



Using Shape Descriptors in Shape Data Classification

Moghimbeygi, M. 

Department of Mathematics, Faculty of Mathematics and Computer Science,
Kharazmi University, Tehran, Iran.

Corresponding author: M. Moghimbeygi, m.moghimbeygi@khu.ac.ir

Received: 13/7/2024 Revised: 2/8/2025 Accepted and Published Online: 4/8/2025.

Introduction

Statistical shape analysis is a growing field that focuses on studying the geometric features of objects. Shape data classification is one of the core challenges in this area, with applications in biology, medical imaging, and pattern recognition. In this paper, we propose a classification model that uses meaningful geometric shape descriptors to enhance interpretability and performance. These descriptors quantify essential properties of shapes such as size, elongation, and convexity, allowing for more accurate distinction between shape classes.

Material and Methods

This study introduces a multinomial logistic regression (MLR) model for shape classification, where the covariates are a set of geometric shape descriptors extracted from labeled configurations. The descriptors include area, elongation (eccentricity), convexity, circularity, centroid size, calliper, length, width, rectangularity, and total absolute curvature (TAC). These features are derived from the raw landmark data using standard shape analysis techniques. The proposed model captures the relationship between these shape characteristics and class membership using MLR, which models the conditional probability of class membership based on descriptor values. Model fitting is achieved via maximum likelihood estimation, and the evaluation is supported by pseudo- R^2 indices (McFadden, Cox-Snell, and Nagelkerke), as well as AIC and BIC criteria. Variable importance is further assessed using the Wald test.

Results and Discussion

Extensive experiments were conducted on both synthetic and real-world datasets. The synthetic study comprised four geometric shape classes—circle, heart, ellipse, and star—with landmark configurations simulated under controlled noise conditions to evaluate classification robustness. Real data included MRI-derived midsagittal brain images (for schizophrenia diagnosis), river and sea sand particle outlines, and mouse molar contours. For all datasets, shape descriptors were extracted using the Momocs and shapes packages in R. The proposed method achieved high classification accuracy across all cases, with overall correct prediction rates exceeding 89% in most scenarios. When compared to a recently published method (Moghimbeygi, 2023), our approach demonstrated superior classification performance and enhanced computational efficiency. Furthermore, the model exhibited strong resilience to noise and variations in dimensionality, although some degree of multicollinearity among shape descriptors was observed in specific datasets.

Conclusion

The classification of shape data presents a significant challenge in statistical shape analysis and machine learning. To address this, we introduce an MLR model that uses geometric shape descriptors, such as area, elongation, convexity, and circularity, as covariates. These descriptors preserve essential geometric information, enhancing classification accuracy. We evaluate the proposed model with results demonstrating high predictive accuracy and effective performance. Therefore, the study confirms that this approach provides a simple, interpretable, and practical method for classifying shape data. Given its strong performance and ease of implementation, the MLR framework with geometric descriptors serves as a valuable tool for shape classification tasks. Future work could extend the method to 3D shape data using volumetric descriptors and incorporate feature selection techniques like PCA or LASSO to mitigate potential multicollinearity.

Keywords: Logistic regression, Shape descriptor, Shape data, Classification.

Mathematics Subject Classification (2020): 62J12, 62H35.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [\(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

استفاده از توصیف‌گرهای شکل در رده‌بندی داده‌های شکل

میشم مقیم بیگی

گروه ریاضی، دانشکده علوم ریاضی و کامپیوتر، دانشگاه خوارزمی

نویسنده مسئول: میثم مقیم بیگی، M.Moghimbeygi@khu.ac.ir

تاریخ دریافت: ۱۴۰۳/۴/۲۳ تاریخ بازنگری: ۱۴۰۴/۵/۱۱ تاریخ پذیرش و انتشار: ۱۴۰۴/۵/۱۳

چکیده: رده‌بندی داده‌های شکل یکی از مسائل مهم در تحلیل آماری اشکال و یادگیری ماشین است. در این مقاله، یک مدل رگرسیون چندجمله‌ای لوژستیک بر اساس توصیف‌گرهای شکل برای رده‌بندی پیکربندی‌های برچسب‌دار معرفی شده است. در این مدل، متغیرهای توضیحی شامل مجموعه‌ای از توصیف‌گرهای هندسی مانند مساحت، کشیدگی، تحدب و دایره‌ای بودن بوده و متغیر پاسخ نشان‌دهنده‌ی رده هر پیکربندی است. استفاده از این توصیف‌گرها امکان حفظ اطلاعات هندسی ضروری را فراهم کرده و دقت رده‌بندی را بهبود می‌بخشد. مدل پیشنهادی در این مقاله با استفاده از داده‌های شبیه‌سازی شده و مجموعه داده‌های واقعی مورد ارزیابی قرار گرفته است. نتایج حاصل نشان‌دهنده‌ی عملکرد مناسب مدل است. علاوه بر این، روش پیشنهادی با یکی از روش‌های موجود در نوشتگان مقایسه شد و نتایج حاکی از برتری آن در دو معیار دقت رده‌بندی و سادگی محاسباتی بود.

واژه‌های کلیدی: رگرسیون لوژستیک، توصیف‌گر شکل، داده شکل، رده‌بندی.

کد موضوع‌بندی ریاضی (۲۰۲۰): 62H35, 62J12.

۱ مقدمه

یکی از حوزه‌های آمار چند متغیره که در آن توجه اصلی بر ساختارهای هندسی اجسام متمرکز است، تحلیل شکل آماری است. تحلیل شکل آماری در اواخر دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰ پس از بحث‌های علمی بین کندانال (۱۹۷۷) و بوکشتین (۱۹۸۶) رایج شد. بر اساس یک تعریف رسمی از شکل که توسط کندانال ارائه شده است: "شکل یک شیء



©نویسندگان). ناشر انجمن آمار ایران است.

این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

تمام اطلاعات هندسی باقی مانده پس از حذف اثرات مکان، مقیاس و دوران از آن شیء است^۱. یکی از موضوعات مورد علاقه محققین آماری در تحلیل شکل، طبقه‌بندی اشکال بر اساس ویژگی‌های خاص آن‌هاست. این طبقه‌بندی معمولاً به دو صورت طبقه‌بندی اشکال بدون برجسب (خوشه‌بندی) و با برجسب (رده‌بندی) انجام می‌گیرد. به عنوان مثال، **دباولار و همکاران (۲۰۲۰)** یک مدل ترکیبی برای خوشه‌بندی داده‌های شکل طولی در کلاس‌های مختلف بدون برجسب معرفی کردند. مدل آن‌ها برای هر خوشه یک مسیر شکل متوسط ایجاد و با استفاده از آن خوشه‌بندی را انجام می‌دهد. از دیگر کاربردهای خوشه‌بندی اشکال، استفاده از آن در تجزیه و تحلیل تصاویر است. برای خوشه‌بندی تصاویر، **سریواستاوا و همکاران (۲۰۰۵)** از اشکال مرزی تصاویر استفاده و ابزارهایی را برای خوشه‌بندی سلسله مراتبی اشکال حاصل از تصاویر ارائه کردند. رده‌بندی اشکال نیز یکی از حوزه‌های مهم در آمار شکل است. برای مثال **سیمو و همکاران (۲۰۲۰)** به رده‌بندی اشکال با استفاده از رگرسیون لوژستیک با متغیرهای کمکی خطی پرداختند و از این مدل در تعیین اندازه لباس کودکان در فروشگاه‌های آنلاین بهره بردند. **مقیم‌بیگی و نودهی (۲۰۲۲)** یک مدل رگرسیون لوژستیک با استفاده از مولفه‌های اصلی و **مقیم‌بیگی (۱۴۰۱)** یک مدل نیمه پارامتری برای رده‌بندی داده‌های شکل معرفی کردند. مطالعات دیگری نیز با ماهیت مدل‌بندی رگرسیونی داده‌های شکل انجام شده است که از جمله این مطالعات می‌توان به **فتوحی و گل‌علی‌زاده (۱۳۹۱)** در بررسی تغییرات شکل با استفاده از تحلیل ژئودزیک اصلی و مطالعه **مقیم‌بیگی و گل‌علی‌زاده (۱۳۹۸)** در مدل‌بندی رگرسیونی داده‌های شکل با استفاده از مثلثی کردن آن اشاره کرد.

توصیف‌گرهای شکل ویژگی‌هایی از شکل هستند که به صورت کمی اطلاعاتی از اشکال را گزارش می‌کنند. امروزه این توصیف‌گرها نقش مهمی را در بینایی ماشین مانند تشخیص الگو، فناوری بیومتریک، تجزیه و تحلیل تصاویر پزشکی بازی می‌کنند (**چاکی و دی، ۲۰۲۰**). یکی از مسائل مهم کاربردی در حوزه تحلیل تصاویر، مقایسه تصاویر و بررسی میزان تفاوت یا تشابه آن‌هاست (**روساکف و همکاران، ۲۰۰۴**). بنابراین طبقه‌بندی تصاویر می‌تواند در حل این مسائل سودمند باشد. توصیف‌گرهای شکل انواع مختلفی دارند که به عنوان مثال می‌توان به میزان دایره‌ای بودن، کشیدگی، تحدب و مساحت شکل اشاره کرد. از آنجایی که رده‌بندی داده‌های شکل با استفاده از توصیف‌گرهای شکل در ادبیات کمتر مورد توجه قرار گرفته است. بنابراین، هدف این مقاله رده‌بندی داده‌های شکل با استفاده از رگرسیون لوژستیک و توصیف‌گرهای شکل است. از این رو در بخش ۲، توصیف‌گرهای شکل معرفی خواهند شد. سپس، مدل رگرسیون لوژستیک چند جمله‌ای بر اساس این توصیف‌گرها در بخش ۳ ارائه می‌شود. عملکرد مدل در یک مطالعه شبیه‌سازی و دو مجموعه داده واقعی در بخش ۴ مورد بررسی قرار خواهد گرفت.

۲ توصیف‌گرهای شکل

در تحلیل آماری شکل به مجموعه نقاط محدودی که هندسه یک شیء را تعیین می‌کنند، نقاط شاخص و مجموعه تمام نقاط شاخص در نظر گرفته شده برای یک شیء را پیکربندی می‌نامند. علاوه بر این، از یک ماتریس $k \times m$ بعدی مانند X برای تعریف پیکربندی استفاده می‌شود که در آن k تعداد نقاط شاخص و m بعد پیکربندی است. طبق

تعریف کندال (۱۹۷۷)، "شکل"، تمام اطلاعات هندسی باقیمانده از یک شیء پس از حذف اثرات مکان، مقیاس و دوران است. اگرچه روش‌های مبتنی بر تعریف شکل کندال (۱۹۷۷) می‌توانند در حل بسیاری از مسائل کاربردی مورد استفاده قرار گیرند اما این عیب را دارند که تطبیق تعاریف و ویژگی‌های آن‌ها دشوار است. یک روش ساده برای تحلیل اشکال استفاده از توصیف‌گرهای شکل است. از این رو در ادامه برخی از توصیف‌گرهای شکل معرفی می‌شوند.

تعریف ۱. مساحت یک شکل: مساحت به سطح محصور به یک شکل غیرمقاطع و محدب گفته می‌شود.

تعریف ۲. دایره‌ای بودن: میزان شباهت به یک دایره کامل را دایره‌ای بودن می‌نامند. میزان دایره‌ای بودن را می‌توان با نسبت مساحت به مجذور محیط شکل ضرب در 4π تعریف کرد. مقدار یک، یک دایره کامل را مشخص می‌کند. با نزدیک شدن مقدار دایره‌ای به صفر، شکل به‌طور فزاینده‌ای کمتر دایره‌ای می‌شود.

تعریف ۳. اندازه مرکزی: این اندازه نرم پیکربندی مرکزی شده است که برای پیکربندی X به‌صورت
$$\sqrt{\sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_j)^2}$$
 محاسبه می‌شود که در آن X_{ij} درایه سطر i -ام و ستون j -ام ماتریس X است.

تعریف ۴. گریز از مرکز کشیدگی: گریز از مرکز مقاطع مخروطی به‌عنوان نسبت فاصله از هر نقطه در مقطع مخروطی به کانون به فاصله عمود از آن نقطه به نزدیکترین جهت تعریف می‌شود. برای هر مقطع مخروطی، گریز از مرکز یک مقطع مخروطی، فاصله هر نقطه از منحنی تا کانون آن تقسیم بر فاصله همان نقطه تا جهت آن مساوی با یک مقدار ثابت است. این مقدار ثابت به‌عنوان گریز از مرکز شناخته می‌شود که با e نشان داده می‌شود. گریز از مرکز یک شکل منحنی تعیین می‌کند که شکل چقدر گرد است. با افزایش گریز از مرکز، انحناها کاهش می‌یابد. به عبارت دیگر $e = c/a$ که در آن c فاصله از هر نقطه در مقطع مخروطی تا کانون آن و a فاصله از هر نقطه در مقطع مخروطی تا جهت آن است.

تعریف ۵. تحدب: با تشکیل نسبت مساحت یک شکل به مساحت بدنه محدب آن، می‌توان میزان تحدب را به‌دست آورد.

تعریف ۶. قطر فرت کولیس^۱: طولانی‌ترین فاصله بین دو نقطه از شکل را قطر فرت می‌نامند.

تعریف ۷. طول و عرض شکل: طول و عرض یک شکل بر اساس محور اینرسی^۲ آن‌ها، یکی از توصیف‌گرهای شکل است. پس از تراز با محور x ، طول محاسبه می‌شود و متناظر با این طول فاصله در امتداد محور y ‌ها به‌عنوان عرض در نظر گرفته می‌شود.

^۱Calliper Feret's diameter

^۲Inertia

۴۵۴ استفاده از توصیف‌گرهای شکل در رده‌بندی داده‌های شکل

تعریف ۸. مستطیلی بودن: مستطیلی بودن یک شکل با تشکیل نسبت مساحت یک شکل به مساحت کوچکترین مستطیل مرزی یک شکل به‌دست می‌آید.

تعریف ۹. یک راست بودن: میزان مطابقت زوایای یک چندضلعی با $\pm 90^\circ$ را یک راست بودن می‌نامند.

تعریف ۱۰. مجموع انحنای مطلق یک شکل: با استفاده از مجموع قدر مطلق مشتق دوم از پیش‌بینی اسپلاین‌های همواری^۱ برای هر نقطه تعریف شده محاسبه می‌شود (روزین، ۲۰۰۵).

توصیف‌گرهای شکل معرفی شده می‌توانند به‌عنوان متغیرهای تبیینی در مدل رگرسیونی مورد استفاده قرار گیرند. محاسبه این توصیف‌گرها توسط بسته Momocs در نرم افزار R به‌راحتی قابل انجام است.

۳ مدل لوژستیک

مدل‌های رگرسیونی در فهم اینکه چگونه مقدار متغیر وابسته با تغییر هر کدام از متغیرهای مستقل و با ثابت بودن دیگر متغیرهای مستقل تغییر می‌کند، کمک می‌کنند. اغلب این مدل‌ها بر اساس متغیرهای مستقل و وابسته کمی معرفی می‌شوند. اما گاهی اوقات لازم است ارتباط متغیرهای تبیینی کمی با متغیرهای پاسخ کیفی سنجیده شود. در این گونه مسائل استفاده از مدل رگرسیون لوژستیک چند جمله‌ای (MLR) پیشنهاد می‌شود. مدل‌بندی رگرسیونی با استفاده از مدل رگرسیون لوژستیک چند جمله‌ای روشی برای رده‌بندی داده‌ها است.

در رگرسیون خطی استاندارد، مقدار مورد انتظار متغیر پاسخ Y به شرط مقدار x را می‌توان به‌صورت $\mathbb{E}(Y|X=x) = \alpha + \beta x$ نوشت که در آن α و β پارامترهای عرض از مبدا و شیب هستند. هنگامی که برآمد Y دو وجهی است، می‌توان از رگرسیون لوژستیک استفاده کرد. روش‌های دیگر برای تحلیل این مدل‌ها را می‌توان در **کاکس و اسنیل (۱۹۸۹)** یافت. با استفاده از تابع لجیت، می‌توان میانگین Y به شرط x را به‌صورت $\pi(x) = \mathbb{E}(Y|x) = e^{\alpha+\beta x} / (1 + e^{\alpha+\beta x})$ تعریف کرد که در آن $\pi(x)$ به‌عنوان احتمال موفقیت نیز شناخته می‌شود. یک تبدیل ساده از $\pi(x)$ که نقش کلیدی در این مقاله ایفا می‌کند تبدیل لجیت $\ln(\pi(x)/(1 - \pi(x))) = \alpha + \beta x$ است. برای محاسبه رگرسیون لوژستیک چند جمله‌ای ابتدا p متغیر کمکی با r رده متفاوت (با پاسخ‌های اسمی Y) و بردار x از متغیرهای تبیینی به طول $p + 1$ را در نظر بگیرید. با انتخاب r -امین رده به‌عنوان یک محور داریم

$$g_m(x) = \ln \frac{Pr(Y = m|x)}{Pr(Y = r|x)} = x^T \beta_m, \quad m = 1, \dots, r-1, \quad (1)$$

که در آن $x = (1, x_1, \dots, x_p)$ برداری به طول $p + 1$ و $\beta_m^T = (\beta_{m0}, \dots, \beta_{mp})$ بردار پارامترها است.

¹Smoothing spline

تعمیمی از احتمال شرطی در r رده ممکن به صورت

$$\pi_j(\mathbf{x}) = Pr(Y = j|\mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{1 + \sum_{k=1}^{r-1} e^{g_k(\mathbf{x})}}, \quad j = 1, \dots, r-1,$$

و برای رده r -ام

$$\pi_r(\mathbf{x}) = Pr(Y = r|\mathbf{x}) = \frac{1}{1 + \sum_{k=1}^{r-1} e^{g_k(\mathbf{x})}},$$

است. برای تابع درستنمایی، از r متغیر تصادفی دو وجهی U_i که مقادیر ۰ یا ۱ را می‌گیرند استفاده می‌شود. بدین صورت که اگر $Y = i$ ، آن‌گاه $U_i = 1$ و برای $i \neq j$ ، $U_i = 0$ باشد. در این صورت بدون در نظر گرفتن مقدار U_i داریم $\sum_{k=1}^r U_k = 1$. به سادگی می‌توان تابع درستنمایی برای n نمونه مستقل از مشاهدات را به صورت $L(\beta) = \prod_{i=1}^n \prod_{k=1}^r \pi_k(\mathbf{x}_i)^{u_{ki}}$ تعریف کرد که در آن u_{ki} مقادیر مشاهده شده از متغیر U است. به منظور برآورد پارامترها می‌توان از لگاریتم تابع درستنمایی استفاده کرد. لذا با در نظر گرفتن شرط $\sum_{k=1}^r U_k = 1$ داریم:

$$\ell(\beta) = \sum_{i=1}^n \sum_{k=1}^{r-1} [u_{ki} \ln \pi_k(\mathbf{x}_i)] - \sum_{i=1}^n \ln \left[1 + \sum_{k=1}^{r-1} e^{g_k(\mathbf{x}_i)} \right]. \quad (2)$$

حال کافی است از تابع $\ell(\beta)$ نسبت به پارامترهای مدل، برای $k = 1, \dots, r$ و $l = 0, \dots, p$ مشتق گرفته شود. در این صورت داریم $\frac{\partial \ell(\beta)}{\partial \beta_{kl}} = \sum_{i=1}^n x_{li}(u_{ki} - \pi_k(\mathbf{x}_i))$ که در آن $x_{li} = 1$ از آنجایی که به دست آوردن یک رابطه روشن برای برآورد پارامترها با استفاده از $\partial \ell(\beta) / \partial \beta_{kl} = 0$ امکان‌پذیر نیست لذا باید به روش‌های عددی متوصل شد. یکی از این روش‌های عددی که کمک می‌کند پارامترها را برآورد کرد شبکه عصبی مصنوعی است. به منظور برآورد پارامترها، شبکه‌های عصبی معمولاً از یک تابع هزینه استفاده می‌کنند. این تابع هزینه در مدل‌های رگرسیونی منفی لگاریتم تابع درستنمایی در نظر گرفته می‌شود. سپس به منظور برآورد پارامترها تابع هزینه را از روش گرادیان کاهشی با حرکت در جهات مختلف کم می‌کنند.

حال فرض کنید بردار $\mathbf{x}$ در رابطه (۱) از توصیف‌گرهای شکل تعریف شده در بخش ۲ تشکیل شده باشد. به عبارت دیگر x_{ij} برداری از متغیرهای تبیینی دایره‌ای بودن، مساحت، اندازه مرکزی، گریز از مرکز، تحذب، قطر فرت کولیس، طول، عرض، مستطیلی بودن و مجموع انحنای مطلق یک شکل برای $i = 1, \dots, K$ و $j = 1, \dots, r$ متناظر با پاسخ‌های اسمی y_{ij} در r رده مختلف است. در نتیجه می‌توان با استفاده از یک مدل لگ-خطی چند جمله‌ای با r رده، یک شبکه عصبی با r خروجی تعریف کرد. این شبکه عصبی با در نظر گرفتن منفی لگاریتم تابع درستنمایی شرطی به صورت $E = - \sum_{i=1}^K \sum_{j=1}^r t_{ij} \log \pi_{ij}$ که در آن $\pi_{ij} = e^{y_{ij}} / \sum_{j=1}^r e^{y_{ij}}$ ، پارامترها را برآورد می‌کند که در آن t_{ij} هدف و y_{ij} خروجی برای ورودی j -ام است (ونابلز و ریپلی، ۲۰۰۲). همچنین نحوه برآورد پارامترها در مدل MLR با استفاده از شبکه عصبی در بسته $nnet$ نرم افزار R موجود است. ارزیابی عملکرد مدل‌های رگرسیونی به وسیله معیارهای مختلفی انجام می‌گیرد. در ادامه برخی از این معیارها

که در این مقاله مورد استفاده قرار گرفته‌اند، معرفی می‌شوند. یکی از معیارهایی که در بررسی مدل‌های آماری کاربرد زیادی دارد، معیار اطلاع آکائیک^۱ (AIC) است. این معیار بر اساس مفهوم آنتروپی بنا شده و تخمین‌گر خطای پیش‌بینی و در نتیجه کیفیت نسبی مدل‌های آماری برای مجموعه‌ای از داده‌ها است (آکائیک، ۱۹۷۳). با استفاده از تابع درستنمایی (۲) معیار آکائیک به صورت $AIC = -2\ell(\hat{\beta}) + 2p$ تعریف می‌شود که در آن p تعداد پارامترها است. در مقایسه بین مدل‌ها، مدلی که دارای مقدار AIC پایین‌تری باشد، ترجیح داده می‌شود. معیار دیگری که از توسیع معیار AIC به دست آمده است معیار اطلاع بیزی^۲ (BIC) است که توسط شوارز (۱۹۷۸) به صورت $BIC = -2\ell(\hat{\beta}) + p \ln(n)$ ارائه شده است که در آن n تعداد نمونه‌ها است.

یکی دیگر از ابزارهایی که در تحلیل مدل‌های رگرسیونی برای بررسی عملکرد مدل از آن استفاده می‌شود معیار ضریب تعیین است. ضریب تعیین نسبت تغییرات در متغیر وابسته است که به وسیله متغیر(های) مستقل قابل پیش‌بینی است. ضریب تعیین را با نماد R^2 نشان می‌دهند و معمولاً در رگرسیون خطی از آن استفاده می‌کنند. استفاده از ضریب تعیین در رگرسیون لوژستیک مرسوم نیست، از این رو ضرایب تعیین‌های مختص این رگرسیون معرفی شده است که از جمله می‌توان به ضرایب تعیین مک‌فادن (MF)، کاکس اسنیل (CS) و ناگلکرک (Na) اشاره کرد، که به ترتیب به صورت

$$R_{MF}^2 = 1 - \frac{\ell(\hat{\beta})_k}{\ell(\hat{\beta})_{null}}, \quad R_{CS}^2 = 1 - \left[\frac{\ell(\hat{\beta})_{null}}{\ell(\hat{\beta})_k} \right]^{2/n}, \quad R_{Na}^2 = \frac{1 - \left[\frac{\ell(\hat{\beta})_{null}}{\ell(\hat{\beta})_k} \right]^{2/n}}{1 - (\ell(\hat{\beta})_{null})^{2/n}},$$

تعریف می‌شوند. در این ضرایب $\ell(\hat{\beta})_{null}$ مقدار تابع درستنمایی تحت فرض $H_0: \beta_{ij} = 0$ برای $i = 1, \dots, r$ و $j = 1, \dots, p$ و $\ell(\hat{\beta})_{Full}$ مقدار تابع درستنمایی تحت فرض تمامی فضای پارامتری است.

۴ مطالعه شبیه‌سازی

در این بخش قصد داریم عملکرد روش معرفی شده را با روش ارائه شده توسط مقیم‌بیگی (۱۴۰۱) مقایسه کنیم. به همین منظور ابتدا چهار پیکربندی مرجع، ستاره، قلب، دایره و بیضی را با استفاده از ده نقطه شاخص ایجاد کرده و سپس به هر نقطه شاخص اجازه می‌دهیم با استفاده از مدل

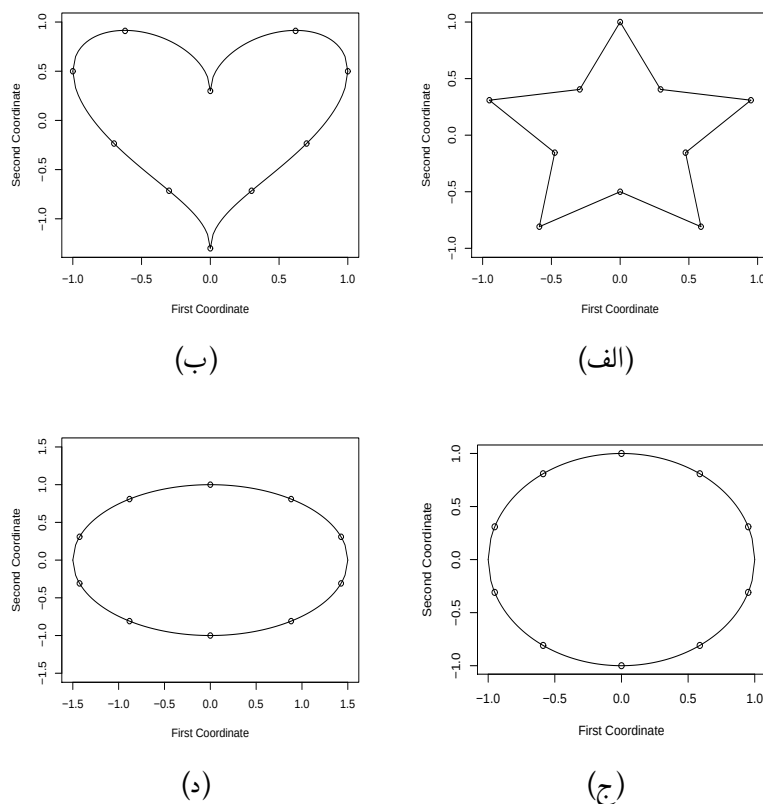
$$X_{ij} = \beta_{ij}\mu_j\Gamma_{ij} + \gamma_{ij}^T + \varepsilon_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, r, \quad (3)$$

که در آن μ_j ، j -امین پیکربندی مرجع و ε_{ij} یک ماتریس تصادفی $k \times m$ به عنوان منبع خطا است که از توزیع نرمال چند متغیره مرکزی با ماتریس کوواریانس $km \times km$ با قطر اصلی σ^2 و عناصر غیر قطری صفر تولید

¹Akaike information criterion

²Bayesian information criterion

می‌شود، جابجا شود. اشکال مرجع در شکل ۱ قابل مشاهده هستند. در این مطالعه پارامترهای مکان γ_{ij} برابر



شکل ۱. نمایی از شکل‌های مرجع الف- ستاره، ب- قلب، ج- دایره و د- بیضی.

صفر، مقیاس β_{ij} برابر ۱ و دوران Γ_{ij} ، ماتریس همانی در نظر گرفته می‌شود. این به این دلیل است که هدف این مطالعه بررسی اثر تغییرات σ بر میزان دقت رده‌بندی است. همچنین باید در نظر داشت که هدف اصلی این مقاله ارائه روشی برای رده‌بندی اشکال است و پارامترهای γ_{ij} ، β_{ij} و Γ_{ij} ، پارامترهای تبدیل تشابه اقلیدسی هستند و فقط پیکربندی‌ها را جابجا، تغییر اندازه و دوران می‌دهند و در ساختار آن‌ها تغییری ایجاد نمی‌کنند. لازم به ذکر است که این روش شبیه‌سازی قبلاً نیز توسط مقیم‌بیگی (۱۴۰۱) انجام گرفته است. شبیه‌سازی داده‌ها با استفاده از مدل (۳) و با مقادیر متفاوت σ انجام می‌گیرد. این شبیه‌سازی برای هر شکل مرجع، به تعداد $10, 30, 50$ نمونه گرفته شده و سپس با استفاده از مدل (۳) رده‌بندی داده‌ها انجام می‌گیرد. به عبارت دیگر از هر یک از پیکربندی‌های ستاره، دایره، قلب و بیضی به تعداد n پیکربندی با استفاده از مدل (۳) تولید می‌شود. نمونه‌ای از این شبیه‌سازی

برای شکل مرجع ستاره با $\sigma = 0.1$ و $\sigma = 0.3$ در شکل ۲ نشان داده شده است. همان‌طور که مشخص است برای $\sigma = 0.3$ ساختار شکل ستاره به‌کلی به‌هم ریخته شده است. برای مقادیر بیشتر $\sigma = 0.3$ عملاً تشخیص چشمی اشکال بسیار دشوار است. این عمل ۱۰۰ بار برای هر مقدار σ و هر تعداد n انجام گرفته و میانگین و انحراف معیار معیارهای ضریب تعیین-نمای^۱ مک‌فادن، کاکس‌اسنیل و ناگلکرک و همچنین معیارهای AIC، BIC و درصد پیش‌بینی صحیح کلی (OCP) برای آن‌ها محاسبه می‌شود. به عبارت دیگر به تعداد n نمونه براساس σ های مختلف از اشکال گرفته شده و رده‌بندی برای آن‌ها انجام می‌شود. این عمل ۱۰۰ بار تکرار می‌شود. این مقادیر در جدول ۱ قابل مشاهده است. به‌وضوح با افزایش میزان σ میانگین ضرایب تعیین‌نما کاهش ولی معیارهای AIC و BIC به‌صورت چشم‌گیری افزایش می‌یابند. شکل ۳ نمایی از این تغییرات را براساس σ مختلف در دو معیار AIC و ضریب تعیین‌نمای کاکس‌اسنیل نشان می‌دهد. لازم به ذکر است که σ نقش تغییرات را در هر نقطه بازی می‌کند. مقادیر بزرگ σ منجر به تغییرات زیاد در نقاط و بالطبع تغییرات زیاد در پیکربندی می‌شود. این تغییرات برای مقادیر کوچک σ توجیه پذیر است ولی برای مقادیر بیشتر از $\sigma = 0.5$ منجر به تغییرات بنیادی در شکل می‌شود و عملاً ماهیت شکل تغییر می‌کند. این تغییرات بنیادی در ساختار شکل منجر به کاهش شدید معیار OCP به حدود ۵۰ درصد برای حالت‌های دودویی می‌شود. در رده‌بندی اشکال واقعی، ما با اشکالی مواجه هستیم که با یکدیگر اشتراک‌هایی (همبستگی) دارند. این اشتراک‌ها عامل اصلی در رده‌بندی است. برای مقادیر بزرگ σ عملاً با اشکالی روبرو هستیم که از یکدیگر مستقلند که این حالات خاص در مسائل کاربردی بی‌فایده است. لذا در این مثال شبیه‌سازی، برای مقادیر بیشتر از $\sigma = 0.5$ شبیه‌سازی انجام نمی‌گیرد. در یک مطالعه دیگر اثر مقدار σ بر دو توصیف‌گر مساحت و کشیدگی بررسی شد. این نتایج در جدول ۲ آورده شده است. شکل ۴ نیز نمایی از اثر مقادیر مختلف σ بر متوسط مقدار مساحت و کشیدگی را در شکل‌های مرجع ستاره، قلب، دایره و بیضی نشان می‌دهند. همان‌گونه که مشخص است میزان کشیدگی در اشکال قلب، ستاره و دایره با افزایش σ زیاد و در بیضی کم می‌شود. از آنجایی که نقاط در پیکربندی با توجه به مقدار σ به‌صورت تصادفی در چهار جهت جابجا می‌شوند لذا میزان مساحت نیز به‌صورت تصادفی کم یا زیاد می‌شود که در شکل ۴ این الگوی تصادفی قابل مشاهده است.

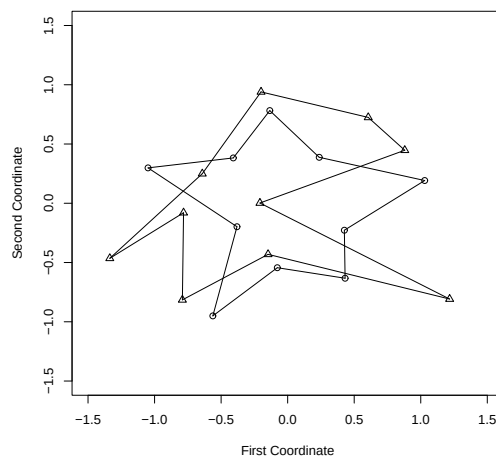
نتایج حاصل از رده‌بندی داده‌ها به این روش را می‌توان با نتایج حاصل از رده‌بندی داده‌های با استفاده از روش گفته شده توسط مقیم‌بیگی (۱۴۰۱) مقایسه کرد. همان‌گونه که مشخص است، روش ارائه شده در این مقاله در مقایسه با روش معرفی شده توسط مقیم‌بیگی (۱۴۰۱) عملکرد بهتری بجز در دو مورد $n = 30, 50$ با $\sigma = 0.5$ ، دارد. در بخش بعد کاربردی از روش ارائه شده در این مقاله در رده‌بندی اشکال ارائه می‌شود.

۴.۱ کاربرد

در این بخش سه مجموعه داده واقعی از دو بسته shapes و Momocs در نرم‌افزار R مورد استفاده قرار گرفته است: اسکیزوفرنی (Schi): برای ثبت این داده‌ها، ابتدا تصویربرداری رزونانس مغناطیسی (MRI) در نزدیکی میدساژیتال^۲

^۱Pseudo-R-Square

^۲Midsagittal



شکل ۲. نمونه‌ای از پیکربندی ستاره، شبیه‌سازی شده با مقادیر $\sigma = 0.1$ (با نقاط شاخص دایره‌ای) و $\sigma = 0.3$ (با نقاط شاخص مثلثی).

از ۱۴ بیمار اسکیزوفرنی و ۱۴ فرد کنترل انجام گرفت. سپس برش‌های دو بعدی از تصاویر در نظر گرفته و ۱۳ نقطه شاخص آن‌ها ثبت شد. این داده‌ها در دو بعد هستند. نمایی از تصویر MRI یک بیمار اسکیزوفرنی به همراه نقاط شاخص ثبت شده از یک بیمار اسکیزوفرنی (مثلث) و یک فرد سالم (دایره) در قسمت الف و ب شکل ۵ نشان داده شده است.

شن (Sand): این داده‌ها شامل ۵۰ نقطه شاخص دو بعدی از ۲۵ دانه شن دریا و ۲۵ دانه شن رودخانه است. مثالی از این داده‌ها در شکل ۵ (ج) نشان داده شده است.

موش (Mouse): این داده‌ها مربوط به ۶۴ نقطه شاخص از سطح بیرونی دندان‌های آسیاب ۳۰ موش درختی است. موش‌های درختی از سه نوع متفاوت بوده که به تعداد ۱۰ نمونه از هر نوع جمع‌آوری شده‌اند. مختصات نقاط شاخص از سطح بیرونی دندان‌های سه نمونه از این سه نوع موش درختی در شکل ۵ (د) نشان داده شده است.

با استفاده از توصیف‌گرهای معرفی شده در بخش ۲ به‌عنوان متغیرهای تبیینی، رده‌بندی داده‌ها انجام گرفت. نتایج حاصل در جدول ۳ (مرحله اول) ارائه شده است. همانگونه که مشاهده می‌شود معیارهای مک‌فادن، کاکس‌اسنیل و ناگلکرک به‌عنوان ضرایب تعیین‌نما مقادیر بزرگی را نشان می‌دهند. همچنین مقدار OCP^1 به‌عنوان معیار دقت، میزان رده‌بندی درست بالایی را نشان می‌دهد. برای مجموعه داده اسکیزوفرنی، میزان دقت به‌دست آمده در روش ارائه شده توسط مقیم‌بیگی (۱۴۰۱) برابر ۸۲ درصد است که این مقدار از میزان دقت ارائه شده در این مقاله کمتر است. بنابراین استفاده از توصیف‌گرها می‌تواند راهکار ساده‌تری در رده‌بندی اشکال در برخی مسائل کاربردی باشد.

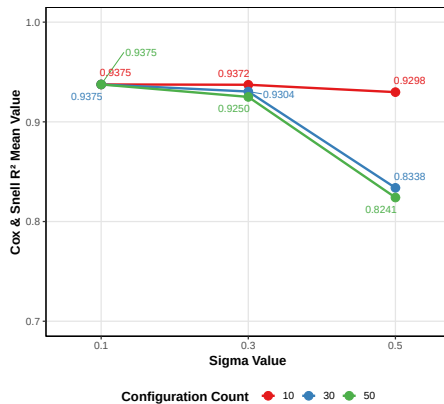
¹Overall correct percent

جدول ۱. میانگین و انحراف معیار ضریب تعیین‌نما و معیارهای برازش مدل (۳).

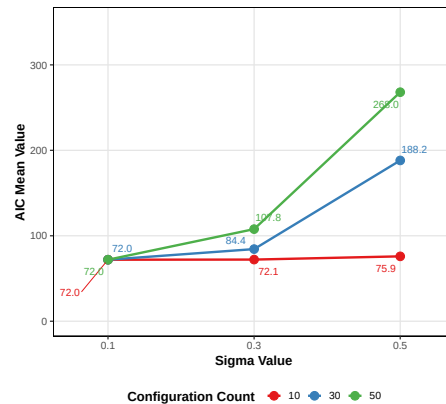
تعداد پیکربندی	σ	برآورد	مک‌فادن	ضریب تعیین‌نما			معیارهای برازش مدل		
				کاکس‌اسنیل	ناگلرک	AIC	BIC	OCP	
۱۰	۰/۱	میانگین	۱	۰/۹۳۷۵	۱	۷۲/۰۰۰۱	۱۰۵/۴۵۰۰	۱۰۰	
		انحراف معیار	۰	۰/۰۰۰۰	۰	۰/۰۰۰۱	۰/۰۰۰۱	۰	
	۰/۳	میانگین	۰/۹۹۸۷	۰/۹۳۷۲	۰/۹۹۹۷	۷۲/۱۴۷۶	۱۰۵/۵۹۷۵	۹۹/۹۵۰۰	
		انحراف معیار	۰/۰۰۹۱	۰/۰۰۱۷	۰/۰۰۱۸	۱/۰۰۷۵	۱/۰۰۷۵	۰/۳۵۱۸	
	۰/۵	میانگین	۰/۹۶۴۵	۰/۹۲۹۸	۰/۹۹۱۸	۷۵/۹۳۴	۱۰۹/۳۸۳۹	۹۸/۱۰۰۰	
		انحراف معیار	۰/۰۶۴۶	۰/۰۱۴۸	۰/۰۱۵۸	۷/۱۶۸	۷/۱۶۸	۳/۸۹۴۴	
۳۰	۰/۱	میانگین	۱	۰/۹۳۷۵	۱	۷۲/۰۰۰۲	۱۰۵/۴۵۰۱	۱۰۰	
		انحراف معیار	۰	۰/۰۰۰۰	۰	۰/۰۰۱۰	۰/۰۰۰۱	۰	
	۰/۳	میانگین	۰/۹۶۲۷	۰/۹۳۰۴	۰/۹۹۲۴	۸۴/۴۱۸۷	۱۱۷/۸۶۸۶	۹۷/۹۹۱۷	
		انحراف معیار	۰/۰۰۳۱۸	۰/۰۰۶۲	۰/۰۰۶۶	۱۰/۵۷۶۹	۱۰/۵۷۶۹	۱/۹۵۲۸	
	۰/۵	میانگین	۰/۶۵۰۸	۰/۸۳۳۸	۰/۸۸۹۴	۱۸۸/۱۸۳۶	۲۲۱/۶۳۳۵	۷۸/۷۴۱۷	
		انحراف معیار	۰/۰۵۰۶	۰/۰۲۳۹	۰/۰۲۵۵	۱۶/۸۲۴۰	۱۶/۸۲۴۰	۳/۸۹۸۸	
۵۰	۰/۱	میانگین	۱	۰/۹۳۷۵	۱	۷۲/۰۰۰۳	۱۰۵/۴۵۰۲	۱۰۰	
		انحراف معیار	۰	۰/۰۰۰۰	۰	۰/۰۰۱۰	۰/۰۰۰۱	۰	
	۰/۳	میانگین	۰/۹۳۵۵	۰/۹۲۵۰	۰/۹۸۶۷	۱۰۷/۷۹۲۰	۱۴۱/۲۴۱۹	۹۶/۵۲۰۰	
		انحراف معیار	۰/۰۰۳۰۰	۰/۰۰۶۲	۰/۰۰۶۶	۱۶/۶۱۷۵	۱۶/۶۱۷۵	۱/۸۵۵۸	
	۰/۵	میانگین	۰/۶۴۶۵	۰/۸۲۴۱	۰/۸۷۹۰	۲۶۸/۰۳۰۱	۳۰۱/۴۸۰۰	۷۸/۹۳۰۰	
		انحراف معیار	۰/۱۳۵۹	۰/۰۴۶۰	۰/۰۴۹۱	۷۵/۳۷۹۳	۷۵/۳۷۹۳	۸/۳۰۹۱	

یکی از چالش‌های مدل‌بندی رگرسیونی داده‌های شکل با استفاده از توصیف‌گرها وجود همخطی است. وجود همبستگی بالا بین توصیف‌گرها لزوم استفاده از توصیف‌گرهای تکراری را زیر سوال می‌برد. لذا در این قسمت به تحلیل همخطی بین توصیف‌گرها برای این سه مجموعه داده پرداخته می‌شود. روش‌های مختلفی برای رفع این مشکل وجود دارد که از جمله می‌توان به حذف متغیرهای تکراری اشاره کرد. در این قسمت و برای این سه مجموعه داده، متغیرهایی که همخطی دارند حذف می‌شوند. تحلیل همخطی برای این مجموعه داده‌ها نشان داد که براساس معیار ضریب تورم واریانس^۱، برای داده‌های اسکیزوفرنی، توصیف‌گرهای طول، عرض و قطر فرت کولیس، برای داده‌های سنگ، اندازه مرکزی، طول و قطر فرت کولیس و برای داده‌های موش، اندازه مرکزی، عرض و قطر فرت کولیس با سایر متغیرها همخطی دارند. بنابراین این متغیرها را برای هر گروه از داده‌ها کنار گذاشته و رده‌بندی داده‌ها را دوباره انجام گرفت. نتایج حاصل از رده‌بندی بعد از حذف متغیرهای اضافی در جدول ۳ (مرحله دوم) قابل مشاهده است. همانگونه که مشخص است نتایج به‌دست آمده تفاوت چندانی با نتایج اولیه (مرحله اول) ندارد هرچند ضرایب تعیین‌نما، AIC و BIC کاهش و میزان انحراف (Dev) افزایش یافته است. همچنین با استفاده از آزمون والد اهمیت نسبی هر یک از توصیف‌گرها نیز مورد ارزیابی قرار گرفت که این نسبت‌ها در شکل ۶ برای سه مجموعه داده نشان داده شده است.

¹Variance Inflation Factor



(ب)



(الف)

شکل ۳. الف- مقادیر متوسط معیارهای AIC و ب- ضریب تعیین نمای کاکس اسنیل.

بحث و نتیجه‌گیری

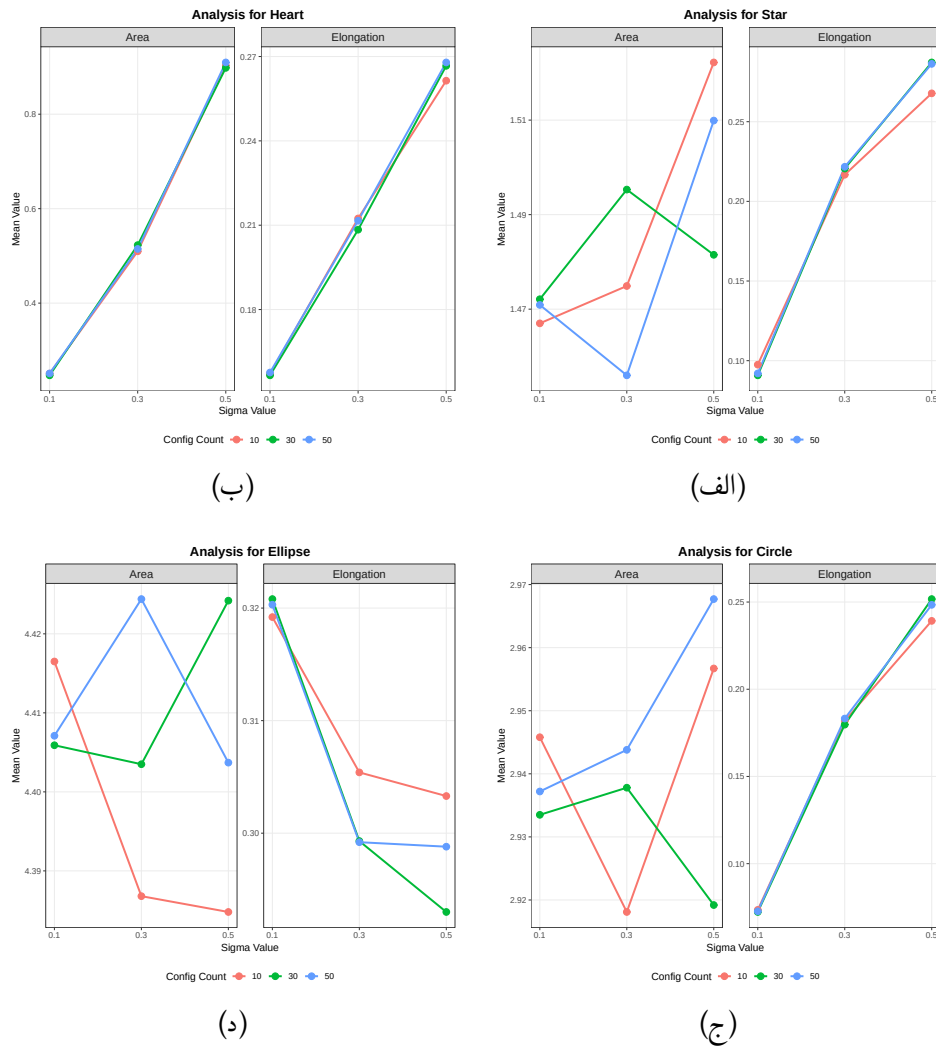
در این مقاله روشی برای رده‌بندی داده‌های شکل با استفاده از مدل رگرسیون لوژستیک چندجمله‌ای و توصیف‌گرهای شکل ارائه شد. یکی از ویژگی‌های این مدل سادگی آن است که از توصیف‌گرهای شکل استفاده می‌کند. توصیف‌گرهای شکل بخش زیادی از اطلاعات درون اشکال را در خود دارند، بنابراین می‌توانند ابزار مناسبی در تحلیل داده‌های شکل باشند. این توصیف‌گرها بیشتر در بینایی ماشین و تحلیل تصویر استفاده می‌شوند. تحلیل انجام گرفته با استفاده از داده‌های شبیه‌سازی شده، نشان داد که عملکرد روش ارائه شده در این مقاله در مقایسه با روش **مقیم‌بیگی** (۱۴۰۱) در بیشتر موارد بهتر بود. همچنین در تحلیل داده‌های اسکیزوفرنی نیز میزان درصد پیش بینی صحیح کلی برابر ۸۹ درصد بود که در مقایسه با روش **مقیم‌بیگی** (۱۴۰۱) مقدار بیشتری را نشان می‌دهد. یکی دیگر از ویژگی‌های روش ارائه شده در این مقاله سادگی آن است. استفاده از ویژگی‌های شکل یک روش مشترک در علوم کامپیوتر و آمار است لذا برای هر دو گروه از محققان قابل درک است.

در این مقاله رده‌بندی داده‌های شکل با استفاده از توصیف‌گرهای شکل در حالت دو بعدی انجام گرفت. یکی از مواردی که می‌توان در تحقیقات آتی مد نظر داشت تعمیم روش گفته شده به حالت سه بعدی است. در حالت سه بعدی نیز می‌توان مدل رگرسیون لوژستیک را برای توصیف‌گرها بکار برد اما باید در نظر داشت که توسعه این روش به ابعاد بالاتر نیازمند تعریف توصیف‌گرهای حجمی جدید مانند سطح رویه، انحنا، گاوسی، حجم و نسبت سطح به حجم است. هرچند توصیف‌گرهای مبتنی بر داده نیز با استفاده از الگوریتم‌های یادگیری ماشین معرفی شده‌اند اما برای استفاده از آن‌ها نیاز به رعایت دستورعمل‌ها و محدودیت‌های خاصی است (**رستمی و همکاران، ۲۰۱۹**). افزایش ابعاد موجب افزایش پیچیدگی محاسباتی و نیاز به الگوریتم‌های بهینه‌سازی کارآمد برای برآورد پارامترها خواهد شد.

جدول ۰.۲. میانگین و انحراف معیار توصیف‌گرهای مساحت و کشیدگی.

توصیف‌گر	تعداد پیکربندی	σ	برآورد	ستاره	قلب	دایره	بیضی
مساحت	۱۰	۰/۱	میانگین	۱/۴۶۷۰	۰/۲۵۰۷	۲/۹۴۵۸	۴/۴۱۶۵
		۰/۳	انحراف معیار	۰/۱۴۵۹	۰/۱۶۱۵	۰/۱۸۶۹	۰/۲۳۶۲
		۰/۳	میانگین	۱/۴۷۴۹	۰/۵۰۹۷	۲/۹۱۸۱	۴/۳۸۶۸
		۰/۵	انحراف معیار	۰/۴۹۳۶	۰/۳۹۳۵	۰/۵۹۴۰	۰/۷۴۲۵
		۰/۵	میانگین	۱/۵۲۲۲	۰/۸۰۳۱	۲/۹۵۶۷	۴/۳۸۴۸
	۳۰	۰/۱	انحراف معیار	۰/۸۸۴۰	۰/۷۱۵۹	۱/۰۳۲۶	۱/۲۸۶۶
		۰/۱	میانگین	۱/۴۷۲۱	۰/۲۴۷۱	۲/۹۳۳۵	۴/۴۰۵۹
		۰/۳	انحراف معیار	۰/۱۴۸۵	۰/۱۶۵۸	۰/۱۹۰۲	۰/۲۳۸۱
		۰/۳	میانگین	۱/۴۹۵۳	۰/۵۲۲۸	۲/۹۳۷۸	۴/۴۰۳۵
		۰/۵	انحراف معیار	۰/۴۷۸۹	۰/۴۰۵۴	۰/۵۸۱۳	۰/۷۴۸۷
کشیدگی	۵۰	۰/۱	میانگین	۱/۴۸۱۵	۰/۸۹۸۱	۲/۹۱۹۲	۴/۴۲۴۲
		۰/۵	انحراف معیار	۰/۸۶۲۵	۰/۶۹۰۴	۱/۰۶۴۱	۱/۲۹۷۲
		۰/۱	میانگین	۱/۴۷۰۹	۰/۲۵۱۱	۲/۹۳۷۲	۴/۴۰۷۱
		۰/۳	انحراف معیار	۰/۱۴۷۸	۰/۱۶۴۳	۰/۱۸۳۷	۰/۲۳۷۶
		۰/۳	میانگین	۱/۴۵۶۰	۰/۵۱۴۷	۲/۹۴۳۸	۴/۴۲۴۴
	۱۰	۰/۵	انحراف معیار	۰/۴۸۳۰	۰/۳۹۷۸	۰/۶۰۱۷	۰/۷۴۴۴
		۰/۵	میانگین	۱/۵۰۹۹	۰/۸۰۹۶	۲/۹۶۷۷	۴/۴۰۳۷
		۰/۵	انحراف معیار	۰/۸۵۷۹	۰/۷۱۰۰	۱/۰۸۵۴	۱/۳۰۶۹
		۰/۱	میانگین	۰/۹۷۵	۰/۱۵۷۱	۰/۰۷۳۵	۰/۳۱۹۲
		۰/۳	انحراف معیار	۰/۰۷۵۶	۰/۰۷۵۸	۰/۰۶۶۹	۰/۰۵۲۳
کشیدگی	۳۰	۰/۳	میانگین	۰/۲۱۶۷	۰/۲۱۲۴	۰/۱۸۱۹	۰/۳۰۵۴
		۰/۵	انحراف معیار	۰/۱۵۱۳	۰/۱۳۹۷	۰/۱۲۸۴	۰/۱۱۶۶
		۰/۵	میانگین	۰/۲۶۷۸	۰/۲۶۱۴	۰/۲۳۹۲	۰/۳۰۳۳
		۰/۵	انحراف معیار	۰/۱۶۳۲	۰/۱۵۹۷	۰/۱۵۳۴	۰/۱۵۳۰
		۰/۱	میانگین	۰/۰۹۰۸	۰/۱۵۶۶	۰/۰۷۲۳	۰/۳۲۰۸
	۵۰	۰/۳	انحراف معیار	۰/۰۷۷۸	۰/۰۷۶۱	۰/۰۶۷۳	۰/۰۴۹۴
		۰/۳	میانگین	۰/۲۲۰۶	۰/۲۰۸۴	۰/۱۷۹۷	۰/۲۹۹۳
		۰/۵	انحراف معیار	۰/۱۴۷۴	۰/۱۴۲۵	۰/۱۲۸۳	۰/۱۲۰۶
		۰/۵	میانگین	۰/۲۸۷۲	۰/۲۶۶۷	۰/۲۵۱۷	۰/۲۹۳۰
		۰/۵	انحراف معیار	۰/۱۶۷۹	۰/۱۵۷۷	۰/۱۵۲۲	۰/۱۵۴۸
	۵۰	۰/۱	میانگین	۰/۰۹۲۰	۰/۱۵۷۶	۰/۰۷۲۷	۰/۳۲۰۳
		۰/۳	انحراف معیار	۰/۰۷۷۷	۰/۰۷۷۱	۰/۰۶۷۱	۰/۰۴۹۴
		۰/۳	میانگین	۰/۲۲۱۷	۰/۲۱۱۶	۰/۱۸۳۲	۰/۲۹۹۲
		۰/۵	انحراف معیار	۰/۱۴۹۰	۰/۱۳۵۸	۰/۱۲۸۱	۰/۱۱۸۹
		۰/۵	میانگین	۰/۲۸۶۳	۰/۲۶۷۹	۰/۲۴۸۴	۰/۲۹۸۸
		۰/۵	انحراف معیار	۰/۱۶۷۶	۰/۱۶۴۳	۰/۱۵۳۲	۰/۱۵۳۸

یکی از چالش‌های مدل‌بندی ارائه شده در این مقاله همخطی بین متغیرها است که افزایش ابعاد ممکن است این همخطی را تشدید کند. این مسئله نیاز به استفاده از روش‌های انتخاب متغیر مانند LASSO یا کاهش ابعاد مانند PCA را ضروری می‌سازد.

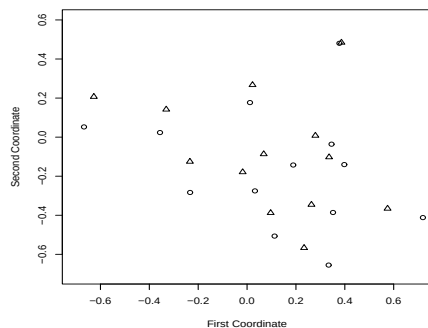


شکل ۴. متوسط مقدار توصیف‌گرهای مساحت و کشیدگی براساس مقادیر مختلف σ در شکل‌های مرجع الف- ستاره، ب- قلب، ج- دایره و د- بیضی.

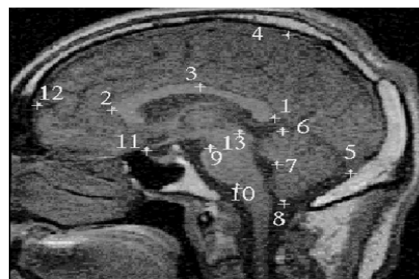
تقدیر و تشکر

نویسنده مقاله، از راهنمایی‌ها، نظرات سازنده و پیشنهادهای ارزشمند داوران گرامی، سردبیر محترم و ویراستار مجله علوم آماری که موجب بهبود و ارتقای کیفیت محتوای علمی این پژوهش شده است، صمیمانه تشکر و قدردانی می‌نماید.

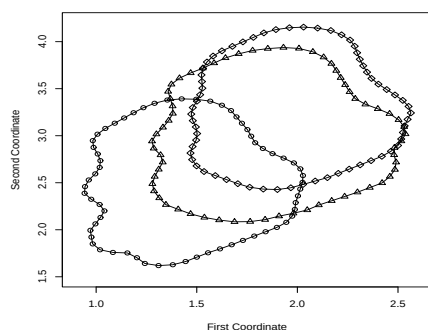
۴۶۴ استفاده از توصیف‌گرهای شکل در رده‌بندی داده‌های شکل



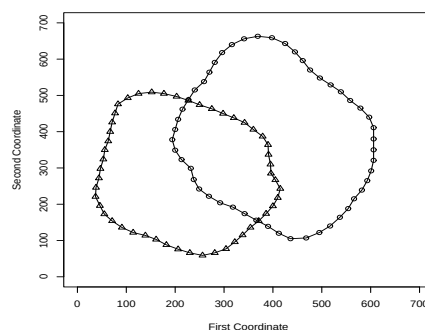
(ب)



(الف)



(د)

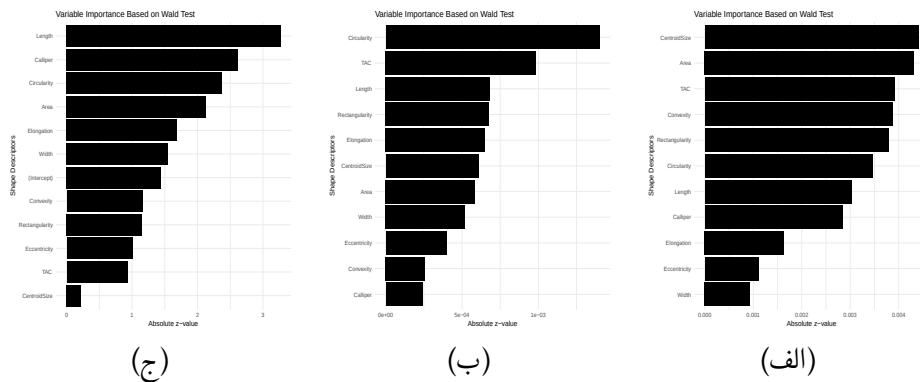


(ج)

شکل ۵. الف- نمایی از تصویر MRI از سر یک بیمار اسکیزوفرنی، ب- مختصات نقاط شاخص بیمار اسکیزوفرنی (مثث) و یک فرد سالم (دایره)، ج- مختصات نقاط شاخص یک دانه شن دریا (دایره) و یک دانه شن رودخانه (مثث)، د- مختصات نقاط شاخص از سطح بیرونی دندان‌های سه نوع موش درختی.

جدول ۳. معیارهای ضریب تعیین نما و برازش مدل رگرسیون لوژستیک به داده‌ها با استفاده از تمام توصیف‌گرها (مرحله اول) و بعد از حذف توصیف‌گرهای اضافی (مرحله دوم).

مرحله	مجموعه داده	ضریب تعیین نما	خلاصه مدل			
			OCP	Dev	BIC	AIC
اول	اسکیزوفرنی	۰/۵۵	۸۹/۲۹	۱۷/۴۸	۵۷/۴۷	۴۱/۴۸
	سنگ	۰/۸۹	۹۵/۹۲	۷/۶۸	۵۴/۳۸	۳۱/۶۸
	موش	۱/۰۰	۱۰۰/۰۰	۰/۰۰	۶۴/۸۲	۴۸/۰۰
دوم	اسکیزوفرنی	۰/۵۰	۸۹/۲۹	۱۹/۵۹	۴۹/۵۸	۳۷/۵۹
	سنگ	۰/۸۴	۹۳/۸۸	۱۰/۶۵	۴۵/۶۸	۲۸/۶۵
	موش	۰/۹۴	۱۰۰/۰۰	۳/۷۳	۵۲/۳۴	۳۹/۷۳



شکل ۶. میزان اهمیت نسبی توصیف‌گرهای مدل برای داده‌های الف- اسکیزوفرنی، ب- سنگ و ج- موش براساس آزمون والد.

مراجع

فتوحی، ح. و گل‌علی‌زاده، م. (۱۳۹۱)، بهبود عملکرد تحلیل ژئودزیک اصلی در تحلیل آمار شکل، مجله علوم آماری، ۲(۶)، ۲۱۹-۲۳۶.

مقیم‌بیگی، م. (۱۴۰۱)، رگرسیون لوژستیک چندجمله‌ای نیمه پارامتری برای رده‌بندی داده‌های شکل، مجله علوم آماری، ۱۶(۲)، ۴۴۹-۴۶۸.

مقیم‌بیگی، م. و گل‌علی‌زاده، م. (۱۳۹۸)، مدل‌بندی رگرسیونی شکل از طریق مثلثی کردن، مجله علوم آماری، ۱۳(۱)، ۱۸۵-۱۹۶.

Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov and F. Csaki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory*.

Bookstein, F. L. (1986), Size and Shape Spaces for Landmark Data in Two Dimensions (With Discussion), *Statistical Science*, 1(2), 181-242.

Chaki, J., and Dey, N. (2020), *Image Color Feature Extraction Techniques: Fundamentals and Applications*, Springer Nature.

Cox, D. R., and Snell, E. J. (1989), *Analysis of Binary Data*, Chapman and Hall/CRC.

- Debavalaere, V., Durrleman, S., and Allasonnière, S. (2020), Learning the Clustering of Longitudinal Shape Data Sets into a Mixture of Independent or Branching Trajectories *International Journal of Computer Vision*, **128**(12), 2794-2809.
- Kendall, D. G. (1977), The Diffusions of Shape, *Advances in Applied Probability*, **9**(3), 428-430.
- Moghimbeygi, M., and Nodehi, A. (2022), Multinomial Principal Component Logistic Regression on Shape Data. *Journal of Classification*, **39**(3), 1-22.
- Rosin, P. L. (2005), Computing Global Shape Measures, *Handbook of Pattern Recognition and Computer Vision* (pp. 177-196),
- Rostami, R., Bashiri, F. S., Rostami, B., and Yu, Z. (2019), A Survey on Data Driven 3D Shape Descriptors. In *Computer graphics forum* (Vol. 38, No. 1, pp. 356-393).
- Russakoff, D. B., Tomasi, C., Rohlfing, T., and Maurer, C. R. (2004), Image Similarity Using Mutual Information of Regions. In *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part III* (Vol. 3022, pp. 596-607), Springer.
- Schwarz, G. (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, **6**, 461-464.
- Simó, A., Ibáñez, M. V., Epifanio, I., and Gimeno, V. (2020), Generalized Partially Linear Models on Riemannian Manifolds, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**(3), 641-661.
- Srivastava, A., Joshi, S. H., Mio, W., and Liu, X. (2005), Statistical Shape Analysis: Clustering, Learning, and Testing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(4), 590-602.
- Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics with S* (4th ed.), Springer.