Study of Fuzzy Linear Regression`s Properties and its Paramatre Estimation

Chen Gang

(Liao Ning Technical University,Fu xin,Liao ning,123000,P.R.China)

Abstract

In this paper ,we have studied the fuzzy Linear regression whose parameters are fuzzy numbers .By introducing the notions of the fitting measure and h-contaning between fuzzy numbers,we have proposed three methods to estimate fuzzy parameter .Finally,we proved some properties.

Keywords:fuzzy number,fuzzy regression,fuzzy parameter estimation

1. Fuzzy number and fuzzy Linear regression

Definition 1. 1 Suppose A is a fuzzy set on R.If A satisfies the following conditions:

(1). $\mu_{A(x)}$ is super continuous

(2). $\mu_{A(x)} = 1 \forall x \notin [c,d]$

(3). $\exists a,b$ such that $c \le a \le b \le d, \mu_{A(x)} = 1$

when $x \in [a,b], \mu_{A(x)}$ is increased when $x \in [c,d]$ and $\mu_{A(x)}$ is decreased when $x \in [b,d]$.Then A is called a fuzzy nember.

Theorem 1. 1 A fuzzy set A is a fuzzy number,only and only there is an interval [a,b] such that

$$\mu_{A(x)} = \begin{cases} 1, x \in [a,b] \\ L(x), x < a \\ R(x), a > b \end{cases}$$

Here $L(x), R(x)$ are increased and right-continuous and decreased and left-continuous respectively,there exists an interval $[c,d] \supseteq [a,b]$,when $x < c, L(x) = 0$,when $x > d, R(x) = 0$. $(0 \le R(x) \le 1, 0 \le L(x) \le 1)$

In this paper we adopt such fuzzy numbers as the following

$$\mu_{A(x)} = \begin{cases} 1 - \dfrac{|x-a|}{c}, |x-a| \le c \\ 0, otherwise \end{cases}$$

Specially,when $c = o, A = a$,denoted A as A=$L(a,c)$. For $\forall \lambda \in [0,1]$,the λ -lever set of fuzzy membership of fuzzy set $A * B$ is defined as

$$\mu_{A*B(x)} = \bigvee_{z*y}\left[\mu_{A(x)} \wedge \mu_{B(y)}\right]$$

Theorem 1.2 Suppose $A_1 = L(a_1, c_1), A_2 = L(a_2, c_2), * \in \{+, -, \times\}$, then $A_1 * A_2$ is a fuzzy number and

$$A_1 + A_2 = L(a_1 \pm a_2, |c_1 \pm c_2|)$$
$$\lambda A_1 = L(\lambda a_1, |\lambda| c_1), (\lambda \in R)$$

Definition 1.3 Suppose A_1, A_2, Λ, A_n are fuzzy numbers, x_1, x_2, Λ, x_n are independent random varibles,then the formula

$$Y = A_1 X_1 + A_2 X_2 + \Lambda + A_n X_n + E$$

is called fuzzy linear regression.

Here Y is a fuzzy set which is called fuzzy out-put, A_1, A_2, Λ, A_n are parameter ,E is a fuzzy set called fuzzy disturbed set.

Theorem1.3 Suppose $A_i = L(a_i, c_i), E = L(e, d)$

$$Y = A_1 X_1 + A_2 X_2 + \Lambda + A_n X_n + E$$

Then Y is a fuzzy number ,and

$$Y = L(\sum_{i=1}^{n} a_i X_i + e, \sum_{i=1}^{n} c_i |X_i| + d)$$

That is,the membership of Y is

$$\mu_{Y(x)} = \begin{cases} 1 - \dfrac{\left|x - \sum_{i=1}^{n} a_i X_i - e\right|}{\sum_{i=1}^{n} c_i |X_i| + d}, & \left|x - \sum_{i=1}^{n} a_1 X_i - e\right| \le \sum_{i=1}^{n} c_i |X_i| + d \\ 0, otherwise \end{cases}$$

2. Fuzzy parameter eatimation

Suppose $X_{i1}, X_{i2}, \Lambda, X_{in}, Y_i (i = 1, 2, \Lambda, N)$ are input-output data,which satisfy fuzzy Linear regression

$$Y_i = A_1 X_{i1} + A_2 X_{i2} + \Lambda + A_n X_{in} + E_i$$

In the following we propose three methods to estimate fuzzy parameters A_i according to these

$$M(A_1,A_2)=\underset{x\subset R}{\vee}\left[\mu_{A_1(x)}\wedge\mu_{A_2(x)}\right]$$

Then call M to be fitting measure of A_1 against A_2 according to Definition 2.1,If A_1,A_2 are two fuzzy numbers, $A_1=L(a_1,c_1),A_2=L(a_2,c_2)$,then

$$M(A_1,A_2)=\begin{cases}1-\dfrac{|a_1-a_2|}{c_1+c_2},|a_1-a_2|\le c_1+c_2\\ 0,otherwise\end{cases}$$

Definition2. 2Suppose A_1,A_2 are two fuzzy numberss,If for a certain $h\in[0,1]$, $[A_1]_h\supseteq[A_2]_h$, then call A_1 h includes A_2 denoted as $A_1\underset{h}{\supseteq}A_2$.

If $A_1=L(a_1,c_1),A_2=L(a_2,c_2)$ then $A_1\underset{h}{\supseteq}A_2$ is only and only the following inequalities are ture.

$$\begin{cases}a_1\le a_2+(1-h)(c_1-c_2)\\ a_1\ge a_2+(1-h)(c_1-c_2)\end{cases}$$

And if $A_1\underset{h}{\supseteq}A_2$,then $A_1\underset{h'}{\supseteq}A_2$.where $0\le h'\le h\le 1$

(1).parameter estimation (I)

Suppose $\hat{A}_i=L(\hat{a}_i,\hat{c}_i)$ are the estimations of $A_i,\hat{Y}_i=\hat{A}_1X_{i1}+\hat{A}_2X_{i2}+\Lambda+\hat{A}_nX_{in}$ are estimation of $Y_i=L(y_i,e_i)(i=1,2,\Lambda\ ,N)$

For the given level h advance.and each i, $\hat{A}_1,Y_i,\hat{Y}_i$ satisfy the follwing conditions

$$\begin{cases}M(\hat{Y}_i,Y_i)\ge h\\ \hat{J}(c,h)=\sum_{i=1}^{n}\omega_ic_i is \min mum\end{cases}$$

$$W_i=\sum_{k=1}^{n}|\omega_{ki}|$$

According to the measure of fuzzy numbers,we can reduce the above problem into the follwing Linear programming

$$\hat{J}(c,h)=Min\sum_{i=1}^{n}\omega_ic_i$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} a_j + (1-h) \sum_{j=1}^{n} |x_{ij}| c_j \geq y_j - (1-h) e_i \\ \sum_{j=1}^{n} x_{ij} a_j - (1-h) \sum_{j=1}^{n} |x_{ij}| c_j \leq y_j - (1-h) e_i \end{cases} \quad (1)$$

$$c_1, c_2, \Lambda\ .c_n > 0, (i = 1, 2, \Lambda\ , N)$$

(2).parameter estimation(II),(III)

Suppose $\overline{A}_i, \underline{A}_i$ are the estimations $A_i, \overline{Y}_i, \underline{Y}_i$ are the estimation of responding to $\overline{A}_i, \underline{A}_i$ .For the given level h in advance,and for each i $\overline{A}_i, \underline{A}_i, \overline{Y}_i, \underline{Y}_i$ satisfy the requirements as the follwing:

$$\begin{cases} Y_i \underset{h}{\subseteq} \overline{Y}_i = \overline{A}_1 X_1 + \overline{A}_2 X_2 + \Lambda + \overline{A}_n X_n \\ \overline{J}(c,h) = \sum_{k=1}^{n} \omega_k c_k is \min mum \end{cases}$$

or

$$\begin{cases} Y_i \underset{h}{\supseteq} \overline{Y}_i = \underset{h}{\underline{A}_1 X_1} + \underline{A}_2 X_2 + \Lambda + \underline{A}_n X_n \\ \underline{J}(c,h) = \sum_{k=1}^{n} \omega_k c_k is \min mum \end{cases}$$

According to h-containing definition2.2 and the h-level set a fuzzy number,We can reduce the above two problems into the folllwing Linear programmings respectively

$$\overline{J}(c,h) = Min \sum_{k=1}^{n} \omega_k c_k$$

$$\begin{cases} y_j - (1-h) e_i \leq \sum_{j=1}^{n} x_{ij} a_j + (1-h) \sum_{j=1}^{n} |x_{ij}| c_j \\ y_j - (1-h) e_i \geq \sum_{j=1}^{n} x_{ij} a_j - (1-h) \sum_{j=1}^{n} |x_{ij}| c_j \end{cases} \quad (2)$$

$$c_1, c_2, \Lambda\ .c_n > 0, (i = 1, 2, \Lambda\ , N)$$

or

$$\underline{J}(c,h) = Max \sum_{k=1}^{n} \omega_k c_k$$

$$\begin{cases} y_j - (1-h) e_i \geq \sum_{j=1}^{n} x_{ij} a_j + (1-h) \sum_{j=1}^{n} |x_{ij}| c_j \end{cases}$$

(3)

parameter A_i ,therefore we can get the preditions of

$$\begin{aligned}\hat{Y}_i &= \hat{A}_1X_{i1} + \hat{A}_2X_{i2} + \Lambda + \hat{A}_nX_{in} \\ \overline{Y}_i &= \overline{A}_iX_{i1} + \overline{A}_2X_{i2} + \Lambda + \overline{A}_nX_{in} \\ \underline{Y}_i &= \underline{A}_1X_{i1} + \underline{A}_2X_{i2} + \Lambda + \underline{A}_nX_{in}\end{aligned} \qquad (4)$$

3. Some properties

In paragraph two ,we have discussed parameter estimation problem,but now we must ask if there exists an optimal solution to the Linear programming(I),(II),(III)?If they have optimal solutions,what are the relations among $\hat{J}(c,h),\bar{J}(c,h)$ and $\underline{J}(c,h)$?and how do they vary with the given h in advance?

The folllwing theorems answer those questions basically.

Theorem 3.1 There are optimal solutions $\hat{A}_1 = L(\hat{a}_1,\hat{c}_1)$ and $\hat{A}_2 = L(\hat{a}_2,\hat{c}_2)$ and in (I),(II).But it is not assured that there exists an optimal solution in(II).

Theorem 3.2 If there is an optimal solution for h' in (III),there exists an optimal solution for $0 < h < h'$ in (III).

Theorem 3.3 For $h < h'$,we have

$$\hat{J}(h',c) \le \hat{J}(h,c), \underline{J}(h',c) \ge \underline{J}(h,c), \bar{J}(h',c) \le \bar{J}(h,c)$$

Theorem 3.4 For $0 < h < 1$, we have the follwing inequalities

$$\bar{J}(h,\bar{c}) \ge \hat{J}(h,c), \underline{J}(h,\underline{c}) \ge \hat{J}(h,\hat{c})$$

References

[1] Fuzzy Forecast of quantity of sold electricity systems Engineering-practice Vol.13.No.1(1993)48~56.

[2] Tanaka,H,Fuzzy data analysis by possibilitic Linear model Fuzzy sets and systems,Vol.24,No.3,(1987)363~375.