INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 12(4), 492-502, August 2017.

The Logistic Regression from the Viewpoint of the Factor Space Theory

Q.F. Cheng, T.T. Wang, S.C. Guo, D.Y. Zhang, K. Jing, L. Feng, Z.F. Zhao, P.Z. Wang

Qifeng Cheng
Research and Development Center,
China Academy of Launch Vehicle Technology,
Fengtai District, Beijing, 100076, P. R. China

Tiantian Wang
School of Science
Liaoning Technical University,
Fuxin, 123000, P. R. China

Sicong Guo, Peizhuang Wang
Institute of Intelligent Engineering and Mathematics
Liaoning Technical University,
Fuxin, 123000, P. R. China

Dayi Zhang*, Kai Jing, Liang Feng, Zhifeng Zhao
General Hospital of Coal Mining Industry Group Fuxin,
Fuxin, 123000, P. R. China
*Corresponding author:18242875460@163.com

Abstract: Logistic regression plays an important role in machine learning. People excitingly use it in conceptual matching yet with some details to be understood further. This paper aims to present a reasonable statement on logistic regression based on fuzzy sets and the factor space theory. An example about breast cancer diagnosis is displayed to show how the factor space theory can be incorporated into the understanding and use of logistic regression.
Keywords: logistic regression, factor space theory, fuzzy sets, logistic membership function

1 Introduction

In 1965, Zadeh put forward the concept of fuzzy subset which made ordinary subset generalize that the range of grade of membership has been relaxed from binary variables {0,1} to continuous variables [0,1] [23]. Today, we are facing the tide of big data, the essential meaning of fuzzy set is shined by its original definition still. Along the way, the factor space theory has been established to provide a deeper mathematical foundation for artificial intelligence [13–15]. The factor space theory builds a bridge connecting certainty and uncertainty and the two sides can be transferred to each other by changing the dimension of related factor space [16].

The falling shadow theory [17] based on factor space was developed to compare fuzziness with randomness. No matter randomness or fuzziness, they are both caused by a lack of factors. Randomness is a kind of uncertainty, which is caused by the lack of conditional factors for predicating. Randomness breaks the law of causality, while probability theory replaces it by a generalized causality law: even though insufficient conditions can not determine whether an event will occur or not, they determine the occurrence of an event with a certain probability. Fuzziness is another kind of uncertainty, which is caused by the lack of identifying factors for recognition. Fuzziness breaks the law of excluded middle, while fuzzy theory replaces it by a generalized law of excluded middle: even though insufficient identifying factors can not determine if an object conforms a concept, they identify an object to a concept with a certain membership degree. An important relationship between fuzziness and randomness is emerged from the similarity: the fuzziness phenomenon on the ground, the universe U, can be described as a randomness phenomenon on the sky, the power of U [22]. Subjective reliability is the non-additive measure since they are the fallen of probability from sky to ground, this is the core idea of the falling shadow theory.

Logistic regression analysis (LRA) is a simple but effective method which is widely used in classification, extending the techniques of multiple regression analysis to research situations in which the outcome variable is categorical [3]. In stratification research, demographic research and social medicine, the use of logistic regression is routine [8]. Some related works have been done on logistic regression. A forward stepwise logistic regression analysis is conducted and the results show that the peer ratings of collaboration can predict you belong to Learning Problem group or Not Learning Problem group [10]. Split-sample, cross-validation and bootstrapping methods for estimation of internal validity of a logistic regression model are compared, which ends up with the recommendation of bootstrapping [12]. The number of events per variable (EPV) in LRA has influences on the validity of the model, and it turns out that low EPV plays a major role [9]. Yet these works only focus on the strategic level of logistic regression, deep understanding hidden behind this common model should be mined.

In artificial intelligence, the regressed curves can be regarded as membership functions. The selection of the types of regressed curves decides the quality of curve fitting partially. Logistic regression is fitted by the logistic membership function. In this article, we will expand the discussion about logistic regression based on factor space theory and fuzzy sets which can present a relatively different state from before.

The paper is proceeded as follows. We introduce the types of membership functions in Section 2. The logistic membership function is put forward in Section 3. In Section 4, logistic regression based on the factor space theory is discussed further and an algorithm is proposed accordingly. In Section 5, an example is used to identify the effectiveness of the proposed algorithm. Finally, a brief conclusion is made in Section 6.

2 Type of membership functions

Definition 1. A fuzzy subset A defined on universe of discussion U is a mapping μ_A: U→[0,1], $\mu_A(u)$ is called the membership degree of u with respect to A [23].

From Definition 1, two fundamental meanings are revealed: firstly, a fuzzy subset stands for the extension of a fuzzy concept, it is a milestone of intelligence mathematics; secondly, a fuzzy subset builds a bridge to step over the gap between quantitative and qualitative phenomena, it was the main bottleneck of information revolution. While the falling shadow theory can provide a deeper mathematical foundation for the revolutionary change [5, 6, 18]. Definitions 2, 3 and 4 are listed to show how the falling shadow of a random subset is generated.

Definition 2. Let $(U, \mathcal{B})$ be a measurable space. $(\mathcal{P}(U), \underline{\mathcal{B}})$ is called a super-measurable space defined on $(U, \mathcal{B})$ if $(\mathcal{P}(U), \underline{\mathcal{B}})$ is a measurable space.

Definition 3. Given a probabilistic field $(\Omega, \mathcal{F}, p)$ and a super-measurable space $(\mathcal{P}(U), \underline{\mathcal{B}})$ on $(U, \mathcal{B})$. A mapping ξ: $\Omega \to \mathcal{P}(U)$ is called a random set if $\xi^{-1}(\mathcal{A}) = \{\omega \in \Omega \mid \xi(\omega) \in \mathcal{A}\} \in \mathcal{F}$ whenever $\mathcal{A} \in \underline{\mathcal{B}}$.

A probability distribution can be induced on $\underline{\mathcal{B}}$ from p on $\mathcal{F}$ through the mapping ξ,

$$P(\mathcal{A}) = P(\xi^{-1}(\mathcal{A})),\ \mathcal{A} \in \underline{\mathcal{B}},$$

which makes $(\mathcal{P}(U), \underline{\mathcal{B}})$ a super probabilistic field [19].

Definition 4. Given a super probabilistic field $(\mathcal{P}(U), \underline{\mathcal{B}})$ and a random set $\xi : \Omega \to \mathcal{P}(U)$, we call a fuzzy subset A_ξ on U the falling shadow of ξ if $\mu_{A_\xi}(u) = P\big(\omega \mid u \in \xi(\omega)\big)$ for all $u \in U$.

It is revealed in Definition 4 that μ_{A_ξ} can be viewed as the covering function of a random set ξ on U. The random set is called clouds, and the fuzzy set is called the falling shadow of the clouds. The thicker the cloud, the higher the darkness of the shadow of the cloud [19]. Meanwhile, according to Definition 1, μ_{A_ξ} is also the membership function with respect to fuzzy subset A_ξ.

Definition 5. A membership function μ_{A_ξ} is also called the possibility distribution of a concept A_ξ on U .

Possibility varies from probability. According to Definition 5, possibility stands for the covering chance of ξ to u and it does not hold exclusiveness; while probability holds the exclusiveness and stands for the chance of monopolization [20].

Theorem 1. *Let $U = (-\infty, +\infty)$ be the one dimensional state space. Given a random interval r with falling shadow μ_A, let ζ be the left extreme point of random interval r, then ζ is a random variable defined on $(U, \mathcal{B})$. And the possibility distribution of the concept A is the same as the distribution function $F(u)$ of ζ : $\mu_A(u) = F(u) = P(\zeta \le u)$.*

Proof: Given probabilistic field $(\Omega, \mathcal{F}, p)$, measurable space $(U, \mathcal{B})$ and super-measurable space $(\mathcal{P}(U), \underline{\mathcal{B}})$.

Since r is a random set, then

$$r^{-1}(\mathcal{A}) = \{\omega \in \Omega \mid r(\omega) \in \mathcal{A}\} \in \mathcal{F}$$

whenever $\mathcal{A} \in \underline{\mathcal{B}}$.

Because ζ is the left extreme point of r, then for each $\omega \in \Omega$, there exists $\beta \in \mathcal{B}$, which makes

$$r(\omega) \in \mathcal{A} \Longrightarrow \zeta(\omega) \in \beta.$$

Since $\zeta(\omega)$ is a real number on the interval $(-\infty, +\infty)$, according to the definition of random variable [21], it is obvious that ζ is a random variable. And

$$\mu_A(u) = P\big\{\omega \mid u \in r(\omega)\big\} = P\big\{\omega \mid \zeta(\omega) \le u\big\} = P(\zeta \le u).$$

□

From Theorem 1, we can distinguish possibility distribution from probability distribution and combine the two terminologies by means of density function and distributed function respectively.

The membership function μ_A determines the extent that u belongs to a concept A. For one concept, different types of membership functions exhibit different membership degree for a certain point u, making it necessary to clarify which type of membership functions should be chosen accordingly.

Definition 6. A is a fuzzy subset defined on universe of U whose membership function is μ_A. If $\mu_A(x) > min\{\mu_A(a), \mu_A(b)\}$ for any $a < x < b$, then A is a convex fuzzy subset [4].

A convex fuzzy set divides A into five parts with four points l^-, l^+, u^-, u^+ ($l^- \leq l^+ \leq u^- \leq u^+$):

$$\mathbf{k} = (-\infty, l^-],\ \mathbf{l} = (l^-, l^+],\ \mathbf{t} = (l^+, u^-],\ \mathbf{u} = (u^-, u^+],\ \mathbf{v} = (u^+, +\infty).$$

$\mu_A \equiv 0$ when $x \in \mathbf{k}$ or $x \in \mathbf{v}$, $\mu_A \equiv 1$ when $x \in \mathbf{t}$. $(l^-, \mu_A(l^-))$ and $(u^+, \mu_A(u^+))$ are the inflection points.

Definition 7. We call $|\,\mathbf{l}\,|$ and $|\,\mathbf{u}\,|$ the lower and upper interim length respectively.

Even though the shape of membership functions is variable with countless changes, the essential variations focus on the two interim periods. The length of interim reflects the degree of fuzziness with respect to a membership function, the narrower the length of interim, the more precise the representation of a concept. For ease of simplicity, we only discuss the membership function on the left fuzzy interval $\mathbf{l}$ formed by the distribution of left extreme point ζ on the interval. There are three common types of probability density functions of ζ: uniform type, cosine type and normal type. Due to space limitation, only uniform distribution of ζ is displayed here to show how the membership function is generated.

For example, ζ is uniformly distributed on fuzzy segment $(l^-, l^+]$, the probability density function of ζ is

$$f_1(x) = \frac{1}{l^+ - l^-}, \quad x \in (l^-, l^+].$$

According to Theorem 1, the membership curve of the concept on this interval is

$$\mu_1(x) = P_1(\zeta \leq x) = \frac{x - l^-}{l^+ - l^-}. \tag{1}$$

It is a straight line ranging from 0 to 1 on fuzzy segment $(l^-, l^+]$.

Curve fitting is the foundation of optimization, the quality of fitting depends on whether the curve chosen is appropriate or not. In this article, we introduce another type of membership function which is frequently used in classification but lacking deeper understanding, as well as expanding the discussion on it from the viewpoint of factor space.

3 Logistic membership function

Let γ be the random variable defined on U, indicating attribute x. And let y be the indication variable of concept α with the extension A defined on U, which takes value 1 when $u \in A$ and 0 else. Denote $P_x = P\{y = 1 | \gamma = x\}$, which is the possibility distribution of the concept A with respect to the variable x. To estimate the possibility, use the *maximal likelihood principle.* Consider a series of sampling points $(x_1, y_1), \cdots, (x_m, y_m)$, the likelihood function is as follows:

$$\underline{L} = \prod_{i=1}^{m} P_{x_i}^{y_i} (1 - P_{x_i})^{1 - y_i}. \tag{2}$$

It is not easy to calculate the derivative, put logarithm on it and get the new likelihood function:

$$\begin{aligned} L = & \ ln\prod_{i=1}^{m} P_{x_i}^{y_i}(1-P_{x_i})^{1-y_i} \\ = & \sum_{i=1}^{m}\big(y_i lnP_{x_i} + (1-y_i)ln(1-P_{x_i})\big) \\ = & \sum_{i=1}^{m}\big(y_i(lnP_{x_i} - ln(1-P_{x_i}))\big) + \sum_{i=1}^{m} ln(1-P_{x_i}) \\ = & \sum_{i=1}^{m}(y_i ln\frac{P_{x_i}}{1-P_{x_i}}) + \sum_{i=1}^{m} ln(1-P_{x_i}). \end{aligned} \tag{3}$$

Since $ln\frac{P_x}{1-P_x}$ varies on the interval $[0,+\infty)$, also being related to the attribute x, we define $ln\frac{P_x}{1-P_x} = ax+b$, then $P_x = \frac{1}{1+e^{-(ax+b)}}$. And Equation (3) can be transferred into

$$L = \sum_{i=1}^{m}(y_i(ax_i+b)) - \sum_{i=1}^{m} ln(1+e^{ax_i+b}). \tag{4}$$

Maximize L so that a and b can be achieved. This is an optimization problem and some strategies solving this type of problems such as *gradient descent* [**?**] can be applied.

Definition 8. The logistic regression function is defined as:

$$\phi(\mathbf{x};\mathbf{a},b) = \frac{1}{1+e^{-(\mathbf{a}^\top\mathbf{x}+b)}}$$

$\mathbf{x},\mathbf{a} \in R^n$, $\mathbf{a}^\top\mathbf{x} = a_1x_1 + \cdots + a_nx_n$. When $n=1$, $\phi(x;a,b) = \frac{1}{1+e^{-(ax+b)}}$; $\phi(x;1,0) = \frac{1}{1+e^{-x}}$.

It is obvious that this method of estimating a and b is kind of complex, while simplicity is what we pursue eventually. Another method which can transfer the problem into a linear problem will be introduced in the next section.

Since $P_A(x) = \frac{1}{1+e^{-(ax+b)}}$ is central symmetric with respect to the point $(-\frac{b}{a},\frac{1}{2})$, we can get

$$P_{\neg A}(x) = P_A(-\frac{2b}{a}-x) = \frac{1}{1+e^{-(a(-\frac{2b}{a}-x)+b)}} = \frac{1}{1+e^{ax+b}}.$$

And it is easy to know

$$P_A(x) + P_{\neg A}(x) = \frac{1}{1+e^{-(ax+b)}} + \frac{1}{1+e^{ax+b}} = 1.$$

With this property, the logistic regression function $P(x) = \frac{1}{1+e^{-(ax+b)}}$ can be seen as the logistic membership function. And this property is available among another three types of membership functions referred in Section 2, the details of which are omitted here.

Definition 9. We call $P(x) = \frac{1}{1+e^{-a(x-\underline{x})}}$ the logistic interim function.

3.1 Logistic regression in risk attributable factor space

The factor space theory paves the way for logistic regression since some concepts find their footholds in the factor space, which contributes to the better understanding of logistic regression.

Definition 10. A factor space defined on universe of discussion U is a family of set $\Psi = \big(\{X(f)\}_{(f\in F)}; U\big)$ satisfying:

(1) $F = (F, \bigvee, \bigwedge,^c, \mathbf{1}, \mathbf{0})$ is a complete Boolean algebra;

(2) $X(\mathbf{0}) = \{\emptyset\}$;

(3) For any $T \subseteq F$, if $\{f|f \in T\}$ are irreducible (*i.e.*, $s \neq t \Rightarrow s \bigwedge t = \mathbf{0}$ $(s, t \in T)$), then

$$X(\{f|f \in T\}) = \prod_{f\in T} X(f)$$

($\prod$ stands for Cartesian product)

(4) $\forall f \in F$, there is a mapping with same symbol $f : U \to X(f)$.

F is called the set of factors, $f \in F$ is called a factor on U. $X(f)$ is called the state space of factor f [22].

Denote $I_j = X(f_j)$ and $I = I_1 \times \cdots \times I_n$.

Definition 11. Given an attribute space $O = \{o_{i_1\cdots i_n}|i_j \in I_j, j = 1, \cdots, n\})$ in factor space $\Psi = (\{X(f)\}_{(f\in F=\{f_1,\cdots,f_n\})}; U)$. For each $i_j \in I_j$,

$$o_{i_1\cdots i_n}$$

is called a granule of $X(F)$ [5]. Denote that $q_{i_1\cdots i_n} = P\big\{u|F(u) = o_{i_1\cdots i_n}\big\}$,

$$\mathbf{P} = \{q_{i_1\cdots i_n}|i_j \in I_j, j = 1, \cdots, n\}$$

is called the probability distribution of attributes.

Assumption 1. $X(f_1), X(f_2)\cdots, X(f_n)$ are all partial ordered sets, it means for one attribute f_i whose range of values is $\{i_{j_1}, i_{j_2}, \cdots, i_{j_r}\}$, $i_{j_1} \le i_{j_2} \le \cdots \le i_{j_r}$ is always correct. It should be clear that if there exists two granules $o_{1\cdots1\cdots1}$ and $o_{1\cdots3\cdots1}$, granule $o_{1\cdots2\cdots1}$ is anyhow certain to appear. Then we naturally suppose that $\boldsymbol{P}$ is convex.

A convex probability distribution of attributes $\boldsymbol{P}$ is called the background distribution with respect to universe U [7].

Definition 12. A factor space $\tilde{\Psi} = \big(\{X(f)\}_{(f\in\tilde{F})}; U\big)$ is called a risk attribute factor space if $\tilde{F} = \{f_1, \cdots, f_n; f_{n+1}\}$, where $f_1, \cdots, f_n$ stand for attribute factors and f_{n+1} stands for a risk factor with binary attribute space $X(f_{n+1}) = \{1, 0\}$.

Given a group of sampling points on $I \times \{1, 0\}$:

$$S = \{(x_{1i}, \cdots, x_{ni}; y_i)\}_{(i=1,\cdots,m)}.$$

For $\big(i_1, \cdots, i_n; i_{n+1}\big) \in I \times \{1, 0\}$, denote

$$q_{i_1\cdots i_n i_{n+1}} = \frac{|\ \{t \mid x_{1t} = i_1, \cdots, x_{nt} = i_n; y_t = i_{n+1}\}\ |}{m}. \tag{5}$$

It is obvious that $\sum_{i_1}\cdots\sum_{i_n}\sum_{i_{n+1}} q_{i_1\cdots i_n i_{n+1}} = 1$.

Definition 13. $\tilde{Q} = \{i_1, \cdots, i_n; i_{n+1} \in I \times \{1, 0\} \mid q_{i_1\cdots i_n i_{n+1}} > 0\}$ is the support set of $Q = \{q_{i_1\cdots i_n i_{n+1}}\}_{(i_1,\cdots,i_n;i_{n+1})\in I\times\{1,0\}}$.

In the risk attribute factor space $\tilde{\Psi}$, logistic regression is mainly based on the support set $\tilde{Q}$. While the condition that $q_{i_1\cdots i_n i_{n+1}} = 0$ still exists and transformation used to handle it will be discussed in the final part of this section. Also, the definition of the logistic membership function fitted out through logistic regression get updated in $\tilde{\Psi}$.

Definition 14. The membership function of illness α:

$$P(X) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n)}} = \frac{1}{1 + e^{-\boldsymbol{\theta}^\top X}},$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_n)^\top$, $X = (1, x_1, \cdots, x_n)^\top$. The parameters $\theta_1, \cdots, \theta_n$ are called risk attribute coefficients of factors.

The bigger the coefficient θ_j, the more important the factor f_j on risk-increasing.

Definition 15. Let α be a concept with extension $A \subseteq U$ and intension $O \subseteq X(F)$, we call conditional probability $P\{u \in A | F(u) = o_{i_1\cdots i_n}\}$ the possibility of concept α under $o_{i_1\cdots i_n}$.

Remark 2. The difference between possibility and probability is clear again. Since possibility is equal to the membership degree of the factorial configuration with respect to concept α, the logistic membership function of illness α is

$$\begin{aligned} p_{i_1\cdots i_n} &= P\{u \in A | F(u) = a_{i_1\cdots i_n}\} \\ &= \frac{|\ \{t \mid x_{1t} = i_1, \cdots, x_{nt} = i_n; y_t = 1\}\ |}{|\ \{t \mid x_{1t} = i_1, \cdots, x_{nt} = i_n\}\ |}. \end{aligned} \tag{6}$$

Also, the probability corresponding to possibility $p_{i_1\cdots i_n}$ is

$$\begin{aligned} q_{i_1\cdots i_n, i_{n+1}=1} &= P\{u \in A | F(u) = o_{i_1\cdots i_n}\} p\{F(u) = o_{i_1\cdots i_n}\} \\ &= \frac{|\ \{t \mid x_{1t} = i_1, \cdots, x_{nt} = i_n; y_t = 1\}\ |}{m}. \end{aligned} \tag{7}$$

It is obvious that

$$\sum p_{i_1\cdots i_n} \neq \sum q_{i_1\cdots i_n, i_{n+1}=1}. \tag{8}$$

Then the approach to estimate θ through logistic regression is developed as Algorithm 1, and the membership degree for the unknown samples can be calculated.

Algorithm 1 Logistic Regression Algorithm

Given sample points $S = \{(x_{1i}, \cdots, x_{ni}; y_i)\}$ $(i = 1, \cdots, m)$.
For $|\ \{t \mid x_{1t} = i_1, \cdots, x_{nt} = i_n\}\ | > 0$ $(i_1 \in I_1, \cdots, i_n \in I_n)$:
if $|\ \{t \mid y_t = 1\}\ | > 0$ and $|\ \{t \mid y_t = 0\}\ | > 0$, **then**
    calculate $P_{i_1\cdots i_n}$
    let

$$y_{i_1\cdots i_n} = ln \frac{P_{i_1\cdots i_n}}{1 - P_{i_1\cdots i_n}}$$

else

$$y_{i_1\cdots i_n} = ln \frac{|\{t|y_t=1\}| + 0.5}{|\{t|y_t=0\}| + 0.5}$$ [1]

end if
Set $y_{i_1\cdots i_n} = \theta_0 + \theta_1 i_1 + \cdots + \theta_n i_n$, do linear regression to get the coefficients $\theta_0, \theta_1, \cdots, \theta_n$.

4 Example

To show how the factor space theory is imbedded into logistic regression, in this section, an example calculating the membership degree of breast cancer using Algorithm 1 is considered. The data extracted from [2] are shown in Table 1 and Table 2. There are 1896 samples in total which form the universe of discussion U. Four attribute factors f_1, f_2, f_3, f_4 are considered and each factor is a categorical variable whose set of value is $I_1 = \{0,1,2\}, I_2 = \{0,1,2,3\}, I_3 = \{0,1\}, I_4 = \{0,1,2\}$ respectively. The meaning of these coded numbers are shown in Table 4. Then the factor space $\Psi = \big(\{X(f)\}_{(f\in F=f_1,f_2,f_3,f_4)}; U\big)$ involving 72 different granules $o_{i_1\cdots i_4}$ of $X(F)$ can be established. The risk attribute factor space $\tilde{\Psi} = \big(\{X(f)\}_{(f\in \tilde{F}=f_1,f_2,f_3,f_4;f_5)}; U\big)$ is also confirmed, where $X(f_5) = \{1,0\}$.

The data must be processed before using Algorithm 1. For the background distribution $\boldsymbol{P}$, there are some granules whose $q_{i_1\cdots i_4} = 0$ which means no samples are included when $f_1 = i_1, \cdots, f_4 = i_4$, then those granules should be deleted. For some granules, the samples are very small and there is little meaning to consider them, so the granules whose total samples are less than 5 are deleted. Hence 33 granules remain and the samples decrease to 1837. Algorithm 1 is applied to the 1837 samples and we can get the membership function of breast cancer as follows:

$$P(X) = \frac{1}{1+e^{-(0.19x_1+0.24x_1+0.36x_3+0.80x_4-0.75)}}. \tag{9}$$

Equation (9) represents a hyperplane in five-dimensional space. Although the interim still exists, we ignore it on account of the complexity. The estimated membership degree for each granule can be calculated. Table 5 depicts the real possibility $P_{i_1\cdots i_4}$, the calculated possibility $P(X_i)$ and probability $q_{i_1\cdots i_4,i_5=1}$. Then Equation (8) can be verified easily:

$$\sum P_{i_1\cdots i_4} = 18.5227, \ \sum q_{i_1\cdots i_4,i_5=1} = 0.4899.$$

In Table 5, the rows that the difference between $P_{i_1\cdots i_4}$ and $P(X_i)$ are over 0.2 are marked in red, only four granules $13th$, $16th$, $21st$ and $32nd$ are included. Since membership degree is estimated through linear regression, it is necessary to compare $y_{i_1\cdots i_4}$ and $ln\frac{P(X_i)}{1-P(X_i)}$, Figure 1 shows the differences of their values. The circles represent the real differences between $y_{i_1\cdots i_4}$ and $ln\frac{P(X_i)}{1-P(X_i)}$. Since *least-square estimation* is applied here, it is necessary to do parameter test. The vertical line segments represent the 95% confidence intervals of random residuals. It is obvious that the confidence intervals of the $16th$, $21st$ and $32nd$ cases don't cover 0, yet not far away from 0. Errors from different angles indicate that logistic regression is applicable in this example.

From Equation (9), we know that θ_i $(i = 1,2,3,4)$ is $0.19, 0.24, 0.36, 0.80$ accordingly. θ_i $(i = 1,2,3,4)$ are all positive numbers which means the risk of having breast cancer will grow when the coded number of factor i gets larger. This is in accordance with the existing knowledge: the risk of having breast cancer will become larger if girls get their first period at a younger age; nulliparous women and those who give the first birth when they are older will have higher risks of having breast cancer; doing previous breast biopsies[1] means you are suspected of having breast cancer, the more you do, the larger the chance of being suspected will be; the risk of having breast cancer will grow if the number of people with breast cancer in near relations goes up. This conclusion verifies the efficiency of the logistic regression from another perspective.

[1] A biopsy is a medical test commonly performed by a surgeon, interventional radiologist, or an interventional cardiologist involving extraction of sample cells or tissues for examination to determine the presence or extent of a disease.

5 Conclusions

The possibility in logistic regression is equal to the membership degree in factor space. Connection between the two sides is established through the membership function. The paper shows that the factor space theory gives logistic regression a relatively different state. Meanwhile, logistic regression forms another foothold of the factor space theory in the big data era. This is of great meaning since it gives us the reason and motivation to explore the factor space theory deeply.

Acknowledgements

This work is supported in part by Natural Science Foundation of Liaoning Province, China [2015020570] and National Natural Science Foundation (NNSF) of China under Grant [61304090].

Bibliography

[1] Bao Y.K (2011); *Data Analysis*, Tsinghua University Press, 2011.

[2] Bruzzi B.P, Sylvan B.G, David P.B, Louise A.B, Catherine S. (1985); Estimating the Population Attributable Risk for Multiple Risk Factors Using Case-Control Data, *American Journal of Epidemiology*, 122(5), 904-914, 1985.

[3] Dayton C.M. (1992); Logistic Regression Analysis, *University of Maryland*, 1992.

[4] Guo S.C, Chen G. (2001); *Method of Soft Computing in Information Science*, Northeastern University Press, 2001.

[5] Li H.X, Wang P.Z. (1994); *A Mathematical Theory on Knowledge Representation*, Tianjin Scientific and Technical Press, 1994.

[6] Liu Z.L, Liu Y.C. (1992); *Theory of Factorial Neural Networks*, Beijing Normal University Press, 1992.

[7] Liu H.T, Wang P.Z; *Background distribution and fuzzy background relationship (to be published).*

[8] Mood C. (2010); Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It, *European Sociological Review*, 26(1), 67-82, 2010.

[9] Peduzzi P., Concato J., Kemper K. (1996);A simulation study of the number of events per variable in logistic regression analysis, *Journal of clinical epidemiology*, 49(12), 1373-1379, 1996.

[10] Prette Z. A. P. D, Prette A.D, Oliveira L.A.D, Gresham F.M, Vance M.J. (2012); Role of social performance in predicting learning problems: Prediction of risk using logistic regression analysis, *School Psychology International*, 33(6), 615-630, 2012.

[11] Russell S.J, Norvig P. (2011); *Artificial Intelligence A modern Approach (Third Edition)*, Tsinghua University Press, 2011.

[12] Steyerberga E.W, Jrb F. E. H, Borsbooma G. J. J. M. (2001); Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis, *Journal of Clinical Epidemiology*, 54, 774-781, 2001.

[13] Wang P.Z. (2013); Factor space and factor data bases, *Journal of Liaoning Engineering and Technology University*, 32(10), 1-8, 2013.

[14] Wang P.Z, Guo S.C, Bao Y.K, Liu H.T. (2014); Causality analysis in factor spaces, *Journal of Liaoning Engineering and Technology University*, 33(7), 1-6, 2014.

[15] Wang P.Z. (2015); Factor space and data science, *Journal of Liaoning Engineering and Technology University*, 34(2), 273-280, 2015.

[16] Wang P.Z. (1981); *In: Hao BL et al (eds) Advance of statistical physics.Science*, Beijing, 1981.

[17] Wang P.Z. (1985); *Fuzzy Sets and Falling Shadow of Random Sets*, Beijing Normal University Press, 1985.

[18] Wang P.Z, Sugeno M. (1982); The factors field and background structure for fuzzy subsets, *Fuzzy Mathematics*, (2), 45-54, 1982.

[19] Wang P.Z, Zhang H.M, Ma X.W, Xu W. (1991); Fuzzy set-operations represented by falling shadow theory, in: Fuzzy Engineering toward Human Friendly Systems, *Proceedings of the International Fuzzy Engineering Symposium '91, Yokohama, Japan*, 1, 82-90, 1991.

[20] Wang P.Z, Liu H.C, Zhang X.H. (1993); Win-win strategy for probability and fuzzy mathematics, *The Journal of Fuzzy Mathematics*, 1(1), 223-231, 1993.

[21] Wang P.Z. (2013); *Fuzzy Mathematics and Optimization*,Beijing Normal University Press, 2013.

[22] Wang P.Z, Liu Z.L, Shi Y, Guo S.C. (2014); Factor Space, the Theoretical Base of Data Science, *Annals of Data Science*, 1(2), 233-251, 2014.

[23] Zadeh L.A. (1965); Fuzzy Subsets, *Information and Control*, 8, 338-353, 1965.

Appendix

Table 1: Detailed distribution of cases and their matched controls in all strata defined by cross-classifying the four risk factors

Risk factor category					
Age at menarche	Age at first live birth	Previous breast biopsies	Mothers plus sisters with breast cancer	No. of cases	No. of controls
0	0	0	0	14	27
0	0	0	1	3	1
0	0	0	2	0	0
0	0	1	0	3	1
0	0	1	1	1	0
0	0	1	2	0	0
0	1	0	0	54	85
0	1	0	1	20	12
0	1	0	2	1	1
0	1	1	0	5	5
0	1	1	1	2	0
0	1	1	2	0	0
0	2	0	0	81	100
0	2	0	1	18	20
0	2	0	2	3	0
0	2	1	0	7	12
0	2	1	1	4	2
0	2	1	2	0	0
0	3	0	0	27	14
0	3	0	1	12	7
0	3	0	2	1	0
0	3	1	0	0	2
0	3	1	1	1	0
0	3	1	2	1	0
1	0	0	0	27	56
1	0	0	1	8	7
1	0	0	2	1	0
1	0	1	0	1	4
1	0	1	1	0	0
1	0	1	2	0	0
1	1	0	0	112	173
1	1	0	1	27	12

Table 2: Table 1 - Continued

Risk factor category					
Age at menarche	Age at first live birth	Previous breast biopsies	Mothers plus sisters with breast cancer	No. of cases	No. of controls
1	1	0	2	4	0
1	1	1	0	14	4
1	1	1	1	1	2
1	1	1	2	0	0
1	2	0	0	187	174
1	2	0	1	41	20
1	2	0	2	10	1
1	2	1	0	11	10
1	2	1	1	5	0
1	2	1	2	0	1
1	3	0	0	41	47
1	3	0	1	15	5
1	3	0	2	4	0
1	3	1	0	4	5
1	3	1	1	1	0
1	3	1	2	1	0
2	0	0	0	9	15
2	0	0	1	3	2
2	0	0	2	2	0
2	0	1	0	1	1
2	0	1	1	0	0
2	0	1	2	0	0
2	1	0	0	43	44
2	1	0	1	14	5
2	1	0	2	1	0
2	1	1	0	3	2
2	1	1	1	2	0
2	1	1	2	0	0
2	2	0	0	53	52
2	2	0	1	9	8
2	2	0	2	2	0

Table 3: Table 2 - Continued

Risk factor category					
Age at menarche	Age at first live birth	Previous breast biopsies	Mothers plus sisters with breast cancer	No. of cases	No. of controls
2	2	1	0	3	1
2	2	1	1	2	1
2	2	1	2	0	0
2	3	0	0	17	4
2	3	0	1	4	3
2	3	0	2	1	0
2	3	1	0	3	0
2	3	1	1	3	0
2	3	1	2	0	0

Table 4: Levels of the risk factors

Risk factor	Range	Coding
	< 12	2
Age at menarche	12-13	1
	≥ 14	0
	< 20	0
Age at first live birth	20-24	1
	25-29(or nulliparous)	2
	≥ 30	3
No. of previous	0 or 1	0
breast biopsies	≥ 1	1
No. of mothers	0	0
plus sisters	1	1
with breast cancer	≥ 2	2

Table 5: $P_{i_1 \cdots i_4}$, $P(X_i)$ and $q_{i_1 \cdots i_4, i_5=1}$ for 33 granules

	Risk factor category						
case number	Age at menarche	Age at first live birth	Previous breast biopsies	Mothers plus sisters with breast cancer	$P_{i_1 \cdots i_4}$	$P(X_i)$	$q_{i_1 \cdots i_4 i_5=1}$
1	0	0	0	0	0.3415	0.3207	0.0076
2	0	1	0	0	0.3885	0.3739	0.0294
3	0	1	0	1	0.6250	0.5711	0.0109
4	0	1	1	0	0.5000	0.4624	0.0027
5	0	2	0	0	0.4475	0.4304	0.0441
6	0	2	0	1	0.4737	0.6275	0.0098
7	0	2	1	0	0.3684	0.5211	0.0038
8	0	2	1	1	0.6667	0.7081	0.0022
9	0	3	0	0	0.6585	0.4887	0.0147
10	0	3	0	1	0.6316	0.6806	0.0065
11	1	0	0	0	0.3253	0.3640	0.0147
12	1	0	0	1	0.5333	0.5606	0.0044
13	1	0	1	0	0.2000	0.4519	0.0005
14	1	1	0	0	0.3930	0.4200	0.0610
15	1	1	0	1	0.6923	0.6175	0.0147
16	1	1	1	0	0.7778	0.5105	0.0076
17	1	2	0	0	0.5180	0.4781	0.1018
18	1	2	0	1	0.6721	0.6713	0.0223
19	1	2	0	2	0.9091	0.8199	0.0054
20	1	2	1	0	0.5238	0.5689	0.0060
21	1	2	1	1	1.0000	0.7463	0.0027
22	1	3	0	0	0.4659	0.5368	0.0223
23	1	3	0	1	0.7500	0.7210	0.0082
24	1	3	1	0	0.4444	0.6254	0.0022
25	2	0	0	0	0.3750	0.4097	0.0049
26	2	0	0	1	0.6000	0.6074	0.0016
27	2	1	0	0	0.4943	0.4675	0.0234
28	2	1	0	1	0.7368	0.6619	0.0076
29	2	1	1	0	0.6000	0.5585	0.0016
30	2	2	0	0	0.5048	0.5263	0.0289
31	2	2	0	1	0.5294	0.7124	0.0049
32	2	3	0	0	0.8095	0.5843	0.0093
33	2	3	0	1	0.5714	0.7580	0.0022

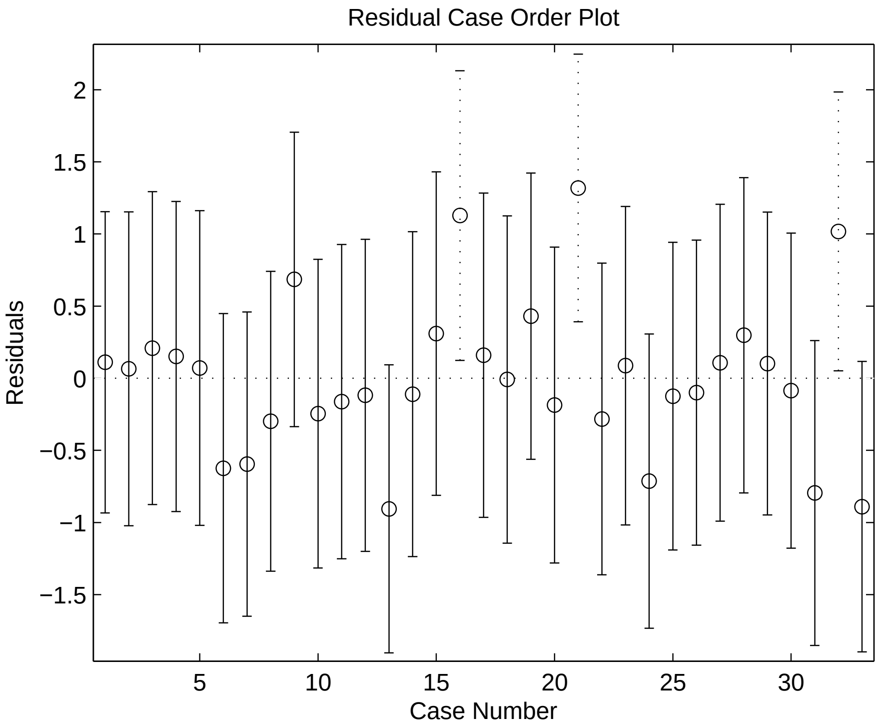

Figure 1: The differences between $y_{i_1 \cdots i_4}$ and $ln \frac{P(X_i)}{1-P(X_i)}$