel dengan satu atau lebih variabel yang lain. Tujuan dari analisa regresi adalah untuk menaksir parameter berdasakan data empiris. Bentuk linier $y = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_n x_n$, Dimana y adalah variabel output, $x_1, x_2, \cdots, x_n$ adalah variabel input, dan $0, 1, \cdots, n$ adalah parameter, yang merupakan bentuk matematika berulang dalam analisa regresi. Suatu permasalahan regresi linier yang lebih luas dimana terdapat parameter fuzzy dan data fuzzy dapat diselesaikan dengan pendekatan *symmetric triangular fuzzy number* dimana titik-titiknya dibagi dan dicocokkan berdasarkan data kemudian diselesaikan dengan masalah pemrograman linier.

Keywords: *regresi fuzzy*, *regresi linier*, *symmetric triangular fuzzy number*

PENDAHULUAN

Istilah fuzzy pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965, kemudian berkembang menjadi logika fuzzyyang didasai teori himpunan fuzzy. Pada teori himpunan fuzzy, peranan derajat keanggotaan sebagai penentu keberadaan dalam suatu himpunan.

TINJAUAN PUSTAKA

ANALISIS REGRESI

Analisis regresi adalah bagian statistik, yang membahas ketergantungan atas satu variabel dengan satu atau lebih variabel yang lain. Ketergantungan biasanya diasumsikan untuk mempunyai suatu bentuk matematis tertentu dengan satu parameter atau lebih. Tujuan dari analisa regresi adalah untuk menaksir parameter berdasakan data empiris. Bentuk linier

$$y = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_n x_n \qquad (1)$$

Dimana y adalah variabel output, $x_1, x_2, \cdots, x_n$ adalah variabel input, dan $0, 1, \cdots, n$ adalah parameter, yang merupakan bentuk matematika berulang dalam analisa regresi. Problem analisa regresi dirumuskan dalam kaitannya dengan bentuk linear yang disebut regresi linear. Diberikan, suatu contoh regresi linear dengan satu variabel.

Selanjutnya diasumsikan bentuk,

$$y = \gamma_0 + \gamma_1 x \qquad (2)$$

merepresentasikan suatu *straight line*.

Diberikan himpunan data yang diamati $< a_1, b_1 >$ $< a_2, b_2 >, \cdots, < a_m, b_m >$ dari pasangan variabel $< x, y >$, harus diperoleh nilai γ_0 dan γ_1, untuk itu *error* total dari titik estimasi pada *straight line* dengan pengaruh korespondensi titik observasi adalah minimal.

Penggunaan metode regresi linear, berdasarkan *least square error*, *total error* dapat ditentukan dengan rumus,

$$\sum_{i=1}^{m} [b_i - (\gamma_0 + \gamma_1 a_1)]^2$$

Sedangkan nilai-nilai optimal dari γ_0 dan γ_1 diperoleh dengan rumus:

$$\gamma_1 = \frac{m \sum_{i=1}^{m} a_i b_i - \sum_{i=1}^{m} a_i \sum_{i=1}^{m} b_i}{m \sum_{i=1}^{m} a_i^2 - (\sum_{i=1}^{m} a_i)^2}$$

$$\gamma_0 = \frac{\sum_{i=1}^{m} b_i - \gamma_1 \sum_{i=1}^{m} a_i}{m}$$

yang bisa digunakan untuk menyelesaikan masalah optimasi.

HASIL DAN PEMBAHASAN

Terdapat dua motivasi untuk mengembangkan analisis regresi fuzzy. Motivasi yang pertama adalah realisasi yang sering tidak realistis untuk mengasumsikan bahwa suatu fungsi *crisp* dari (1), merepresentasikan hubungan antara variabel-variabel yang diberikan. Relasi fuzzy, meskipun kurang tepat namun secara intuitif cukup realistis. Motivasi yang kedua adalah berdasarkan data sebenarnya, yang mana dalam aplikasi-aplikasi Fuzzy secara tak terpisah. Dua motivasi ini akan menjelaskan analisis regresi fuzzy. Suatu massalah yang melibatkan parameter-parameter fuzzy dan data *crisp*, sementara yang lain melibatkan parameter-parameter crisp dan data fuzzy. Dalam penulisan ini, akan diperlihatkan dua jenis regresi fuzzy linear.

1. Regresi Linier Dengan Parameter-Parameter Fuzzy.

Dalam regresi fuzzy tipe ini, ketergantungan variabel output terhadap variabel input di tunjukkan dalam bentuk,

$$Y = C_1x_1 + C_2x_2 + \cdots + C_nx_n \tag{3}$$

dimana $C_1, C_2, \cdots, C_n$ adalah *fuzzy number*, dan $x_1, x_2, \cdots, x_n$ adalah variabel input bernilai riil; untuk setiap pasangan-n dari nilai variabel-variabel input, nilai dari variabel output didefinisikan oleh (3) adalah *fuzzy number* dari Y. Diberikan titik-titik dari himpunan data *crisp* $< a_1, b_1 >, < a_2, b_2 > \cdots, < a_n, b_n >$, tujuan dari masalah regresi ini adalah mendapatkan parameter-parameter fuzzy $C_1x_1 + C_2x_2 + \cdots + C_nx_n$, yang mana (3) menunjukkan kecocokan terbaik dari titik-titik data ini, sesuai dengan beberapa kriteria kebaikannya.

Diasumsikan bahwa parameter-parameter dalam (3) adalah *symmetric triangular* fuzzy *number* sebagai berikut:

$$C_i(c) = \begin{cases} 1 - \frac{|c-c_i|}{s_i} & untuk\ c_i - s_i \le c \le c_i + s_i \\ 0 & untuk\ yang\ lain \end{cases} \tag{4}$$

dimana c_i titik dimana $C_i(c_i) = 1$ dan $s_i > 0$ merupakan *spread* dari C_i (setengah dari panjang *support* himpunan C_i). Diberikan C_i, yang ditunjukkan dalam perkiraan kondisi *linguistic* c_i atau sekitar c_i, dinotasikan dengan $C_i =< c_i, s_i >$ untuk semua $i \in \mathbb{N}_n$. Maka, ini sangat mudah untuk membuktikan dengan *extension principle* bahwa Y dalam (3) adalah juga *symmetric triangular fuzzy number* diberikan oleh,

$$Y(y) = \begin{cases} 1 - \frac{|y-x^Tc|}{s^T|x|} & untuk\ x \ne 0 \\ 1 & untuk\ x = 0, y \ne 0 \\ 0 & untuk\ x = 0, y = 0 \end{cases} \tag{5}$$

untuk semua $y \in \mathbb{R}$, dimana

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}, |x| = \begin{bmatrix} |x_1| \\ |x_2| \\ \vdots \\ |x_n| \end{bmatrix},$$

dan T dinotasikan sebagai transpos.

Masalah sebenarnya dari pencarian parameter-parameter fuzzy $C_1, C_2, \cdots, C_n$ dapat dikonversi ke masalah untuk mencari vektor-vektor c dan s seperti $Y(y)$ (5) dengan mencocokkan data. Dua kriteria terbaik biasanya digunakan dalam masalah ini. Berdasarkan kriteria pertama, untuk setiap data yang diberikan $< a_j, b_j >$, dimana a_j adalah nilai-nilai vektor dari variabel-variabel input, b_j pasti bagian dari koresponding *fuzzy number* b_j dengan *grade* lebih besar atau sama dengan suatu nilai $h \in [0,1]$. Untuk $Y_j(b_j) \ge h$ untuk setiap $j \in \mathbb{N}_m$, dimana Y_j adalah *fuzzy number* didefinisikan dalam (5) untuk $x = a_j$. Berdasarkan kriteria ke dua, nonspesifik total dari parameter-parameter fuzzy haruslah diminimumkan/ diperkecil. Ini yang dibutuhkan untuk memperoleh ekspresi spesifik (3) untuk kecocokan yang cukup (dilihat dari nilai h). Kenonspesifikan dari tiap-tiap parameter *fuzzy* C_j diberikan oleh (4) mungkin ditunjukkan dengan *spread* s_i.

Untuk masalah regresi *fuzzy* dapat diformulasikan dengan kondisi masalah pemrograman linier dasar sebagai berikut .

minimumkan $$\sum_{i=1}^{m} s_i$$

berdasarkan

$$\begin{aligned} (1-h)s^T|a_j| - |b_j - a_j^T c| \ge 0, j \in \mathbb{N}_m \\ s_i \ge 0, i \in \mathbb{N}_M \end{aligned} \tag{6}$$

Contoh Regresi Linier Dengan Parameter-Parameter Fuzzy

Diberikan $\langle 1,1 \rangle, \langle 2,2 \rangle, \langle 3,2 \rangle, \langle 4,3 \rangle$ adalah data yang merepresentasi ketergantungan varibel y terhadap variabel x. Untuk mengilustrasikan regresi fuzzy linear dari data ini, diasumsikan bentuk,

$$Y = Cx$$

dimana $C = \langle c, s \rangle$ adalah parameter fuzzy ditunjukan sebagai anggota(*member*) *symmetric triangular* fuzzy. Maka, masalah pemrograman linear memiliki bentuk:

minimumkan s
berdasarkan

$$\begin{aligned} (1-h)s - |1-c| \ge 0 \\ 2(1-h)s - |2-2c| \ge 0 \\ 3(1-h)s - |2-3c| \ge 0 \\ 4(1-h)s - |3-4c| \ge 0 \end{aligned}$$

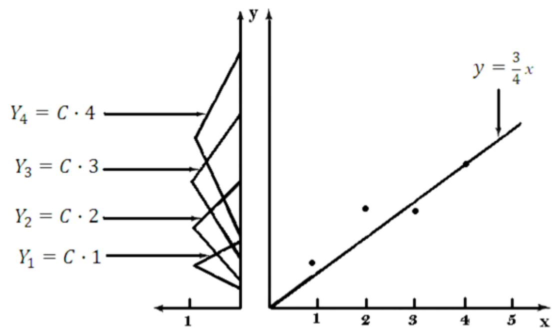
masalah ini dapat ditunjukkan dalam bentuk sederhana sperti yang diberikan di bawah ini.

minimumkan s
berdasarkan

$$s \ge \frac{1}{1-h} \max\left(|1-c|, \left|\tfrac{2}{3} - c\right|, \left|\tfrac{3}{4} - c\right|\right)$$
$$h \in [0,1] \text{ adalah } \textit{fixed number}$$

Menyelesaikan masalah ini, akan didapat nilai optimalnya adalah,

$$c^* = \frac{5}{6} \text{ dan } s^* = \frac{1}{6(1-h)}$$



Gambar 1: Ilustrasi dari contoh regresi linier dengan parameter Fuzzy

Sehingga,

$$C = \langle \frac{5}{6}, \frac{1}{6(1-h)} \rangle$$

Memilih $h = \frac{2}{3}$ diperoleh $C = \langle \frac{5}{6}, \frac{1}{2} \rangle$ himpunan fuzzy $Y_j = C_{a_j}$ untuk $a_1 = 1, a_2 = 2, a_3 = 3$ dan $a_4 = 4$ ditunjukkan dalam gambar 1. Juga yang ditunjukkan dalam gambar tersebut adalah *linear least-square fitting*.

2. Regresi Linier dengan data Fuzzy

Dalam regresi fuzzy tipe ini, ketergantungan variabel output terhadap variabel input di tunjukkan dalam bentuk,

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n \qquad (7)$$

dimana nilai-nilai variabel-variabel input dan output adalah *fuzzy number*, diasumsikan untuk *triangular* dan *symmetric*, dan $a_1, a_2, \cdots, a_n$ adalah parameter-parameter bernilai riil. Diberikan $X_i = \langle x_i, s_i \rangle$ untuk semua $i \in \mathbb{N}_n$. Maka

$$Y(y) = \begin{cases} 1 - \frac{|y - a^T x|}{s^T |a|} & untuk\ a \neq 0 \\ 1 & untuk\ a = 0, y \neq 0 \\ 0 & untuk\ a = 0, y = 0 \end{cases} \qquad (8)$$

untuk semua $y \in \mathbb{R}$, dimana

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix},$$

Data diberikan dari pasangan $\langle X^{(j)} Y^{(j)} \rangle$, dimana $X^{(j)}$ adalah n-tupel dari *symmetric triangular fuzzy number*, dan $Y^{(j)}$ adalah *symmetric triangular fuzzy number* untuk setiap $j \in \mathbb{N}_m$. Tujuan dari masalah regresi ini adalah memperoleh parameter-parameter $a_1, a_2, \cdots a_n$ sebagaimana fungsi linear fuzzy (7) mencocokkan data yang diberikan sebaik mungkin. Dua kriteria terbaik biasanya sering digunakan. berdasarkan kriteria pertama, perbedaan total antara area dari *actual fuzzy number* $Y^{(j)}$ dan area dari *fuzzy number* Y_j diperoleh untuk $X^{(j)}$ berdasarkan (7), dimana $j \in \mathbb{N}_m$, harusnya diminimumkan/diperkecil. berdasarkan kriteria ke dua, *fuzzy number* $Y^{(j)}$ dan Y_j seharusnya kompatibel paling tidak untuk beberapa derajat tertentu $h \in [0,1]$; kompatibel(*com*), didefinisikan,

$$\left(Y^{(j)}, Y_j\right) = \sup_{y \in \mathbb{R}} \min\left[Y^{(j)}(y), Y_j(y)\right]$$

Menggunakan dua kriteria ini, dijelaskan masalah regresi fuzzy dapat diformulasikan dengan kondisi dari masalah optimasi berikut:

Minimumkan

$$\sum_{j=1}^{m} \left| \int_{\mathbb{R}} Y^{(j)}(y)dy - \int_{\mathbb{R}} Y_j(y)dy \right| \qquad (9)$$

Berdasarkan

$$\min_{j \in \mathbb{N}_m} \mathrm{com}(Y^{(j)}, Y_j)] \geq h$$

diberikan $X_i^{(j)} = \langle x_i^{(j)}, s_i^{(j)} \rangle$ untuk semua $i \in \mathbb{N}_n$ dan $Y^{(j)} = \langle y^{(j)}, s^{(j)} \rangle$ maka, masalah regresi *fuzzy* dari tipe ini dapat juga diformulasikan dengan bentuk,

minimumkan

$$\sum_{j=1}^{m} \left| s^{(j)} - \sum_{i=1}^{n} |a_i| s_i^{(j)} \right|$$

berdasarkan

$$\begin{aligned} -\textstyle\sum_{i=1}^{n} |a_i| s_i^{(j)} + \sum_{i=1}^{n} a_i x_i^{(j)} &\leq y^{(j)} - s^{(j)} \\ \textstyle\sum_{i=1}^{n} |a_i| s_i^{(j)} + \sum_{i=1}^{n} a_i x_i^{(j)} &\geq y^{(j)} - s^{(j)} \end{aligned} \qquad (10)$$

$a_i \in \mathbb{R}$ untuk semua $i \in \mathbb{N}_n$ dan semua $j \in \mathbb{N}_m$

Contoh Regresi Linier Dengan DataFuzzy

Untuk mengilustrasikan pendeskripsian metode regresi linear dengan data fuzzy, selanjutnya diberikan bentuk linear simpel,

$$Y = aX$$

dan berikut data dalam kondisi atas pasangan input/output *fuzzy number*:

$$\langle\langle 5/6, 1/2\rangle, \langle 1,0\rangle\rangle$$
$$\langle\langle 5/3, 1/2\rangle, \langle 2,1/2\rangle\rangle$$
$$\langle\langle 5/2, 1/2\rangle, \langle 3,1/2\rangle\rangle$$
$$\langle\langle 10/3, 1/2\rangle, \langle 4,0\rangle\rangle$$

Menerapkan (10) ke dalam contoh ini diperoleh:

Minimumkan $|a| + |1 - a|$

Berdasarkan

$$-\frac{|a|}{2} + \frac{5a}{6} \leq 1,$$
$$\frac{|a|}{2} + \frac{5a}{6} \geq 1,$$
$$-\frac{|a|}{2} + \frac{5a}{3} \leq 1.5,$$

$$\frac{|a|}{2}+\frac{5a}{3}\geq 1.5,$$
$$-\frac{|a|}{2}+\frac{5a}{2}\leq 2.5,$$
$$\frac{|a|}{2}+\frac{5a}{2}\geq 2.5$$
$$-\frac{|a|}{2}+\frac{10a}{3}\leq 4$$
$$\frac{|a|}{2}+\frac{10a}{3}\geq 4$$
$$a\in\mathbb{R}$$

Formula ini dapat disederhanakan menjadi bentuk,

Minimumkan $|a|+|1-a|$

Berdasarkan

$$a\in[6/8\,,3],$$
$$a\in[9/13\,,9/7],$$
$$a\in[5/6\,,5/4],$$
$$a\in[24/23\,,24/17],$$

Hasil dari masalah ini, nilai optimal dari parameter a adalah $a^{*}=24/23$. Sedemikian sehingga bentuk dari $Y=aX$ dengan kecocokan terbaik (berdasarkan kriteria yang dipilih) adalah $Y=24x/23$.

KESIMPULAN

Setiap permasalahan Regresi linear yang memuat parameter-parameter fuzzy ataupun data fuzzy dapat diselesaikan dengan pendekatan *symmetric triangular fuzzy number*

DAFTAR PUSTAKA

Klir, J.G & Yuan, B., 1995, *Fuzzy Sets and Fuzzy Logic:Theory and Application.* Prentice Hall Inc, New Jersey.

Zimmerman., 1991, *Fuzzy Set Theory and its application*, second edition, Kluwer Academic Publisher, Massachusetts-USA.