

DOI 10.54596/2958-0048-2026-1-294-308

УДК 004.896

МРНТИ 28.23.15

**МНОГОМЕРНЫЙ АНАЛИЗ И КЛАССИФИКАЦИЯ ГЕОМЕТРИЧЕСКИХ
ОБЪЕКТОВ ПО СОВОКУПНОСТИ ПАРАМЕТРОВ****Куликова В.П.^{1*}, Чупчиков И.А.¹**¹*НАО «Северо-Казахстанский университет имени Манаша Козыбаева»,
Петропавловск, Казахстан***Автор для корреспонденции: valentina@ku.edu.kz***Аннотация**

В статье рассматривается возможность автоматизированной классификации геометрических объектов на основе многомерного статистического анализа их параметров без применения сложных нелинейных методов машинного обучения. Объекты описываются совокупностью взаимосвязанных геометрических и физических характеристик, включая линейные размеры, площади, объемы, массу и плотность, что приводит к высокой коррелированности признаков и усложняет их интерпретацию.

Для решения поставленной задачи применяются классические методы многомерного анализа данных: корреляционный и регрессионный анализ, метод главных компонент, кластеризация и линейный дискриминантный анализ. Исследование проводится на синтетической выборке, сформированной на основе реалистичных диапазонов параметров, полученных с опорой на реальные геометрические соотношения.

Показано, что предварительное снижение размерности с использованием PCA позволяет устранить мультиколлинеарность признаков и получить компактное, интерпретируемое представление данных. Кластеризация в пространстве главных компонент выявляет устойчивую групповую структуру объектов, а применение линейного дискриминантного анализа обеспечивает высокое качество классификации. Полученные результаты подтверждают, что интерпретируемые статистические методы остаются эффективным инструментом анализа и классификации геометрических объектов и могут использоваться в задачах предварительного анализа данных, поддержки экспертных и интеллектуальных систем в области ИКТ.

Ключевые слова: многомерный анализ; геометрические объекты; PCA; кластеризация; линейный дискриминантный анализ; интерпретируемые модели.

**ПАРАМЕТРЛЕР ЖИЫНТЫҒЫ НЕГІЗІНДЕ ГЕОМЕТРИЯЛЫҚ ОБЪЕКТІЛЕРДІ
КӨП ӨЛШЕМДІ АНАЛИЗ ЖӘНЕ САНАТТАУ****Куликова В.П.^{1*}, Чупчиков И.А.¹**¹*«Манаш Қозыбаев атындағы Солтүстік Қазақстан университеті» КеАҚ,
Петропавл, Қазақстан***Хат-хабар үшін автор: valentina@ku.edu.kz***Аннотация**

Мақалада геометриялық объектілерді күрделі бейсызықтық машиналық оқыту әдістерін қолданбай, олардың параметрлерін көпөлшемді статистикалық талдау негізінде автоматтандырылған жіктеу мүмкіндігі қарастырылады. Объектілер сызықтық өлшемдер, аудандар, көлемдер, масса және тығыздық сияқты өзара байланысты геометриялық және физикалық сипаттамалар жиынтығы арқылы сипатталады, бұл ерекшеліктердің жоғары корреляциясына әкеліп, олардың интерпретациясын қиындатады.

Бұл мәселені шешу үшін көпөлшемді деректерді талдаудың классикалық әдістері қолданылады: корреляциялық және регрессиялық талдау, басты компоненттер әдісі (PCA), кластерлеу және сызықтық

дискриминанттық талдау (LDA). Зерттеу нақты геометриялық қатынастарға негізделген параметрлердің шынайы диапазондары бойынша қалыптасқан синтетикалық деректер жиынтығында жүргізіледі.

РСА қолдану арқылы алдын ала өлшемді азайту ерекшеліктер арасындағы мультиколлинеарлықты жоюға және деректердің ықшам, интерпретацияланатын бейнесін алуға мүмкіндік беретіні көрсетілген. Басты компоненттер кеңістігінде кластерлеу объектілердің тұрақты топтық құрылымын анықтайды, ал сызықтық дискриминанттық талдауды қолдану классификацияның жоғары сапасын қамтамасыз етеді. Алынған нәтижелер интерпретацияланатын статистикалық әдістер геометриялық объектілерді талдау және классификациялау үшін тиімді құрал болып қала беретінін және оларды деректерді алдын ала талдау, эксперттік жүйелерді қолдау және ЗКТ негізіндегі интеллектуалды жүйелерді дамыту тапсырмаларында қолдануға болатынын растайды.

Кілт сөздер: көпөлшемді анализ; геометриялық объектілер; РСА; кластерлеу; сызықтық дискриминанттық талдау; интерпретацияланатын модельдер.

MULTIVARIATE ANALYSIS AND CLASSIFICATION OF GEOMETRIC OBJECTS BASED ON A SET OF PARAMETERS

Kulikova V.P.^{1*}, Chupchikov I.A.¹

^{1*}Manash Kozybayev North Kazakhstan University NPLC, Petropavlovsk, Kazakhstan

*Corresponding author: valentina@ku.edu.kz

Abstract

This study examines the feasibility of automated classification of geometric objects using multivariate statistical analysis of their parameters, without relying on complex nonlinear machine learning techniques. Objects are described by a set of interrelated geometric and physical characteristics, including linear dimensions, areas, volumes, mass, and density, which leads to high feature correlation and complicates their interpretation.

To address this task, classical methods of multivariate data analysis are applied: correlation and regression analysis, principal component analysis (PCA), clustering, and linear discriminant analysis (LDA). The study is conducted on a synthetic dataset formed based on realistic parameter ranges derived from actual geometric relationships.

It is demonstrated that preliminary dimensionality reduction using PCA can eliminate multicollinearity among features and produce a compact, interpretable representation of the data. Clustering in the principal component space reveals a stable group structure of objects, while the application of linear discriminant analysis ensures high classification quality. The results confirm that interpretable statistical methods remain an effective tool for analyzing and classifying geometric objects and can be used in tasks of preliminary data analysis, support of expert systems, and development of intelligent ICT-based systems.

Keywords: multivariate analysis; geometric objects; PCA; clustering; linear discriminant analysis; interpretable models.

Введение

Задачи анализа и классификации геометрических объектов возникают в различных областях информационно-коммуникационных технологий, включая компьютерное зрение, обработку трехмерных данных, системы автоматизированного проектирования и интеллектуальные системы поддержки принятия решений. Геометрические объекты, как правило, описываются набором взаимосвязанных параметров, отражающих их размеры, форму и физические свойства. Такая параметризация приводит к высокой размерности признакового пространства и выраженной коррелированности признаков, что затрудняет прямое применение методов классификации и интерпретацию полученных результатов [1-4].

В современных исследованиях для решения подобных задач часто используются сложные методы машинного обучения, однако их применение сопровождается снижением интерпретируемости моделей и повышенными требованиями к объему данных. В то же время классические методы многомерного статистического анализа сохраняют актуальность благодаря своей прозрачности, устойчивости и возможности анализа структуры данных [5-7].

Целью работы является выявление информативных признаков и латентных факторов, обеспечивающих устойчивое семантически значимое разделение геометрических объектов, а также разработка интерпретируемой процедуры их автоматизированной классификации в условиях высокой коррелированности признаков и ограниченности наблюдаемых данных. Особое внимание уделяется анализу структуры данных, снижению размерности и интерпретации полученных компонент и классификационных правил.

Для достижения поставленной цели необходимо:

1. исследовать корреляционную структуру параметров геометрических объектов и выявить избыточные и взаимосвязанные признаки;
2. проанализировать возможность восстановления интегральных и латентных характеристик объектов на основе ограниченного набора наблюдаемых параметров с использованием регрессионного анализа;
3. выполнить снижение размерности признакового пространства с применением метода главных компонент и интерпретировать полученные компоненты;
4. исследовать устойчивость группирования объектов с использованием методов кластерного анализа в пространстве главных компонент;
5. оценить качество автоматизированной классификации геометрических объектов с применением линейного дискриминантного анализа и интерпретировать полученные классификационные правила.

Научная новизна работы заключается в разработке интерпретируемого подхода к анализу геометрических объектов, основанного на выявлении латентных факторов, формирующих их структурные свойства, и оценке устойчивости семантически значимого разделения объектов в условиях высокой коррелированности признаков и ограниченности наблюдаемых данных. В отличие от традиционного применения отдельных методов, предложенный подход рассматривает их как взаимосвязанную аналитическую процедуру, направленную на выделение информативных признаков и обеспечение воспроизводимого группирования объектов.

Практическая значимость результатов состоит в возможности использования предложенного подхода в задачах предварительного анализа данных, поддержки экспертных систем и разработки интеллектуальных модулей обработки геометрической информации.

Расчеты и визуализация выполнялись с использованием языка Python и стандартных библиотек анализа данных.

Материалы и методы

Для решения поставленной задачи в работе используются классические методы многомерного статистического анализа, широко применяемые в задачах анализа данных и распознавания образов [1-4, 8]. Применение данных методов позволяет исследовать структуру многомерных данных, выявлять скрытые зависимости между параметрами и формировать интерпретируемые признаки для последующей классификации объектов.

В работе основное внимание уделяется не выводу известных формул, а обоснованию их применимости для анализа высококоррелированных геометрических параметров и выявления информативных признаков, обеспечивающих устойчивое разделение объектов. Приводимые математические выражения используются как компактное описание используемого аппарата.

В рамках работы сознательно не рассматриваются методы глубокого обучения и сложные ансамблевые алгоритмы, что позволяет сосредоточиться на анализе структуры данных и интерпретируемости результатов.

Корреляционный анализ

Корреляционный анализ применяется для выявления линейных зависимостей между параметрами геометрических объектов и оценки степени их взаимосвязи. Наличие высоких корреляций между признаками указывает на избыточность параметров и потенциальную мультиколлинеарность, что может негативно сказываться на устойчивости статистических моделей [2, 3, 5].

Поскольку анализируемые геометрические параметры имеют количественную природу и допускают линейную интерпретацию, для оценки зависимости используется коэффициент корреляции Пирсона:

$$r_{ij} = \frac{\text{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} \quad (1)$$

где x_i, x_j – анализируемые признаки, $\sigma_{x_i}, \sigma_{x_j}$ – их стандартные отклонения.

Высокие значения $|r_{ij}|$ свидетельствуют о наличии выраженной линейной зависимости и указывают на необходимость снижения размерности признакового пространства.

В контексте рассматриваемой задачи корреляционный анализ выступает инструментом выявления избыточности геометрических параметров и обоснования необходимости снижения размерности признакового пространства.

Регрессионный анализ

Для оценки влияния геометрических параметров на интегральные характеристики объектов используется линейный регрессионный анализ. В общем виде модель представляется следующим образом:

$$y = \beta_0 + \sum_1^n \beta_i x_i + \varepsilon \quad (2)$$

где y – целевая переменная, x_i – геометрические параметры, β_i – коэффициенты регрессии, ε – случайная ошибка.

Анализ коэффициентов регрессии позволяет оценить вклад отдельных параметров в формирование целевой характеристики и определить их статистическую значимость [2, 5].

В рамках настоящей работы регрессионный анализ используется не как инструмент прогнозирования, а как средство оценки восстановимости латентных характеристик объекта по ограниченному набору наблюдаемых признаков, аналогичный задаче реконструкции признаков формы и структуры в системах визуального распознавания.

Метод главных компонент

Метод главных компонент (Principal Component Analysis, PCA) применяется для снижения размерности признакового пространства за счет перехода к новому набору ортогональных компонент, упорядоченных по доле объясненной дисперсии. Использование PCA позволяет устранить мультиколлинеарность исходных признаков и

получить компактное представление данных при сохранении основной информации [4, 5, 9].

Переход к пространству главных компонент осуществляется по формуле:

$$\mathbf{z} = \mathbf{W}^T(\mathbf{x} - \bar{\mathbf{x}}) \quad (3)$$

где \mathbf{W} – матрица собственных векторов ковариационной матрицы исходных данных \mathbf{x} .

Доля объясненной дисперсии k -й компоненты определяется выражением:

$$\eta_k = \frac{\lambda_k}{\sum_1^n \lambda_k} \quad (4)$$

где λ_k – соответствующее собственное значение.

Данный показатель используется для количественной оценки информативности компонент и выбора их числа.

Применение PCA в данной задаче позволяет перейти от физически зависимых параметров к латентным факторам, отражающим масштабные и структурные свойства объектов.

Линейный дискриминантный анализ

Для классификации объектов применяется линейный дискриминантный анализ (Linear Discriminant Analysis, LDA), основанный на максимизации критерия Фишера [7, 10, 14, 15]:

$$J(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}} \quad (5)$$

где \mathbf{S}_B и \mathbf{S}_W – межклассовая и внутриклассовая матрицы рассеяния.

Максимизация данного критерия обеспечивает наилучшее линейное разделение классов в пространстве признаков. Преимуществом LDA является его интерпретируемость и устойчивость при ограниченном объеме данных.

Использование LDA оправдано необходимостью получения интерпретируемых линейных границ разделения, позволяющих анализировать вклад признаков в семантическое различие групп объектов.

Кластерный анализ

Кластерный анализ используется для выявления групп геометрических объектов, обладающих сходными характеристиками в пространстве признаков или главных компонент. В рамках настоящего исследования применяется алгоритм k-means, ориентированный на минимизацию внутрикластерной дисперсии и широко используемый в задачах анализа многомерных данных при ограниченном объеме выборки.

Выбор числа кластеров осуществляется на основе метода локтя, позволяющего оценить зависимость суммарной внутрикластерной дисперсии от числа кластеров, а также коэффициента силуэта, характеризующего степень разделимости кластеров. Совместное использование указанных критериев позволяет обоснованно определить число кластеров и оценить устойчивость результатов кластеризации [4, 9, 12, 13].

Для повышения устойчивости группирования кластерный анализ выполняется в пространстве главных компонент, что позволяет снизить влияние мультиколлинеарности исходных признаков и повысить интерпретируемость полученных кластеров.

В рассматриваемой задаче кластеризация используется как инструмент выявления естественной групповой структуры объектов, формируемой совокупностью геометрических характеристик [9, 13].

Описание выборки

В качестве объекта исследования используется синтетическая выборка, состоящая из 250 геометрических объектов, сформированная на основе параметров реальных геометрических форм и их физических характеристик. Каждый объект описывается набором параметров (таблица 1), характеризующих его геометрические и физические свойства, включая линейные размеры, площадь, объем, массу и плотность.

Таблица 1. Диапазоны значений параметров геометрических объектов (фрагмент)

<i>Параметр</i>	<i>Миним.</i>	<i>Максим.</i>	<i>Комментарий</i>
Длина (Length), ед.	0.5	5.0	Линейный размер.
Ширина (Width), ед.	0.5	5.0	Диаметр основания.
Высота (Height), ед.	0.5	8.0	Варьирование пропорций.
Площадь основания (Base Area), ед ²	0.2	78.5	Геометрическая формула.
Общая площадь поверхности (Total Area), ед ²	1.5	200.0	Суммарная площадь.
Объем (Volume), ед ³	0.1	500.0	Производная характеристика.
Плотность материала (Density), усл. ед.	0.5	5.0	Случайное распределение.
Масса (Mass), усл. ед.	0.1	2500.0	Расчет Volume×Density

На основе исходных параметров были дополнительно рассчитаны производные характеристики, что позволило сформировать целостное и согласованное признаковое пространство для последующего многомерного статистического анализа. Использование синтетически расширенной выборки, сформированной на базе реалистичных диапазонов параметров, обеспечивает контроль условий эксперимента, воспроизводимость результатов и возможность систематического исследования влияния отдельных признаков на структуру данных. Для проверки реалистичности выборки были проанализированы распределения значений признаков и их диапазоны вариации. Полученные распределения соответствуют ожидаемым закономерностям и не содержат аномальных значений, что позволяет использовать выборку для дальнейшего статистического анализа и моделирования. Диапазоны варьирования параметров выбирались с учетом физических и геометрических ограничений, характерных для рассматриваемых типов объектов [3, 6, 13, 14].

Такой подход снижает влияние случайных ошибок измерений, характерных для реальных данных на ранних этапах исследования, и обеспечивает корректность интерпретации результатов.

Корреляционный и регрессионный анализ

Анализ корреляционной матрицы (рисунок 1) показал наличие сильных линейных зависимостей между рядом геометрических параметров, в частности между линейными размерами, площадью и объемом, что указывает на мультиколлинеарность признаков. Наличие мультиколлинеарности может негативно сказываться на устойчивости моделей

классификации и интерпретации результатов, в связи с чем дальнейший анализ выполнялся с использованием методов снижения размерности.

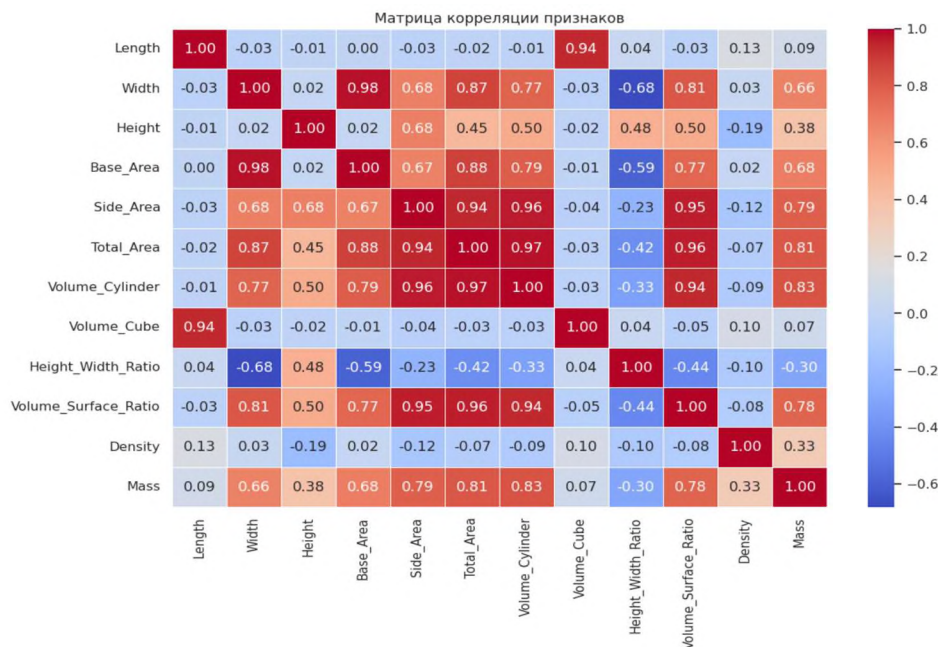


Рисунок 1. Корреляционная матрица параметров

Регрессионный анализ подтвердил статистическую значимость основных геометрических параметров [2, 4, 5, 14]. Фрагмент анализа регрессионной модели приведен на рисунке 2 и используется для демонстрации восстановления латентного интегрального свойства объекта по ограниченному набору наблюдаемых геометрических признаков, что методологически соответствует задачам восстановления скрытых характеристик объекта при анализе визуальных данных.

OLS Regression Results						
Dep. Variable:	Mass	R-squared:	0.851			
Model:	OLS	Adj. R-squared:	0.845			
Method:	Least Squares	F-statistic:	152.1			
Date:	Sun, 16 Mar 2025	Prob (F-statistic):	6.94e-94			
Time:	19:12:59	Log-Likelihood:	-2865.2			
No. Observations:	250	AIC:	5750.			
Df Residuals:	240	BIC:	5786.			
Df Model:	9					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-4.782e+04	1.53e+04	-3.131	0.002	-7.79e+04	-1.77e+04
Length	184.0126	114.498	1.607	0.109	-41.536	409.562
Width	-1048.4999	1352.656	-0.775	0.439	-3713.094	1616.094
Height	437.9615	1053.898	0.416	0.678	-1638.110	2514.033
Base_Area	8.0507	8.892	0.905	0.366	-9.465	25.566
Side_Area	-7.0733	14.393	-0.491	0.624	-35.426	21.279
Total_Area	9.0281	11.633	0.776	0.438	-13.887	31.944
Volume_Cylinder	2.1114	1.155	1.828	0.069	-0.164	4.386
Height_Width_Ratio	-2962.8769	5075.790	-0.584	0.560	-1.3e+04	7035.909
Volume_Surface_Ratio	-215.8370	5449.421	-0.040	0.968	-1.1e+04	1.05e+04
Density	1.876e+04	1202.774	15.598	0.000	1.64e+04	2.11e+04
Omnibus:	73.202	Durbin-Watson:	1.947			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	628.118			
Skew:	0.873	Prob(JB):	4.04e-137			
Kurtosis:	10.566	Cond. No.	8.72e+16			

Рисунок 2. Анализ регрессионной модели массы объекта с предикторами – геометрическими характеристиками

PCA и интерпретация компонент

Для снижения размерности признакового пространства был применен метод главных компонент [4, 5, 14]. Анализ объясненной дисперсии показал, что первые две главные компоненты аккумулируют значительную часть информации, содержащейся в исходных данных (рисунок 3). Исследование факторных нагрузок (рисунок 4) позволило интерпретировать полученные компоненты. Первая главная компонента в наибольшей степени отражает масштабные характеристики объектов, тогда как вторая компонента связана с плотностными и относительными параметрами.

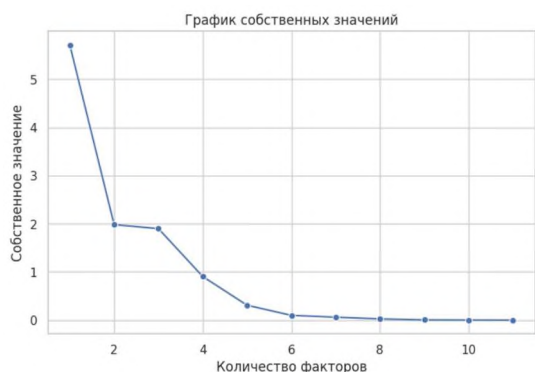


Рисунок 3. График распределения объясненной дисперсии

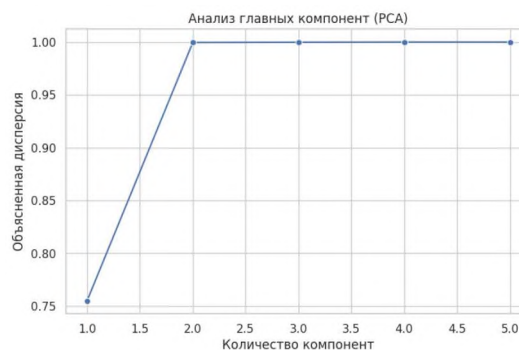


Рисунок 4. Факторные нагрузки главных компонент

Таким образом, применение PCA позволяет перейти от исходного набора коррелированных признаков к компактному и интерпретируемому представлению

данных, удобному для последующих этапов анализа. Кроме того, PCA выступает не только инструментом снижения размерности, но и средством выявления латентных структурных факторов, определяющих различия между геометрическими объектами.

Кластеризация и визуализация

Кластеризация геометрических объектов выполнялась в пространстве главных компонент с использованием алгоритма k-means. Для выбора оптимального числа кластеров применялись метод локтя и коэффициент силуэта.

Результаты анализа показали, что выбранное число кластеров обеспечивает устойчивое группирование объектов с близкими геометрическими характеристиками. Визуализация кластеров в пространстве двух главных компонент демонстрирует четкое разделение групп и позволяет наглядно интерпретировать структуру данных, что подтверждает существование устойчивой латентной структуры признакового пространства.

Использование PCA перед кластеризацией способствует повышению устойчивости группирования за счет устранения избыточных и коррелированных признаков [4, 12, 14].

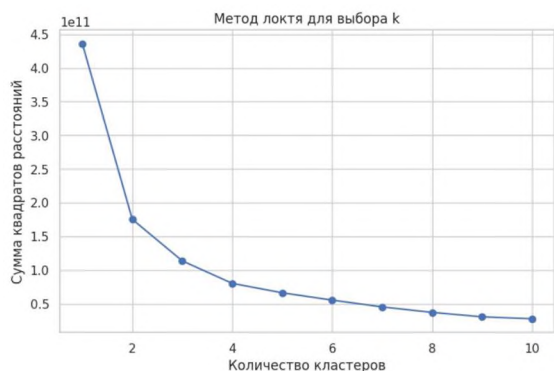


Рисунок 5. Выбор числа кластеров методом локтя

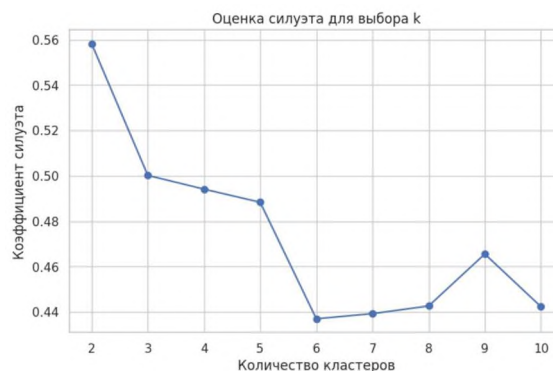


Рисунок 6. Оценка качества кластеризации по коэффициенту силуэта

Кластеризация методом k-means показала устойчивое группирование объектов с близкими геометрическими характеристиками (рисунок 7).



Рисунок 7. Кластеры объектов в пространстве главных компонент

Полученные результаты подтверждают, что предварительное снижение размерности способствует повышению устойчивости кластеризации за счет устранения избыточных признаков.

LDA и качество классификации

Для классификации геометрических объектов по заданным классам была обучена модель линейного дискриминантного анализа (LDA) для автоматизированной классификации объектов [7, 10, 14]. Качество классификации оценивалось с использованием стандартных метрик (таблица 2), а также показателя точности:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

где TP, TN, FP, FN – количества истинно положительных, истинно отрицательных, ложно положительных и ложно отрицательных классификаций соответственно.

Таблица 2. Результаты метрического исследования классификации

Класс	Precision	Recall	F1-score	Support
0	0.87	1.00	0.93	41
1	1.00	0.89	0.94	18
2	0.92	0.69	0.79	16

В качестве входных данных использовались все ранее рассчитанные признаки, включая главные компоненты PCA.



Рисунок 8. Матрица ошибок

Разделение данных на обучающую и тестовую выборки (70/30) позволило объективно оценить качество модели. Итоговая точность классификации составила 90.67%, что подтверждается значениями метрик качества (таблица 2) и соответствует хорошему качеству классификации для линейных моделей при ограниченном объеме данных. При этом полученные дискриминантные функции сохраняют интерпретируемость, позволяя анализировать вклад признаков в формирование семантических различий между группами объектов.

Анализ матрицы ошибок (рисунок 8) показал, что наибольшее число неверных классификаций связано с объектами, имеющими близкие значения параметров, что соответствует ожидаемым свойствам линейных моделей.

Обсуждение результатов

Полученные результаты показывают, что применяемые методы обеспечивают достижение поставленной цели исследования – выявление информативных признаков и анализ факторов, обеспечивающих устойчивое разделение геометрических объектов на семантически значимые группы. Существенным результатом работы является демонстрация возможности восстановления и интерпретации скрытых (латентных) характеристик объекта на основе ограниченного набора наблюдаемых признаков, формируемых при неполном и фрагментарном описании объекта, обладающих высокой коррелированностью.

Корреляционный анализ показал наличие выраженных линейных зависимостей между геометрическими параметрами, такими как линейные размеры, площади и объемы. Данная особенность характерна для многих задач анализа сложных объектов, в том числе в компьютерном зрении, где наблюдаемые признаки, извлекаемые из изображений, оказываются статистически зависимыми и избыточными. Это обосновывает необходимость применения методов снижения размерности и выделения информативных компонент.

Использование регрессионного анализа в работе следует рассматривать не как решение прикладной задачи прогнозирования физической величины, а как иллюстрацию возможности восстановления интегрального латентного свойства объекта по совокупности наблюдаемых геометрических признаков. В этом контексте масса объекта выступает в роли абстрактной скрытой характеристики, аналогичной неявным признакам формы, объема или структуры, которые в задачах компьютерного зрения восстанавливаются по ограниченному числу визуальных наблюдений. Таким образом, данный пример следует рассматривать как абстрактную модель задачи восстановления скрытых свойств объекта в условиях ограниченного числа наблюдений, а не как анализ конкретной физической величины.

Применение метода главных компонент позволило перейти от исходного набора коррелированных признаков к компактному пространству ортогональных компонент,

отражающих основные скрытые факторы, определяющие структуру объекта. Анализ факторных нагрузок показал, что полученные компоненты допускают содержательную интерпретацию и могут рассматриваться как латентные признаки формы и структуры объекта. Такой подход концептуально аналогичен восстановлению внутренних характеристик объекта по ограниченному числу изображений, полученных с различных ракурсов и обладающих неполным покрытием формы объекта, что имеет принципиальное значение для задач визуального распознавания при дефиците данных.

Кластеризация объектов, выполненная в пространстве главных компонент, продемонстрировала устойчивое разделение выборки на группы с близкими геометрическими свойствами. Использование метода локтя и коэффициента силуэта позволило обоснованно определить число кластеров и оценить качество группирования. Полученные результаты показывают, что предварительное снижение размерности способствует повышению устойчивости кластеризации за счет устранения избыточных и коррелированных признаков.

Линейный дискриминантный анализ, примененный для классификации геометрических объектов, продемонстрировал высокие значения точности и сбалансированных метрик качества. Это свидетельствует о возможности эффективного использования линейных классификаторов при условии корректного отбора и предварительной обработки признаков. Существенным преимуществом данного подхода является прозрачность модели, позволяющая анализировать вклад отдельных параметров в процесс принятия решения и обеспечивать интерпретируемость результатов классификации.

Таким образом, результаты работы демонстрируют универсальность интерпретируемых статистических методов для анализа объектов в условиях ограниченности наблюдаемых данных и подтверждают возможность их использования в качестве базового этапа при построении систем распознавания, ориентированных на минимизацию числа входных наблюдений (в том числе изображений, полученных с различных ракурсов) без существенной потери информации о структуре объекта.

Вместе с тем следует отметить ряд ограничений проведенного исследования. Использование синтетической выборки, обеспечивая контроль диапазонов параметров и воспроизводимость экспериментов, не учитывает возможные шумы и искажения, характерные для данных реальных измерений. Кроме того, применяемые методы основаны на линейных предположениях, что может ограничивать их применимость при анализе объектов со сложной нелинейной структурой признаков.

В качестве направлений дальнейших исследований представляется целесообразным расширение анализа за счет использования данных реальных измерений, а также увеличение объема выборки, сформированной на их основе, с целью проверки устойчивости полученных результатов, проведение анализа ошибок классификации для выявления пограничных объектов, а также сопоставление интерпретируемых статистических методов с нелинейными моделями машинного обучения [3, 6, 14, 16].

Отдельный интерес представляет интеграция полученных результатов в состав экспертных и гибридных интеллектуальных систем, где предварительная статистическая обработка данных может использоваться для повышения устойчивости и объяснимости принимаемых решений.

Тем самым подтверждается, что устойчивое разделение объектов определяется не отдельными параметрами, а совокупностью латентных признаков, выявляемых методами многомерного анализа.

Заключение

В работе предложен интерпретируемый подход к анализу и автоматизированной классификации геометрических объектов на основе многомерного статистического анализа их параметров без использования сложных методов машинного обучения [1-16]. В ходе исследования решены задачи анализа корреляционной структуры данных, восстановления латентных характеристик объектов, снижения размерности и устойчивого кластерного разделения объектов.

Проведенный анализ подтверждает применимость рассматриваемого подхода к задачам визуального распознавания объектов, в которых латентные признаки формы и структуры восстанавливаются по ограниченному числу изображений, полученных с различных ракурсов и при различном качестве наблюдений.

Предложенный подход может быть использован при предварительном анализе геометрических данных, в системах поддержки экспертных решений и при разработке интеллектуальных модулей обработки информации в области ИКТ. В качестве направлений дальнейших исследований рассматриваются использование данных реальных измерений и интеграция статистических методов с более сложными моделями анализа.

Полученные результаты расширяют возможности применения интерпретируемых статистических методов в задачах анализа геометрических данных и могут служить методологической основой для построения объяснимых систем распознавания.

Литература:

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях: учебник для вузов. – М.: ЮНИТИ ДАНА, 2001. – 270 с. – ISBN 5-238-00303-X.
2. Дрейпер Н.Р., Смит Г. Прикладной регрессионный анализ. – М.: Диалектика, 2017. – 912 с. – ISBN 978-5-8459-0963-3.
3. Федоров В.В. Методы многомерного статистического анализа. – СПб.: Питер, 2018. – 504 с. – ISBN 978-5-44610-000-0.
4. Guo L., Qin Y. An exploration of the application of principal component analysis in big data processing // *Applied Mathematics and Nonlinear Sciences*. – 2024. – Vol. 9, No. 1. – P. 1–24. – DOI: <https://doi.org/10.2478/amns-2024-0664>
5. Jolliffe I.T. *Principal Component Analysis*. – 2-е изд. – New York: Springer, 2016. – 487 p. – DOI: <https://doi.org/10.1007/9781493944708>
6. Николенко С.И. *Машинное обучение: основы*. – СПб.; М.; Минск: Питер, 2025. – 608 с. – ISBN 978-5-4461-4191-3.
7. Gyamerah S., Soori G.T., Korda D.R., Tawiah J.K., Akolgo E.A., Dapaah E.O. Comparative analysis of feature extraction for high dimensional data reduction using machine learning techniques // *American Journal of Electrical and Computer Engineering*. – 2023. – Vol. 7, No. 2. – P. 27–39. – DOI: <https://doi.org/10.11648/j.ajece.20230702.12>
8. Mathivanan N.M.N., Md. Ghani N.A., Janor R.M. A comparative study on dimensionality reduction between principal component analysis and K means clustering // *Indonesian Journal of Electrical Engineering and Computer Science*. – 2023. – Vol. 16, No. 2. – P. 752–758. – DOI: <https://doi.org/10.11591/ijeecs.v16.i2.pp752-758>
9. Ma Z., López Oriona Á., Ombao H., Sun Y. ROBPCPA: robust multivariate time series clustering method based on common principal component analysis // *Journal of Classification*. – 2026. – Vol. 43, No. 1. – DOI: <https://doi.org/10.1007/s00357-026-09536-7>

10. Oshternian S.R., Loipfinger S., Bhattacharya A., Fehrmann R.S.N. Exploring combinations of dimensionality reduction, transfer learning, and regularization methods for predicting binary phenotypes with transcriptomic data // *BMC Bioinformatics*. – 2024. – Vol. 25, Art. 167. – DOI: <https://doi.org/10.1186/s12859-024-05795-6>
11. Fisher R.A. The use of multiple measurements in taxonomic problems // *Annals of Eugenics*. – 1936. – Vol. 7. – P. 179-188.
12. Sinaga B. Multivariate data analysis for customer segmentation using principal component analysis and K means clustering // *Jurnal Info Sains: Informatika dan Sains*. – 2024. – Vol. 3. – P. 1-10.
13. Cheng G., Lin R., Peng L. High dimensional multivariate analysis of variance via geometric median and bootstrapping // *Biometrics*. – 2024. – Vol. 80, No. 3, ujae088. – DOI: <https://doi.org/10.1093/biomtc/ujae088>
14. Shen Z. Comparison and evaluation of classical dimensionality reduction methods // *Highlights in Science, Engineering and Technology*. – 2023. – Vol. 70. – P. 411-418. – DOI: <https://doi.org/10.54097/hset.v70i.13890>
15. Qu L., Pei Y. A comprehensive review on discriminant analysis for addressing challenges of class level limitations, small sample size, and robustness // *Processes*. – 2024. – Vol. 12, No. 7, Art. 1382. – DOI: <https://doi.org/10.3390/pr12071382>
16. Muñoz Pichardo J.M., Pino Mejías R., Cubiles de la Vega M.D., Enguix González A. Dimensionality reduction through clustering of variables and canonical correlation // *Journal of the Korean Statistical Society*. – 2024. – Vol. 54, No. 1. – P. 63-90. – DOI: <https://doi.org/10.1007/s42952-024-00290-3>

References:

1. Aivazyan, S. A., & Mkhitarian, V. S. (2001). *Prikladnaya statistika v zadachakh i uprazhneniyakh: Uchebnik dlya VUZov*. Moscow: UNITY DANA.
2. Draper, N. R., & Smith, G. (2017). *Prikladnoy regressiynnyy analiz* (3rd ed.). Moscow: Dialektika.
3. Fedorov, V. V. (2018). *Metody mnogomernogo statisticheskogo analiza*. St. Petersburg: Piter.
4. Guo, L., & Qin, Y. (2024). An exploration of the application of principal component analysis in big data processing. *Applied Mathematics and Nonlinear Sciences*, 9(1), 1-24. <https://doi.org/10.2478/amns-2024-0664>
5. Jolliffe, I. T. (2016). *Principal Component Analysis* (2nd ed.). New York, NY: Springer. <https://doi.org/10.1007/9781493944708>
6. Nikolenko, S. I. (2025). *Mashinnoye obucheniye: osnovy*. Saint Petersburg; Moscow; Minsk: Piter.
7. Gyamerah, S., Soori, G. T., Korda, D. R., Tawiah, J. K., Akolgo, E. A., & Dapaah, E. O. (2023). Comparative analysis of feature extraction for high dimensional data reduction using machine learning techniques. *American Journal of Electrical and Computer Engineering*, 7(2), 27-39. <https://doi.org/10.11648/j.ajece.20230702.12>
8. Mathivanan, N. M. N., Md. Ghani, N. A., & Janor, R. M. (2023). A comparative study on dimensionality reduction between principal component analysis and K means clustering. *Indonesian Journal of Electrical Engineering and Computer Science*, 16(2), 752-758. <https://doi.org/10.11591/ijeecs.v16.i2.pp752-758>
9. Ma, Z., López Oriona, Á., Ombao, H., & Sun, Y. (2026). ROBPCPCA: robust multivariate time series clustering method based on common principal component analysis. *Journal of Classification*, 43(1). <https://doi.org/10.1007/s00357-026-09536-7>
10. Oshternian, S. R., Loipfinger, S., Bhattacharya, A., & Fehrmann, R. S. N. (2024). Exploring combinations of dimensionality reduction, transfer learning, and regularization methods for predicting binary phenotypes with transcriptomic data. *BMC Bioinformatics*, 25, 167. <https://doi.org/10.1186/s12859-024-05795-6>
11. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
12. Sinaga, B. (2024). Multivariate data analysis for customer segmentation using principal component analysis and K means clustering. *Jurnal Info Sains: Informatika dan Sains*, 3, 1-10.
13. Cheng, G., Lin, R., & Peng, L. (2024). High dimensional multivariate analysis of variance via geometric median and bootstrapping. *Biometrics*, 80(3), ujae088. <https://doi.org/10.1093/biomtc/ujae088>
14. Shen, Z. (2023). Comparison and evaluation of classical dimensionality reduction methods. *Highlights in Science, Engineering and Technology*, 70, 411-418. <https://doi.org/10.54097/hset.v70i.13890>

15. Qu, L., & Pei, Y. (2024). A comprehensive review on discriminant analysis for addressing challenges of class level limitations, small sample size, and robustness. *Processes*, 12(7), 1382. <https://doi.org/10.3390/pr12071382>
16. Muñoz Pichardo, J. M., Pino Mejías, R., Cubiles de la Vega, M. D., & Enguix González, A. (2024). Dimensionality reduction through clustering of variables and canonical correlation. *Journal of the Korean Statistical Society*, 54(1), 63-90. <https://doi.org/10.1007/s42952-024-00290-3>

Information about the authors

V.P. Kulikova – corresponding author, Professor, Candidate of Technical Sciences, Associate Professor, Department of Information and Communication Technologies, Manash Kozybayev North Kazakhstan University NPLC, Petropavlovsk, Kazakhstan; e-mail: y4lentina@mail.ru;

I.A. Chupchikov – master's student, Department of Information and Communication Technologies, Manash Kozybayev North Kazakhstan University NPLC, Petropavlovsk, Kazakhstan; e-mail: igor.chupchikov.03@gmail.com.