

رگرسیون نیمه پارامتری فازی بر اساس خوشه‌بندی فازی

معصومه اسداللهی و محمدقاسم اکبری^۱

گروه آمار، دانشگاه بیرجند

تاریخ دریافت: ۱۳۹۵/۱۰/۹ تاریخ پذیرش: ۱۳۹۶/۴/۲۷

چکیده: تحلیل خوشه‌ای از مهم‌ترین روش‌های طبقه‌بندی محسوب می‌شود. در تحلیل خوشه-بندی تلاش می‌شود تا مشاهدات واقع در هر خوشه بیشترین تشابه را از نظر متغیرهای موردنظر باهم داشته باشند. به‌طور کلی روش‌های خوشه‌بندی به دو دسته قطعی و فازی تقسیم می‌شوند. در روش‌های متداول خوشه‌بندی، هر مشاهده تنها در یک خوشه قرار می‌گیرد، اما در خوشه‌بندی فازی، یک مشاهده هم‌زمان در دو یا چند خوشه جای می‌گیرد. در سال ۱۹۶۶، یانگ^۲ و کو^۳[۱۶] خوشه‌بندی k - میانگین معمولی برای حالتی است که داده‌ها به‌صورت فازی مشاهده شده‌اند. یک مدل رگرسیون فازی، برای رابطه‌ی بین متغیرهای مستقل و متغیر وابسته به‌کار می‌رود؛ اما در برخی از موارد پراکندگی و ناهمگنی برخی از مشاهدات باعث می‌شود که یک معادله رگرسیونی نتواند به داده‌ها برازش خوبی داشته باشد. برای رفع این مشکل یانگ و کو[۱۶] داده‌ها را خوشه-بندی نموده و برای هر خوشه یک معادله رگرسیونی بر اساس داده‌های فازی، برازش نموده است. در این مقاله، ابتدا معادله رگرسیون نیمه پارامتری که توسط حسامیان و همکاران [۸] معرفی شده را بیان نموده و سپس با استفاده از آن نویسندگان برای اولین بار از این معادله در خوشه‌بندی با داده‌های فازی استفاده نموده‌اند. لازم به ذکر است که نتایج حاصل از این روش با روش یانگ و کو بر اساس معیارهای نیکویی برازش پیشنهادی، مقایسه می‌کنیم.

واژه‌های کلیدی: عدد فازی، آلفا-شک، خوشه‌بندی فازی، رگرسیون تعمیم‌یافته‌ی یک مرحله-ای، رگرسیون نیمه پارامتری.

رده‌بندی ریاضی (۲۰۱۰): ۹۴D۰۵، ۶۲G۰۸

۱- آدرس الکترونیکی نویسنده مسئول مقاله: g_z_akbari@birjand.ac.ir

2-Yang

3-Ko

۱- مقدمه

برای دسته‌بندی افراد، اشیاء و داده‌ها در گروه‌های (خوشه‌های) مختلف از روش‌های خوشه‌بندی استفاده می‌شود هر خوشه شامل تعدادی از افراد یا اشیاء است که دارای خصوصیات مشابهی باشند. در روش‌های خوشه‌بندی قطعی، هر داده به یک و فقط یک خوشه نسبت داده می‌شود، درحالی‌که در خوشه‌بندی فازی، یک تفکیک فازی صورت می‌گیرد، به این معنی که هر داده با یک درجه عضویت به یک خوشه متعلق است.

پس از تعریف مجموعه‌های فازی توسط زاده [۱۸]، اولین گام در این جهت در سال ۱۹۶۹ توسط رسیپینی^۱ [۱۱] برداشته شد؛ اما اکثر روش‌های خوشه‌بندی فازی، برای بررسی داده‌های دقیق، با استفاده از ایده نظریه‌ی مجموعه‌های فازی مطرح شدند. برای مطالعه بیشتر به یانگ [۱۴] مراجعه کنید. یکی از روش‌های خوشه‌بندی که بر مبنای تابع هدف و استفاده از فاصله‌ی اقلیدسی بنا شده بود، در سال ۱۹۷۴ توسط دان^۲ [۷] ارائه و سپس در سال ۱۹۸۱ توسط بزدک^۳ [۵] تعمیم داده شد. برای بررسی برخی از دیدگاه‌های خوشه‌بندی فازی به بلمن^۴ و همکاران [۴]، اشناتر^۵ [۱۲]، کانن^۶ و همکاران [۶]، یانگ [۱۵] و عارفی [۳] مراجعه کنید. یانگ و کو [۱۶] در سال ۱۹۹۶، یک روش خوشه‌بندی فازی را ارائه کردند. این روش تعمیمی از روش خوشه‌بندی k - میانگین معمولی، برای حالتی است که داده‌ها به صورت فازی مشاهده شده باشند.

در این مقاله، پس از بیان مفاهیم اولیه، به تخمین پارامترهای مدل رگرسیون نیمه پارامتری با استفاده از متر پیشنهادی K_{LR} که بر اساس تفاضل α -شک‌ها، توسط نویسندگان معرفی شده، به روش کمترین قدرمطلق خطا و با استفاده از روش‌های خوشه‌بندی فازی بر اساس روش یانگ و کو، تحت روش رگرسیون فازی تعمیم یافته یک مرحله‌ای می‌پردازیم و نتایج حاصل را با روش یانگ و کو بر اساس معیارهای نیکویی برازش پیشنهادی، مقایسه می‌کنیم. لازم به ذکر است که روش یانگ و کو به طور مفصل در منبع [۱] ذکر شده است.

۲- مفاهیم اولیه

در این بخش به بیان برخی مفاهیم اساسی که در ادامه مقاله مورد نیاز است، می‌پردازیم. برای جزئیات بیشتر به طاهری [۲] مراجعه شود.

-
- 1-Ruspini
 - 2-Dunn
 - 3-Bezdek
 - 4-Bellman
 - 5-Schnatter
 - 6-Cannon

مجموعه‌ی فازی \tilde{A} از مجموعه‌ی مرجع X به صورت $\{(x, \tilde{A}(x)) : x \in X\}$ تعریف می‌شود که $\tilde{A}(x) : X \rightarrow [0, 1]$ تابع عضویت x در \tilde{A} است. مجموعه‌ی فازی \tilde{A} از R مفروض است. اگر \tilde{A} تک‌نمایی باشد (تنها یک $x_0 \in R$ وجود داشته باشد که $\tilde{A}(x_0) = 1$) و α -برش‌های آن به ازای هر $\alpha \in [0, 1]$ بازه‌های بسته و کران‌دار باشند، \tilde{A} را یک عدد فازی گوییم. نوع خاصی از اعداد فازی که به واسطه سادگی در عملیات جبری آن‌ها در بسیاری از مقالات فازی استفاده می‌شود، اعداد فازی LR می‌باشند. اگر عدد فازی \tilde{A} دارای تابع عضویتی به صورت زیر باشد:

$$\tilde{A}(x) = \begin{cases} L\left(\frac{m-x}{l}\right) & x \leq m \\ R\left(\frac{x-m}{r}\right) & x > m \end{cases}$$

که در آن L و R توابعی غیر صعودی از R^+ به $[0, 1]$ هستند و $L(0) = R(0) = 1$ ، آنگاه \tilde{A} را یک عدد فازی LR نامیده و با نماد $\tilde{A} = (m; l, r)_{LR}$ نشان می‌دهیم. L و R توابع مرجع نامیده می‌شود. عدد حقیقی m مقدار نما (میان) و اعداد مثبت l و r به ترتیب پهنای چپ و پهنای راست \tilde{A} نامیده می‌شود. اگر $L(x) = R(x) = \max\{0, 1-x\}$ ، آنگاه \tilde{A} را یک عدد فازی مثلثی نامیده و با نماد $\tilde{A} = (m; l, r)_T$ نشان می‌دهیم.

قضیه ۱: اگر $\tilde{A} = (m; l, r)$ و $\tilde{B} = (n; l', r')$ و $\lambda \in \mathbb{R}$ آنگاه

$$\tilde{A} \oplus \tilde{B} = (m; l, r)_{LR} \oplus (n; l', r')_{LR} = (m+n; l+l', r+r')_{LR},$$

$$\tilde{A} \ominus \tilde{B} = (m; l, r)_{LR} \ominus (n; l', r')_{LR} = (m-n; l+r', r+l')_{LR},$$

$$\lambda \otimes \tilde{A} = \begin{cases} (\lambda m; \lambda l, \lambda r)_{LR} & \lambda > 0 \\ (\lambda m; -\lambda r, -\lambda l)_{RL} & \lambda < 0 \end{cases}$$

تعریف ۱: لیو^۱ [۹] معیاری را برای مقایسه‌ی اعداد فازی و اعداد حقیقی معرفی کرده که از آن به‌عنوان درجه‌ی اعتبار، میزان کوچکی عدد فازی \tilde{A} نسبت به مقدار حقیقی x یاد کرده و به صورت $C : F(\mathbb{R}) \rightarrow [0, 1]$ نمایش داده است که در آن

$$C(\tilde{A} \leq x) = \frac{\sup_{y \leq x} \tilde{A}(y) + 1 - \sup_{y > x} \tilde{A}(y)}{2}$$

تعریف ۲: بزرگ‌ترین کران پایین مجموعه‌ی تمام عناصری از تکیه‌گاه مجموعه‌ی فازی \tilde{A} را \tilde{A}_α می‌دهیم، یعنی

$$\tilde{A}_\alpha = \inf \{x \in \sup p(\tilde{A}) \mid C(\tilde{A} \leq x) \geq \alpha\}$$

که برای $\alpha \in (0, 1]$ یک تابع غیر نزولی است.

تذکر ۱: فرض کنید $\tilde{A} = (m; l, r)_{LR}$ یک عدد فازی LR و $x \in R$ باشد، تابع درجه‌ی اعتبار آن به صورت زیر است:

$$C(\tilde{A} \leq x) = \begin{cases} 0 & x < m - l \\ \frac{1}{2}L\left(\frac{m-x}{l}\right) & m - l \leq x \leq m \\ 1 - \frac{1}{2}R\left(\frac{x-m}{r}\right) & m \leq x \leq m + r \\ 1 & x > m + r \end{cases}$$

و همچنین α -شک آن عبارت است از

$$\tilde{A}_\alpha = \begin{cases} m - lL^{-1}(\alpha) & 0 < \alpha \leq 0.5 \\ m + rR^{-1}(\alpha) & 0.5 < \alpha < 1 \end{cases}$$

تعریف ۳: فرض کنید $\tilde{M} = (m; l, r)$ و $\tilde{N} = (n; l', r')$ دو عدد فازی LR باشند. در این صورت فاصله‌ی بین آن‌ها بر اساس α -شک را با K_{LR} نمایش داده و به صورت زیر تعریف می‌کنیم:

$$K_{LR}(\tilde{M}, \tilde{N}) = \int_0^1 |\tilde{M}_\alpha - \tilde{N}_\alpha| d\alpha.$$

تذکر ۲: اگر \tilde{M} و \tilde{N} دو عدد فازی مثلثی غیرمتقارن باشند، داریم:

$$K_{LR}(\tilde{M}, \tilde{N}) = \frac{1}{4} \int_0^1 |(m-n) + u(l'-l)| du + \frac{1}{4} \int_0^1 |(m-n) + u(r-r')| du.$$

تعریف ۴: برای دو عدد فازی LR ، $\tilde{M} = (m; l, r)$ و $\tilde{N} = (n; l', r')$ ، اگر $l \geq l'$ و $r \geq r'$ آنگاه تفاضل هاکوهارا به صورت زیر تعریف می‌شود:

$$\tilde{M} \ominus_H \tilde{N} = (m - n; l - l', r - r')_{LR}$$

که در صورت عدم شرایط فوق، این تفاضل وجود نخواهد داشت.

۳- رگرسیون نیمه پارامتری در محیط فازی

در این بخش تعمیم روش رابینسن^۱ [۱۰] را برای برآورد پارامتر خطی مدل رگرسیون نیمه پارامتری در محیط فازی با پارامترهای خطی غیر فازی و داده‌های متغیر مستقل و متغیر وابسته فازی ارائه می‌کنیم.

مدل رگرسیون نیمه پارامتری یک متغیره در محیط فازی به صورت زیر تعریف می‌شود:

$$\tilde{y} = (a_0 \otimes \tilde{x}) \oplus \tilde{f}(t) \quad (۱)$$

که در آن $\tilde{y}_i = (y_i; l_{y_i}, r_{y_i})_{LR}$ و $\tilde{f}(t_i) = (f_i; l_{f_i}, r_{f_i})_{LR}$ که برای $i = 1, \dots, n$ مقادیر $t_i = \frac{i}{n}$ به صورت صعودی روی بازه‌ی $[0, 1]$ توزیع شده‌اند. با استفاده از روش هموارساز کرنل، برآورد $\tilde{f}(t_i)$ را به صورت زیر به دست می‌آوریم:

$$\begin{aligned} \hat{\tilde{f}}(t_i) &= \sum_{j=1}^n w_j(t_i) \otimes (\tilde{y}_j \ominus_H (a_0 \otimes \tilde{x}_j)) \\ &= \sum_{j=1}^n w_j(t_i) \otimes \tilde{y}_j \ominus_H \sum_{j=1}^n w_j(t_i) a_0 \otimes \tilde{x}_j. \end{aligned}$$

با جایگذاری این برآورد در رابطه‌ی (۱) به رابطه‌ی زیر می‌رسیم:

$$\begin{aligned} \tilde{y}_i &= (a_0 \otimes \tilde{x}_i) \oplus \sum_{j=1}^n (w_j(t_i) \otimes \tilde{y}_j) \ominus_H \sum_{j=1}^n w_j(t_i) a_0 \otimes \tilde{x}_j \\ &= \sum_{j=1}^n (w_j(t_i) \otimes \tilde{y}_j) \oplus (\tilde{x}_i \ominus_H \sum_{j=1}^n w_j(t_i) \otimes \tilde{x}_j) \otimes a_0 \\ &= \left(\sum_{j=1}^n w_j(t_i) \times y_j; \sum_{j=1}^n w_j(t_i) \times l_{y_j}, \sum_{j=1}^n w_j(t_i) \times r_{y_j} \right)_{LR} \end{aligned}$$

$$\otimes(a_1 \otimes \left(x_i - \sum_{j=1}^n w_j(t_i) \otimes x_j \right));$$

$$sa_1 \otimes \left(l_{x_i} - \sum_{j=1}^n w_j(t_i) \otimes l_{x_j} \right) - (1-s)a_1 \otimes \left(r_{x_i} - \sum_{j=1}^n w_j(t_i) \otimes r_{x_j} \right),$$

$$sa_1 \otimes \left(r_{x_i} - \sum_{j=1}^n w_j(t_i) \otimes r_{x_j} \right) - (1-s)a_1 \otimes \left(l_{x_i} - \sum_{j=1}^n w_j(t_i) \otimes l_{x_j} \right)$$

که در آن

$$s = \begin{cases} 1 & a_1 \geq 0 \\ 0 & a_1 < 0 \end{cases}$$

۴- روش خوشه‌بندی فازی

فرض کنید φ مجموعه‌ای از n داده و c یک عدد صحیح بزرگ‌تر از یک باشد. گروه‌بندی φ به c قسمت، به صورت مجموعه‌های دوبه‌دو مجزای $\varphi_1, \dots, \varphi_c$ با خصوصیت $\varphi_1 \cup \dots \cup \varphi_c = \varphi$ و برای هر $i, j \in \{1, 2, \dots, c\}$ که $i \neq j$ $\varphi_i \cap \varphi_j = \emptyset$ را یک افراز گویند. تابع نشانگر را به صورت زیر در نظر بگیرید:

$$\mu_i(X) = \begin{cases} 1 & X \in \varphi_i \\ 0 & X \notin \varphi_i \end{cases} \quad i = 1, 2, \dots, c$$

بر اساس تابع نشانگر فوق، اگر متغیر X در خوشه‌ی i -ام قرار گیرد، مقدار تابع نشانگر آن یک و در غیر این صورت صفر است. یکی از نارسایی‌های روش‌های خوشه‌بندی قطعی این است که در این روش‌ها هر شیء می‌تواند متعلق به یک و فقط یک خوشه باشد، بنابراین هر شیء به گونه‌ای در یکی از خوشه‌ها جای می‌گیرد، هر چند آن شیء با آن خوشه همگون نباشد. برای حل این مشکل می‌توان از روش‌های خوشه‌بندی فازی استفاده کرد. فرض کنید یک متغیر بتواند به بیش از یک خوشه با یک درجه عضویتی تعلق داشته باشد. فرض کنید $\mu_i(X)$ یک تابع عضویت باشد، به طوری که برای هر $X \in \varphi$ $\sum_{i=1}^c \mu_i(X) = 1$. فرض کنید اگر $\varphi = (\tilde{X}_1, \dots, \tilde{X}_n)$ مجموعه‌ای از n عدد فازی باشد، هدف در این روش خوشه‌بندی φ در c خوشه، بر اساس معیار مینیمم کردن تابع هدف زیر است:

$$F(\mu, \vec{W}) = \sum_{j=1}^n \sum_{i=1}^c \mu_i(\tilde{X}_j) d_f(\tilde{X}_j, \vec{W}_i)$$

که در آن، \tilde{W}_i مرکز خوشه‌ی i -ام، d_f یک فاصله‌ی مناسب بین هر شیء از مرکز خوشه و (\tilde{X}_j) μ_i درجه عضویت شیء j -ام در خوشه‌ی i -ام است. مفهوم پایه‌ای در خوشه‌بندی به روش فازی، درجه عضویت است که نشان می‌دهد یک شیء چه اندازه به خوشه‌ی موردنظر تعلق دارد. تحلیل رگرسیون با استفاده از روش‌های خوشه‌بندی فازی براساس نگرش یانگ و کو [۱۷] و تحت دو روش زیر موردبررسی قرار می‌گیرد:

(۱) رگرسیون فازی وزنی دومرحله‌ای:

به‌منظور یافتن معادله رگرسیونی فازی با داده‌های ناهمگون با استفاده از روش وزنی دومرحله‌ای، ابتدا مشاهدات مربوط به متغیرهای مستقل و متغیر وابسته را به کمک یک روش خوشه‌بندی فازی، خوشه‌بندی کرده و مقادیر درجه عضویت را به‌دست می‌آوریم. سپس این مقادیر را به‌عنوان وزن برای برآورد ضرایب رگرسیونی کمترین مربعات خطا مورداستفاده قرار می‌دهیم.

(۲) رگرسیون فازی تعمیم‌یافته‌ی یک مرحله‌ای:

در این روش خوشه‌بندی فازی و رگرسیون کمترین مربعات در یک روش یک مرحله‌ای با یکدیگر تلفیق می‌شوند، به‌گونه‌ای که در تابع هدف خوشه‌بندی فازی به‌جای مرکز هر خوشه، معادله‌ی خط رگرسیونی مربوط به آن خوشه قرار می‌گیرد و داده‌ها بر مبنای معادله‌ی خط رگرسیونی که بهترین برازش را به داده‌ها داشته باشد، خوشه‌بندی می‌شوند. لازم به ذکر است که در این روش، داده‌ها به‌طور هم‌زمان خوشه‌بندی شده و معادله‌ی خط رگرسیونی برای هر خوشه به‌دست می‌آید. فرض کنید مدل رگرسیونی به‌صورت زیر در نظر گرفته شود:

$$\tilde{y} = a_0 \oplus a_1 \otimes \tilde{x}$$

برای برآورد ضرایب مدل رگرسیونی فوق، بر اساس روش تعمیم‌یافته‌ی یک مرحله‌ای و با استفاده از متر d ، تابع هدف به‌صورت زیر تعریف می‌گردد:

$$H(a_{o_i}, a_{v_i}) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} d(a_{o_i} \oplus a_{v_i} \otimes \tilde{x}_j, \tilde{y}_j),$$

که مشابه تابع هدف روش وزنی دومرحله‌ای است، با این تفاوت که در تابع هدف روش تعمیم‌یافته‌ی یک مرحله‌ای مقادیر μ_{ij} مجهول‌اند و به ازای $j=1, 2, \dots, n$ ، $\sum_{i=1}^c \mu_{ij} = 1$. درواقع در روش وزنی دومرحله‌ای، مقادیر μ_{ij} از قبل از روی یک روش خوشه‌بندی مشخص می‌گردد ولی در این روش، مقادیر μ_{ij} در حین حل مسئله برآورد خواهند شد.

حال مسئله‌ی رگرسیون نیمه پارامتری فازی (رابطه‌ی ۱) بر اساس خوشه‌بندی فازی تحت روش رگرسیون تعمیم‌یافته‌ی یک مرحله‌ای، به صورت زیر در نظر می‌گیریم:

$$\tilde{y}_j = (a_{i_j} \otimes \tilde{x}_j) \oplus \tilde{f}(t_j) \quad t_j = \frac{j}{n}, \quad j = 1, 2, \dots, n, \quad i = 1, 2, \dots, c.$$

به منظور اینکه برآورد معادلات رگرسیونی را بر اساس c خوشه‌ی فازی، مورد بررسی قرار دهیم. مسئله‌ی بهینه‌سازی زیر را در نظر می‌گیریم:

$$\min \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} K_{LR} \left(a_{i_j} \otimes \tilde{x}_j \oplus \tilde{f}(t_j), \tilde{y}_j \right) = \min \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} \int_0^1 \left| \hat{y}_{j\alpha} - \tilde{y}_{j\alpha} \right|,$$

$$s.t. \quad \sum_{i=1}^c \mu_{ij} = 1 \quad j = 1, \dots, n.$$

که در آن μ_{ij} مقادیر درجه عضویت مربوط به خوشه‌بندی مشاهدات است.

در این مقاله جهت برآورد پارامترها و مقادیر درجه عضویت، با مینیمم کردن مسئله‌ی بهینه‌سازی فوق از نرم‌افزار Minitab 13 در مثال‌های عددی استفاده شده است.

۵- نیکویی برازش

در رگرسیون یک موضوع مهم، ارزیابی عملکرد مدل است. در این بخش تعدادی معیار نیکویی برازش برای اندازه‌گیری برازش یک مدل رگرسیونی معرفی می‌کنیم (طاهری و کلکین‌نما [۱۳]).

(۱) معیار خطا

• خطای نسبی برآورد: فرض کنید \tilde{y}_i و $\hat{\tilde{y}}_i$ به ترتیب متغیر وابسته‌ی مشاهده شده و برآورد شده از یک مدل رگرسیونی فازی باشند. تفاضل نسبی بین توابع عضویت بین این دو عدد به صورت زیر تعریف می‌شود:

$$E_s(\tilde{y}_i, \hat{\tilde{y}}_i) = \frac{\int |\tilde{y}_i(x) - \hat{\tilde{y}}_i(x)| dx}{\int \tilde{y}_i(x) dx}$$

که در آن انتگرال‌گیری روی $S_{\tilde{y}_i} \cup S_{\hat{\tilde{y}}_i}$ است که در آن $S_{\tilde{y}_i}$ و $S_{\hat{\tilde{y}}_i}$ به ترتیب تکیه‌گاه \tilde{y}_i و $\hat{\tilde{y}}_i$ هستند.

- خطای برآورد: در تعریف خطای نسبی برآورد اگر فقط تفاضل بین توابع عضویت \tilde{y}_i و \hat{y}_i را در نظر بگیریم و مجموع تابع عضویت عدد فازی مشاهده شده را یک در نظر بگیریم، معیار دیگری به نام خطای برآورد به صورت زیر به دست می‌آید

$$E_r(\tilde{y}_i, \hat{y}_i) = \int |\tilde{y}_i(x) - \hat{y}_i(x)| dx$$

مقادیری که $E_r(\tilde{y}_i, \hat{y}_i)$ و $E_v(\tilde{y}_i, \hat{y}_i)$ می‌توانند اختیار کنند در بازه‌ی $[0, +\infty)$ است که با یک تغییر متغیر به شکل زیر معیارهای $G_v(\tilde{y}_i, \hat{y}_i)$ و $G_r(\tilde{y}_i, \hat{y}_i)$ به دست می‌آید که محدوده‌ی تغییر آن‌ها بازه‌ی $[0, 1]$ است.

$$G_v(\tilde{y}_i, \hat{y}_i) = \frac{1}{1 + E_v(\tilde{y}_i, \hat{y}_i)} \quad G_r(\tilde{y}_i, \hat{y}_i) = \frac{1}{1 + E_r(\tilde{y}_i, \hat{y}_i)}$$

- معیار دیگری که بر اساس تفاضل مقادیر بالایی و پایینی $-\alpha$ برش‌های اعداد فازی مشاهده شده و برآورد شده به دست می‌آید، به شکل زیر تعریف می‌شود:

$$D(\tilde{y}_i, \hat{y}_i) = \frac{1}{2m} \sum_{k=1}^m (|\hat{y}_i^L[\alpha_k] - y_i^L[\alpha_k]| + |\hat{y}_i^U[\alpha_k] - y_i^U[\alpha_k]|)$$

که در آن بازه‌ی $[0, 1]$ (مقادیر α)، به m قسمت تقسیم می‌شود و به طور مثال $\hat{y}_i^L[\alpha_k]$ ، $-i$ امین برآورد کران پایین k -امین $-\alpha$ برش \tilde{y}_i است.

در عمل به منظور ارزیابی نیکویی برازش مدل، از روابط زیر استفاده می‌کنیم:

$$G_v = \frac{\sum_{i=1}^n G_v(\tilde{y}_i, \hat{y}_i)}{n} \quad G_r = \frac{\sum_{i=1}^n G_r(\tilde{y}_i, \hat{y}_i)}{n}$$

$$ME = \frac{\sum_{i=1}^n E_v(\tilde{y}_i, \hat{y}_i)}{n} \quad MD = \frac{\sum_{i=1}^n D(\tilde{y}_i, \hat{y}_i)}{n}$$

(۲) اندازه مشابهت

اگر \tilde{A} و \tilde{B} دو عدد فازی باشند، مقدار مشابهت آن‌ها به شکل زیر تعریف می‌شود:

$$S_{UI}(\tilde{A}, \tilde{B}) = \frac{Card(\tilde{A} \cap \tilde{B})}{Card(\tilde{A} \cup \tilde{B})}$$

که در آن اگر x گسسته باشد $Card(\tilde{A}) = \sum_x \tilde{A}(x)$ و اگر x پیوسته باشد $Card(\tilde{A}) = \int \tilde{A}(x) dx$.

هر چه میانگین اندازه مشابهت اعداد فازی برازش شده و اعداد فازی مشاهده شده پاسخ که به صورت

$$MSM = \frac{\sum_{i=1}^n S_{UI}(\tilde{y}_i, \hat{\tilde{y}}_i)}{n}$$

تعریف می شود به یک نزدیک باشد، مدل بهتر عمل کرده است.

تعریف ۵: میانگین وزنی معیارهای نیکویی برازش به صورت زیر تعریف می شوند.

$$\begin{aligned} G_1(c) &= \sum_{i=1}^c w_i G_{1i} & G_r(c) &= \sum_{i=1}^c w_i G_{ri} \\ ME(c) &= \sum_{i=1}^c w_i ME_i & MD(c) &= \sum_{i=1}^c w_i MD_i \\ MSM(c) &= \sum_{i=1}^c w_i MSM_i \end{aligned}$$

که در آن $w_i = \frac{\sum_{j=1}^n \hat{\mu}_{ij}}{\sum_{i=1}^c \sum_{j=1}^n \hat{\mu}_{ij}}$ است. در ادامه برای ارزیابی عملکرد مدل ها با خوشه های متفاوت از این معیارها استفاده می کنیم.

تذکر ۳: در این مقاله برای مقایسه مدل های مختلف، ابتدا اعداد فازی \tilde{y}_i ($\hat{\tilde{y}}_i$) را به اعداد غیر فازی $M_{\tilde{y}}$ ($M_{\hat{\tilde{y}}}$) تبدیل می کنیم. برای این کار از روش غیر فازی سازی مرکز ثقل نام استفاده می کنیم که به شکل زیر تعریف می شود:

$$M_{\tilde{A}} = \frac{\int x \tilde{A}(x) dx}{\int \tilde{A}(x) dx}$$

که انتگرال گیری روی تکیه گاه \tilde{A} انجام می گیرد.

۶- مثال‌های کاربردی

مثال ۱: داده‌های فازی جدول (۱) که در [۱] آمده را در نظر بگیرید. مقادیر این جدول شامل وزن (برحسب پوند) و فشار سیستولیک خون (BP) یک نمونه تصادفی از افراد در گروه‌های سنی ۲۵ تا ۳۰ سال است. به دلیل عدم دقت کافی در اندازه‌گیری، مشاهدات به صورت اعداد فازی مثلثی در نظر گرفته شده‌اند.

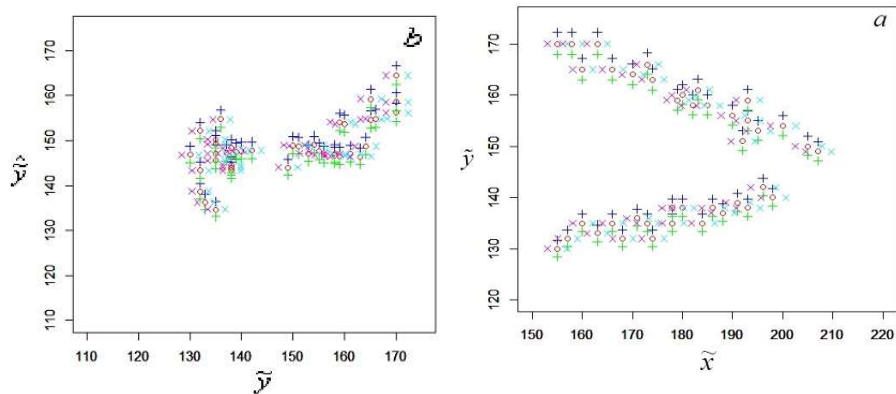
جدول (۱): داده‌های وزن و فشار سیستولیک خون در مثال ۱

شماره	فشار خون	وزن	شماره	فشار خون	وزن
	$\tilde{y} = (m; l, r)_T$	$\tilde{x} = (n; l', r')_T$		$\tilde{y} = (m; l, r)_T$	$\tilde{x} = (n; l', r')_T$
۱	(۱۷, ۱/۶; ۱۳۰)	(۲/۰, ۱/۹; ۱۵۵)	۲۱	(۱/۸, ۱/۷; ۱۴۰)	(۲/۶, ۲/۴; ۱۹۸)
۲	(۲/۱, ۲/۰; ۱۶۵)	(۲/۱, ۱/۹; ۱۶۰)	۲۲	(۲/۰, ۱/۸; ۱۵۴)	(۲/۶, ۲/۴; ۲۰۰)
۳	(۲/۲, ۲/۰; ۱۷۰)	(۲/۱, ۲/۰; ۱۶۳)	۲۳	(۲/۰, ۱/۸; ۱۵۰)	(۲/۷, ۲/۵; ۲۰۵)
۴	(۱/۷, ۱/۶; ۱۳۲)	(۲/۲, ۲/۰; ۱۶۸)	۲۴	(۱/۹, ۱/۸; ۱۴۹)	(۲/۷, ۲/۵; ۲۰۷)
۵	(۲/۱, ۲/۰; ۱۶۴)	(۲/۲, ۲/۰; ۱۷۰)	۲۵	(۱/۷, ۱/۶; ۱۳۲)	(۲/۰, ۱/۹; ۱۵۷)
۶	(۱/۸, ۱/۶; ۱۳۵)	(۲/۲, ۲/۱; ۱۷۳)	۲۶	(۱/۸, ۱/۶; ۱۳۵)	(۲/۱, ۱/۹; ۱۶۰)
۷	(۲/۱, ۲/۰; ۱۶۳)	(۲/۳, ۲/۱; ۱۷۴)	۲۷	(۱/۷, ۱/۶; ۱۳۳)	(۲/۱, ۲/۰; ۱۶۳)
۸	(۱/۷, ۱/۶; ۱۳۲)	(۲/۳, ۲/۱; ۱۷۴)	۲۸	(۱/۸, ۱/۶; ۱۳۵)	(۲/۲, ۲/۰; ۱۶۶)
۹	(۱/۸, ۱/۷; ۱۳۸)	(۲/۳, ۲/۱; ۱۷۸)	۲۹	(۲/۲, ۲/۰; ۱۷۰)	(۲/۰, ۱/۹; ۱۵۵)
۱۰	(۲/۱, ۱/۹; ۱۵۹)	(۲/۳, ۲/۱; ۱۷۹)	۳۰	(۲/۲, ۲/۰; ۱۷۰)	(۲/۱, ۱/۹; ۱۵۸)
۱۱	(۱/۸, ۱/۷; ۱۳۸)	(۲/۳, ۲/۲; ۱۸۰)	۳۱	(۲/۱, ۲/۰; ۱۶۵)	(۲/۲, ۲/۰; ۱۶۶)
۱۲	(۲/۱, ۱/۹; ۱۵۸)	(۲/۴, ۲/۲; ۱۸۲)	۳۲	(۱/۸, ۱/۶; ۱۳۶)	(۲/۲, ۲/۱; ۱۷۱)
۱۳	(۱/۸, ۱/۶; ۱۳۵)	(۲/۴, ۲/۲; ۱۸۴)	۳۳	(۲/۲, ۲/۰; ۱۶۶)	(۲/۲, ۲/۱; ۱۷۳)
۱۴	(۲/۱, ۱/۹; ۱۵۸)	(۲/۴, ۲/۲; ۱۸۵)	۳۴	(۲/۱, ۱/۹; ۱۶۰)	(۲/۳, ۲/۲; ۱۸۰)
۱۵	(۱/۸, ۱/۶; ۱۳۷)	(۲/۴, ۲/۳; ۱۸۸)	۳۵	(۱/۸, ۱/۶; ۱۳۵)	(۲/۳, ۲/۱; ۱۷۸)
۱۶	(۲/۰, ۱/۹; ۱۵۶)	(۲/۵, ۲/۳; ۱۹۰)	۳۶	(۲/۱, ۱/۹; ۱۶۱)	(۲/۴, ۲/۲; ۱۸۳)
۱۷	(۱۹/۸, ۱/۷; ۱۳۹)	(۲/۵, ۲/۳; ۱۹۱)	۳۷	(۱/۸, ۱/۷; ۱۳۸)	(۲/۴, ۲/۲; ۱۸۶)
۱۸	(۲/۰, ۱/۹; ۱۵۵)	(۲/۵, ۲/۳; ۱۹۳)	۳۸	(۱/۸, ۱/۷; ۱۳۸)	(۲/۵, ۲/۳; ۱۹۳)
۱۹	(۱/۸, ۱/۷; ۱۴۲)	(۲/۵, ۲/۴; ۱۹۶)	۳۹	(۲/۰, ۱/۸; ۱۵۱)	(۲/۵, ۲/۳; ۱۹۲)
۲۰	(۲/۰, ۱/۸; ۱۵۳)	(۲/۵, ۲/۳; ۱۹۵)	۴۰	(۲/۱, ۱/۹; ۱۵۹)	(۲/۵, ۲/۳; ۱۹۳)

ابتدا رگرسیون نیمه پارامتری را با تمام داده‌ها ($c = 1$) انجام داده و معیارهای نیکویی برازش آن را محاسبه می‌کنیم. نتایج در جدول (۲) ارائه شده‌اند.

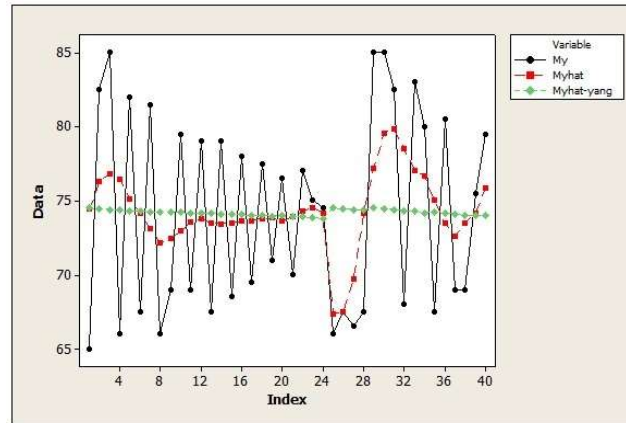
جدول (۲): معیارهای نیکویی برازش با $c = 1$ و $h = 0/1$ در مثال ۱

مورد	MSM	G_1	G_2	ME	MD
مدل پیشنهادی با کرنل سه وزنی $\tilde{y} = (-0.233 \otimes \tilde{x}) \oplus \tilde{f}(t)$	۰/۰۲۷۶	۰/۳۴۷۸	۰/۲۲۶۰	۱/۹۳۲۴	۹/۳۹۱۲
مدل یانگ و کو $\tilde{y} = 153/2796 \oplus (-0.271) \otimes \tilde{x}$	۰/۰۰۰۴	۰/۴۹۲۱	۰/۳۴۳۷	۱/۰۳۲۱	۱۲/۰۴۷۶



شکل (۱): نمودار مشاهدات و نمودار \tilde{y} در مقابل \tilde{x} با $c = 1$ در مثال ۱

نمودار مربوط به مشاهدات فازی که با مقادیر فازی \tilde{x} و \tilde{y} نمایش داده شده، در شکل (۱-a) و نمودار \tilde{y} در مقابل \tilde{x} در شکل (۱-b) رسم شده‌اند. همان‌طور که در شکل (۱-b) در نمودار \tilde{y} در مقابل \tilde{x} ، واضح است داده‌ها حول نیمساز ربع اول به‌گونه‌ای پراکنده‌اند که می‌توان آن‌ها را به دو دسته تقسیم کرد که این موضوع، نیاز به خوشه‌بندی داده‌ها به دو دسته را می‌رساند. اختلاف مقادیر واقعی مشاهدات (My) از مقادیر برآورد شده به دو روش یانگ و کو (Myhat-Yang) و روش رگرسیون نیمه پارامتری (Myhat) در شکل (۲) قابل مشاهده است.



شکل (۲): نمودار $M_{\hat{y}}$ با $M_{\tilde{y}}$ به روش یانگ و کو و رگرسیون نیمه پارامتری با $c = 1$ در مثال ۱

حال داده‌ها را به دو دسته خوشه‌بندی کرده و معادله خط رگرسیونی را برای هر خوشه به دست می‌آوریم. نتایج معیارهای نیکویی برازش برای هر خوشه به‌طور جداگانه، به روش رگرسیون نیمه پارامتری با توابع کرنل مختلف و روش یانگ و کو در جدول (۳) آمده است.

در نهایت معیارهای نیکویی برازش برای داده‌های این مثال در جدول (۴) آمده است. با مقایسه‌ی جدول (۲) با جدول (۴)، ملاحظه می‌شود که اگر مدل رگرسیونی را با دو خوشه برآورد نماییم نتایج بهتری به دست می‌آید و روش پیشنهادی عملکرد بهتری نسبت به روش یانگ و کو داشته است. نمودار مربوط به مشاهدات فازی \tilde{x} و \tilde{y} به همراه خطوط رگرسیونی آن‌ها در شکل (۳) a و نمودار \tilde{y} در مقابل \hat{y} در شکل (۳-b) رسم شده‌اند. در شکل (۳-b) در مقایسه با شکل (۱-b) واضح است که پس از استفاده روش رگرسیون نیمه پارامتری، داده‌ها به‌خوبی حول نیمساز ربع اول پراکنده شده‌اند.

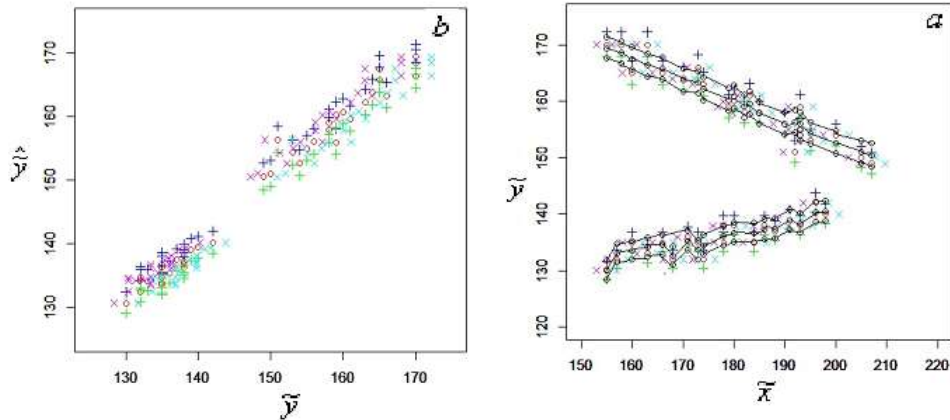
نمودار (۴) اختلاف مشاهدات از برآوردشان در دو روش یانگ و کو و رگرسیون نیمه پارامتری را نشان می‌دهد. واضح است که برآوردها به روش رگرسیون نیمه پارامتری به مقادیر واقعی نزدیک‌ترند و این بیانگر این است که مدل رگرسیون نیمه پارامتری عملکرد بهتری نسبت به مدل یانگ و کو دارد، معیارهای نیکویی برازش نیز گواه بر این ادعا هستند.

جدول (۳): معیارهای نیکویی برازش با $c = 2$ در مثال ۱

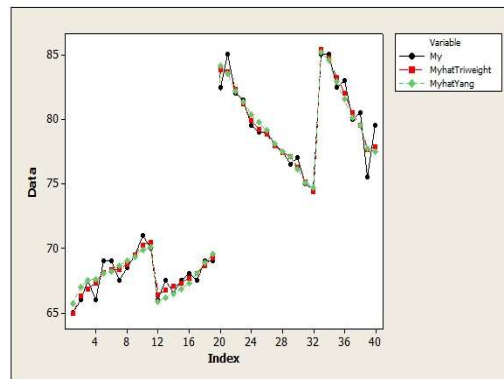
MD	ME	G_p	G_s	MSM	h	مورد
					۰/۱۵۸	خوشه‌ی اول
۱/۰۰۳۳	۰/۹۳۸۲	۰/۴۳۸۲	۰/۵۵۴۷	۰/۴۰۱۶		مدل پیشنهادی با کرنل سه وزنی $\tilde{t} = (0/1996 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۰۴۴۳	۰/۹۶۶۶	۰/۴۲۲۳	۰/۵۴۳۱	۰/۳۸۶۲		مدل پیشنهادی با کرنل مثلثی $\tilde{t} = (0/2236 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۱۷۷۸	۱/۰۵۰۴	۰/۴۰۹۱	۰/۵۲۶۶	۰/۳۵۴۲		مدل پیشنهادی با کرنل نرمال $\tilde{t} = (0/2005 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۴۶۸۱	۱/۰۱۹۹	۰/۳۷۲۱	۰/۵۰۰۹	۰/۱۱۷۴		مدل یانگ و کو $\tilde{t} = 99/6183 \oplus 0/2047 \otimes \tilde{x}$
					۰/۱۴۳	خوشه‌ی دوم
۱/۲۹۴۱	۰/۹۲۶۰	۰/۴۲۳۲	۰/۵۷۲۳	۰/۴۲۴۲		مدل پیشنهادی با کرنل سه وزنی $\tilde{t} = (-0/4313 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۲۳۲۴	۰/۸۹۴۷	۰/۴۴۰۳	۰/۵۸۵۴	۰/۴۴۲۸		مدل پیشنهادی با کرنل مثلثی $\tilde{t} = (-0/4640 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۴۰۴۹	۰/۹۴۸۵	۰/۴۴۶۷	۰/۵۸۳۳	۰/۴۳۰۳		مدل پیشنهادی با کرنل نرمال $\tilde{t} = (-0/4339 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۶۱۰۶	۰/۹۳۴۲	۰/۳۷۴۳	۰/۵۳۷۰	۰/۲۴۸۹		مدل یانگ و کو $\tilde{t} = 332/8386 \oplus (-0/4029) \otimes \tilde{x}$

جدول (۴): میانگین وزنی معیارهای نیکویی برازش با $c = 2$ در مثال ۱

$MD(c = 2)$	$ME(c = 2)$	$G_p(c = 2)$	$G_s(c = 2)$	$MSM(c = 2)$	مدل
۱/۱۲۲۸	۰/۹۱۵۵	۰/۴۳۹۳	۰/۵۷۰۷	۰/۴۲۳۱	مدل پیشنهادی
۱/۵۴۲۴	۰/۹۷۵۲	۰/۳۷۳۲	۰/۵۱۹۷	۰/۱۸۶۰	مدل یانگ و کو



شکل (۳): نمودار مشاهدات و برآورد آن‌ها و نمودار \hat{y} در مقابل \hat{y} با $c = 2$ مثال ۱



شکل (۴): نمودار $M_{\hat{y}}$ با $M_{\hat{y}}$ به روش یانگ و کو و رگرسیون نیمه پارامتری با $c = 2$ در مثال ۱

مثال ۲: داده‌های جدول (۵) را در نظر بگیرید.

می‌خواهیم یک مدل بهینه به روش تعمیم یافته‌ی یک مرحله‌ای، به این داده‌ها برازش دهیم. ابتدا داده‌ها را به یک خوشه و سپس به دو خوشه تقسیم می‌کنیم، نتایج معیارهای نیکویی برازش در جداول (۶) تا (۸) آمده است.

جدول (۵): داده‌های مربوط به مثال ۲

$\tilde{y} = (m; l, r)_T$	x	شماره	$\tilde{y} = (m; l, r)_T$	x	شماره
(۱۰, ۱۰; ۶)	۱۱	۱۱	(۲, ۲; ۱۲)	۱	۱
(۹, ۹; ۸)	۱۲	۱۲	(۲, ۲; ۱۳)	۲	۲
(۸, ۸; ۱۶)	۱۳	۱۳	(۳, ۳; ۲۱)	۳	۳
(۶, ۶; ۱۹)	۱۴	۱۴	(۴, ۴; ۲۹)	۴	۴
(۶, ۶; ۲۶)	۱۵	۱۵	(۴, ۴; ۳۱)	۵	۵
(۴, ۴; ۲۹)	۱۶	۱۶	(۶, ۶; ۳۴)	۶	۶
(۴, ۴; ۳۷)	۱۷	۱۷	(۶, ۶; ۴۵)	۷	۷
(۳, ۳; ۳۹)	۱۸	۱۸	(۸, ۸; ۴۷)	۸	۸
(۲, ۲; ۴۶)	۱۹	۱۹	(۹, ۹; ۴۸)	۹	۹
(۲, ۲; ۴۹)	۲۰	۲۰	(۱۰, ۱۰; ۵۴)	۱۰	۱۰

جدول (۶): معیارهای نیکویی برازش با $c = 1$ و $h = 0/15$ در مثال ۲

MD	ME	G_r	G_l	MSM	مورد
۳/۱۴۰۵	۰/۸۱۹۳	۰/۲۹۶۷	۰/۶۰۲۷	۰/۴۷۷۹	مدل پیشنهادی با کرنل سه وزنی $\tilde{y} = (0/9962 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۹۰۶۶	۲/۰۴۴۹	۰/۱۰۹۸	۰/۳۵۱۸	۰/۰۸۱۳	مدل یانگ و کو $\tilde{y} = (22/044, 5/387, 5/4021) \oplus (0/1957, 0/001, 0/0000) \otimes \tilde{x}$

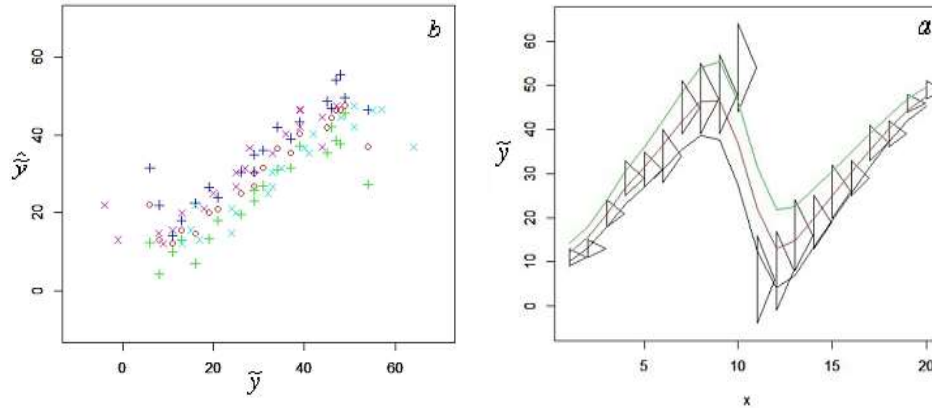
نمودار مربوط به مشاهدات x و \tilde{y} به همراه خطوط رگرسیونی آنها در شکل (۵-a) و نمودار \tilde{y} در مقابل $\hat{\tilde{y}}$ در شکل (۵-b) رسم شده‌اند. نمودار (۶) اختلاف $M_{\tilde{y}}$ با $M_{\hat{\tilde{y}}}$ به دو روش مذکور وقتی داده‌ها را در یک خوشه قرار دادیم را نشان می‌دهد.

جدول (۷): معیارهای نیکویی برازش با $c = 2$ و $h = 0.2$ در مثال ۲

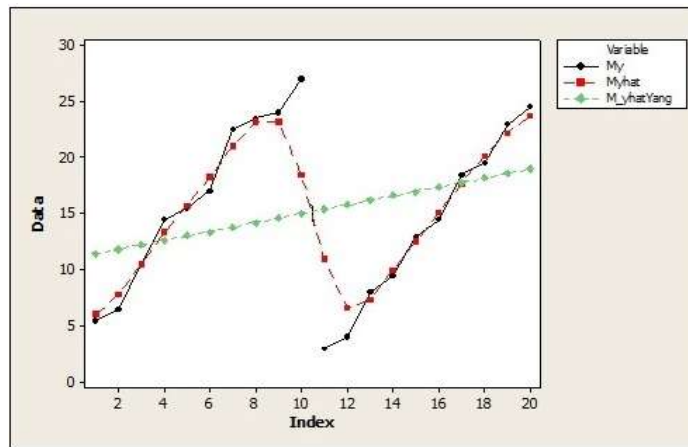
MD	ME	G_+	G_1	MSM	مورد
					خوشه‌ی اول
۱/۰۵۳۶	۰/۴۲۰۲	۰/۴۴۱۷	۰/۷۳۹۶	۰/۶۸۳۰	مدل پیشنهادی با کرنل سه وزنی $\tilde{y} = (0/3195 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۱۵۶۸	۰/۴۵۶۴	۰/۴۲۵۸	۰/۷۲۵۱	۰/۶۶۱۹	مدل پیشنهادی با کرنل مثلثی $\tilde{y} = (0/3361 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۲/۹۴۱۸	۱/۰۳۱۲	۰/۲۲۹۴	۰/۵۵۰۱	۰/۴۰۴۶	مدل پیشنهادی با کرنل نرمال $\tilde{y} = (0/9262 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۸۷۲۷	۰/۶۷۲۱	۰/۳۴۱۶	۰/۶۵۶۳	۰/۵۵۲۹	مدل یانگ و کو $\tilde{y} = (6/3016; 0/5447, 0/329) \oplus (4/911; 0/1199, 0/968) \otimes \tilde{x}$
					خوشه‌ی دوم
۰/۹۹۶۴	۰/۴۵۰۸	۰/۳۶۴۷	۰/۷۰۶۴	۰/۶۴۸۵	مدل پیشنهادی با کرنل سه وزنی $\tilde{y} = (0/7474 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۰۹۳۸	۰/۴۹۱۷	۰/۳۴۴۲	۰/۶۸۹۰	۰/۶۲۳۷	مدل پیشنهادی با کرنل مثلثی $\tilde{y} = (0/7175 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۲/۴۸۵۰	۰/۹۱۳۶	۰/۲۷۵۰	۰/۵۹۴۹	۰/۴۷۰۵	مدل پیشنهادی با کرنل نرمال $\tilde{y} = (0/18970 \otimes \tilde{x}) \oplus \tilde{f}(t)$
۱/۸۸۳۹	۰/۷۹۴۴	۰/۲۵۳۸	۰/۶۰۰۸	۰/۵۳۹۴	مدل یانگ و کو $\tilde{y} = (-48/9935; 5/5109, 4/668) \oplus (4/9409; 0/1000, 0/352) \otimes \tilde{x}$

جدول (۸): میانگین وزنی معیارهای نیکویی برازش با $c = 2$ در مثال ۲

MD (c = 2)	ME (c = 2)	$G_+(c = 2)$	$G_1(c = 2)$	MSM (c = 2)	مدل
۱/۰۲۵۴	۰/۴۳۵۲	۰/۴۰۳۸	۰/۷۲۳۳	۰/۶۶۶۰	مدل پیشنهادی
۱/۸۷۸۲	۰/۷۳۲۳	۰/۲۹۸۴	۰/۶۲۹۰	۰/۵۴۶۳	مدل یانگ و کو

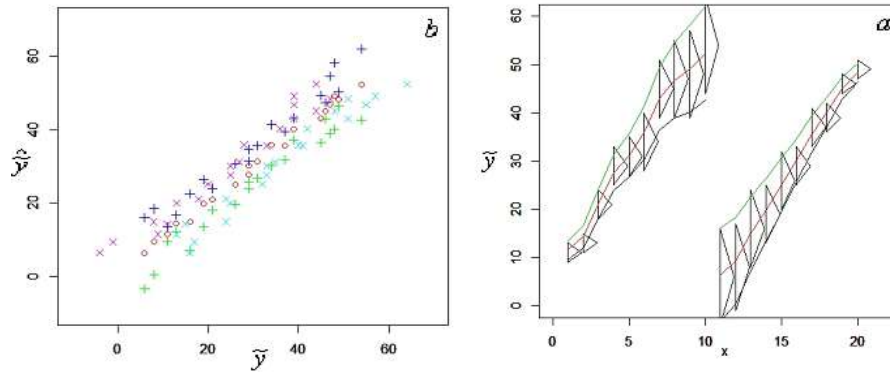


شکل (۵): نمودار مشاهدات و برآورد آن‌ها و نمودار \hat{y} در مقابل \hat{y} با $c = 1$ در مثال ۲

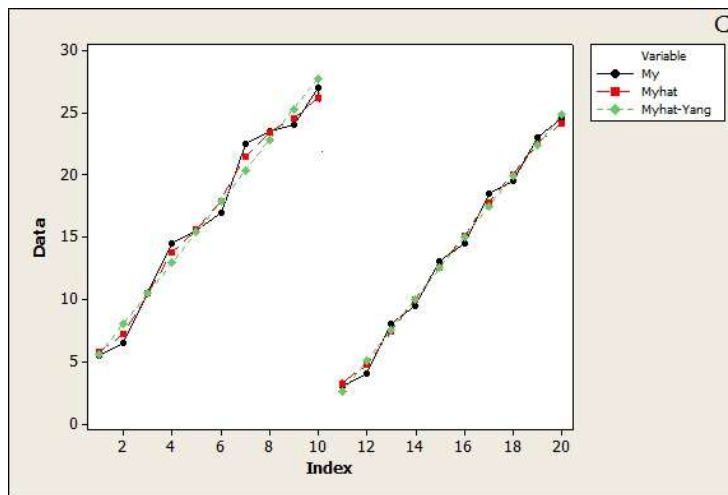


شکل (۶): نمودار $M_{\hat{y}}$ با $M_{\hat{y}}$ به روش یانگ و کو و رگرسیون نیمه پارامتری با $c = 1$ در مثال ۲

نمودارهای (۷-ا) خطوط رگرسیونی برازش شده به داده‌ها، (۷-ب) مقادیر \hat{y} در مقابل \hat{y} و (۷-ج) اختلاف $M_{\hat{y}}$ با $M_{\hat{y}}$ به دو روش رگرسیون نیمه پارامتری و روش یانگ و کو زمانی که $c = 2$ است را نشان می‌دهند. با مقایسه‌ی این نمودار با نمودار (۵-ا)، (۵-ب) و شکل (۶)، ملاحظه می‌شود که اگر داده‌ها به دو خوشه تقسیم شوند نتایج بهتری به دست می‌آید، ضمن اینکه مدل بر اساس روش رگرسیون نیمه پارامتری نسبت به مدل یانگ و کو عملکرد بهتری داشته است.



شکل (۷-ا): نمودار خطوط رگرسیونی برازش شده به داده‌ها با $c = 2$ در مثال
 شکل (۷-ب): نمودار مقادیر \hat{y} در مقابل \hat{y} به داده‌ها با $c = 2$ در مثال



شکل (۷-ج): نمودار اختلاف $M_{\hat{y}}$ با $M_{\hat{y}}$ به دو روش رگرسیون نیمه پارامتری و روش یانگ و کو با $c = 2$ در مثال ۲

نتیجه‌گیری

در این مقاله، چگونگی تعمیم مدل رگرسیون نیمه پارامتری به داده‌های فازی و تحلیل آن با استفاده از خوشه‌بندی فازی بیان شد. همچنین در قالب دو مثال، روش پیشنهادی را با روش

یانگ و کو مقایسه کرده و مشاهده نموده‌ایم که روش بیان شده دارای عملکرد بسیار بهتری در مقایسه با روش یانگ و کو است.

منابع

- [۱] خالقی گزیک، سارا (۱۳۹۳). مدل‌های رگرسیون خطی بر اساس خوشه‌بندی فازی، پایان‌نامه کارشناسی ارشد آمار، دانشکده علوم ریاضی و آمار، دانشگاه بیرجند.
- [۲] طاهری، سید محمود (۱۳۷۵). *آشنایی با نظریه مجموعه‌های فازی*، انتشارات جهاد دانشگاهی دانشگاه فردوسی مشهد.
- [3] Arefi, M. (2016). Clustering regression based on interval-valued fuzzy outputs and interval-valued fuzzy parameters. *Journal of Intelligent and Fuzzy Systems*, **30**, 1339–1351.
- [4] Bellman, R. Kalaba, R. and Zadeh, L.A. (1966). Abstraction and pattern classification, *Journal of Mathematical Analysis and Applications*, **13**, 1-7.
- [5] Bezdek, J.C. (1981). *Pattern recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.
- [6] Cannon, R. Dave, J. and Bezdek, J.C. (1986). Efficient implementation of the fuzzy c-means clustering algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 248-255.
- [7] Dunn, J.C. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics*, **3**, 32-57.
- [8] Hesamian, G. Akbari, M.G. and Asadollahi, M. (2017). Fuzzy semi-parametric partially linear model with fuzzy inputs and fuzzy outputs, *Expert Systems with Applications*, **71**, 230-239.
- [9] Liu, B. (2013). *Uncertainty Theory*, 4th ed., Springer, Berlin.
- [10] Robinson, P. (1988). Root-N-Consistent Semi-parametric Regression, *Econometrics*, **56**, 931–954.
- [11] Ruspini, E. (1969). A new approach to clustering, *Information and Control*, **15**, 22-32.
- [12] Schnatter, S.F. (1992). On statistical inference for fuzzy data with applications to descriptive statistics, *Fuzzy Sets and Systems*, **50**, 143-165.
- [13] Taheri, S.M. and Kelkinnama, M. (2012). Fuzzy linear regression based on least absolute deviations, *Iranian Journal of Fuzzy Systems*, **9**, 121-140.

-
- [14] Yang, M.S. (1993). A survey of fuzzy clustering, *Math. Computer Modeling*, **18**, 1-16.
- [15] Yang, M.S. (1993). On a class of fuzzy classification maximum likelihood procedures, *Fuzzy Sets and Systems*, **57**, 365-375.
- [16] Yang, M.S. and Ko, C.H. (1996). On a class of fuzzy c-numbers clustering procedures for fuzzy data, *Fuzzy Sets and Systems*, **84**, 49-60.
- [17] Yang M.S. and Ko C.H. (1997). On cluster-wise fuzzy regression analysis, *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, **27**, 1-13.
- [18] Zadeh, L.A. (1956). Fuzzy sets, *Information and Control*, **8**, 338-353.

Fuzzy Semi-Parametric Partially Cluster-Wise Regression Analysis

Masoumeh Asadollahi and Mohammad Ghasem Akbari

Department of Statistics, University of Birjand, Birjand, Iran

Abstract

Cluster analysis is one of the most important methods in classification in which the observations of each cluster has maximum similarity in terms of some desirable variables. In general the clustering methods are divided into two parts, crisp and fuzzy. In usual clustering methods an observation is in only one cluster whereas in fuzzy clustering it may fall into two or more clusters simultaneously. Yang and Ko (1996) introduced a fuzzy clustering method. Their method is an extension of the usual k-means clustering method as they assumed that the observations are fuzzy. A fuzzy regression model is used for studying the relationship between the explanatory variables and dependent variable. In some situations when some observations are dispersed and are heterogeneous, the regression model may not have a goodness of fit for data. To solve this problem Yang and Ko classified data and then based on fuzzy observations fitted a regression model to each cluster. In this paper we first explain the semi-parametric regression model introduced by Hesamian et al. [2017] and then use their model to perform our clustering method for fuzzy observations. Finally, based on some suggested goodness of fit criterions. We compare our results with those of Yang and Ko.

Keywords: Fuzzy number, α -pessimistic, Fuzzy clustering, One-stage generalized regression, Semi-parametric regression.

Mathematics Subject Classification (2010): 94D05, 62G08.