Krassimir Markov, Vladimir Ryazanov,
Vitalii Velychko, Levon Aslanyan
(editors)

New Trends
in
Classification and Data Mining

I T H E A
SOFIA
2010

Krassimir Markov, Vladimir Ryazanov, Vitalii Velychko, Levon Aslanyan (ed.)
New Trends in Classification and Data Mining
ITHEA®
Sofia, Bulgaria, 2010

First edition

Recommended for publication by The Scientific Concil of the Institute of Information Theories and Applications FOI ITHEA

This book maintains articles on actual problems of classification, data mining and forecasting as well as natural language processing:
- new approaches, models, algorithms and methods for classification, forecasting and clusterisation. Classification of non complete and noise data;
- discrete optimization in logic recognition algorithms construction, complexity, asymptotically optimal algorithms, mixed-integer problem of minimization of empirical risk, multi-objective linear integer programming problems;
- questions of complexity of some discrete optimization tasks and corresponding tasks of data analysis and pattern recognition;
- the algebraic approach for pattern recognition - problems of correct classification algorithms construction, logical correctors and resolvability of challenges of classification, construction of optimum algebraic correctors over sets of algorithms of computation of estimations, conditions of correct algorithms existence;
- regressions, restoring of dependences according to training sampling, parametrical approach for piecewise linear dependences restoration, and nonparametric regressions based on collective solution on set of tasks of recognition;
- multi-agent systems in knowledge discovery, collective evolutionary systems, advantages and disadvantages of synthetic data mining methods, intelligent search agent model realizing information extraction on ontological model of data mining methods;
- methods of search of logic regularities sets of classes and extraction of optimal subsets, construction of convex combination of associated predictors that minimizes mean error;
- algorithmic constructions in a model of recognizing the nearest neighbors in binary data sets, discrete isoperimetry problem solutions, logic-combinatorial scheme in high-throughput gene expression data;
- researches in area of neural network classifiers, and applications in finance field;
- text mining, automatic classification of scientific papers, information extraction from natural language texts, semantic text analysis, natural language processing.

It is represented that book articles will be interesting as experts in the field of classifying, data mining and forecasting, and to practical users from medicine, sociology, economy, chemistry, biology, and other areas.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Bulgaria

ISBN 978-954-16-0042-9

C\o Jusautor, Sofia, 2010

БАЗОВЫЕ СТРУКТУРЫ ЕВКЛИДОВЫХ ПРОСТРАНСТВ: КОНСТРУКТИВНЫЕ МЕТОДЫ ОПИСАНИЯ И ИСПОЛЬЗОВАНИЯ

Владимир Донченко, Юрий Кривонос, Виктория Омардибирова

Аннотация: *Предложена и детально рассмотрена концепция базовых структур линейного пространства, включающих основные линейные и основные нелинейные объекты. Развиты конструктивные методы их построения, описания, перехода и использования на основе систематического развития и применения аппарата псевдообращения по Муру - Пенроузу. Важность приведённых результатов и возможности их применения проиллюстрирована на примере широкого спектра прикладных задач.*

Ключевые слова*: Псевдообращение по Муру - Пенроузу, сингулярное представление матрицы, метод наименьших квадратов, линейная регрессия, системы оптимального управления, прогноз, кластеризация, искусственные нейронные сети.*

ACM Classification Keywords*: G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization; G.2.m. Discrete mathematics: miscellaneous.*

Вступление

В работе предложена и обоснована концепция базовых структур евклидового пространства, к которым предлагается отнести основные линейные структуры, а также – основные нелинейные. И те и другие структуры евклидова пространства проявляются либо во множественной форме: подпространства и гиперплоскости, либо в единичной: линейные операторы. То же относится к нелинейным структурам. К таким нелинейным структурам относятся с одной стороны: матрицы квадратичных форм (в работе – неотрицательно определённых). С другой – поверхности уровня, отвечающие единичному значению соответствующей квадратичной формы. Такими поверхностями уровня являются эллипсоиды, точнее: цилиндрические эллипсоиды. Описаны конструктивные способы задания взаимного перехода от одних типов структур к другим: от линейных подпространств и гиперплоскостей к матрицам и наоборот, а также от набора векторов к матрицам квадратичных форм и эллипсам группировки. В числе других рассмотрены конструктивные способы порождения подпространств и гиперплоскостей, а также ортогональных проекторов, связанных с указанными объектами. В том же русле лежит рассмотрение группирующих операторов. Упомянутая конструктивность обеспечивается применением псевдообращения по Муру – Пенроузу (ПдО), а также новыми результатами в этой области, берущими своё начало и опирающимися на фундаментальную работу Н.Ф Кириченко [Кириченко, 1997]. Важность и эффективность использования приведённых результатов проиллюстрирована на широком спектре задач, включающем линейную регрессию, в том числе векторную, теорию оптимального управления, кластеризацию, прогноз и функциональные сети, являющиеся обобщение искусственных нейронных сетей.

Как отмечалось в работе [Донченко,2009], «структура объекта» ассоциируется с представлением об объекте, как чём-то едином, составленном из взаимодействующих, связанных между собою частей. Математическое описание - моделирование объекта связано с передачей представления о «структуре» объекта математически: средствами математического описания «связей». Это означает, что структура объекта должна быть передана, отражена в математической модели средствами математического структурирования. К последним относятся четыре базовых структуры, а также их комбинации. К четырём базовым следует отнести отношения, функции, операции и наборы подмножеств. К комбинациям относятся, например, структуры линейного и евклидового пространства.

«Линейные структуры»: структуры линейного пространства, занимают ведущее место среди важнейших математических структур. Именно в рамках линейных пространств осуществляется уточнение понятия линейной структуры. В рамках рассмотрения линейных пространств к ним относят линейные подпространства и гиперплоскости а также – линейные операторы и функционалы. Евклидово пространство: конечномерное линейное плюс скалярное произведение, занимает ведущее место среди линейных структур по богатству возможностей использования связей. Богатство свойств линейных структур: и в варианте линейных пространств и подпространств, и в варианте операторов соответствующего вида, – трудно переоценить в математическом моделировании объектов. Это касается как абстрактных математических исследований, так и исследований прикладного характера. В полной мере сказанное выше относится, в частности, к алгебре, регрессионному анализу, теории случайных процессов, теории дифференциальных и интегральных уравнений, систем оптимального управления, прикладным задачам классификации, прогноза и т.д. Важную роль в прикладных исследованиях играют конструктивные методы описания соответствующих объектов. В том, что касается линейных операторов и линейных функционалов, вопрос конструктивности решается построением матриц соответствующих объектов, а для операций – использованием операций матричной алгебры. В том, что касается подпространств, порождённых теми или иными совокупностями векторов, дело обстоит сложнее. Их конструктивное описание можно получить, связав указанные объекты с пространством значений подходящей матрицы, которое в свою очередь описывают подходящим ортогональным проектором. Именно этот подход развивается ниже. Отметим, что ортогональные проекторы играют важную роль в исчерпывающем исследовании систем линейных алгебраических уравнений (СЛАУ). Принципиально важны они также в постановке и решении важных оптимизационных задач с квадратичными функционалами качества в евклидовых пространствах. Это в полной мере относится к задаче построении наилучших квадратичных приближений правой части СЛАУ значениями левой, когда СЛАУ не имеет точных решений. Такие наилучшие приближения называют также псевдорешениями. Конструктивное описание ортогональных проекторов в связи с естественными подпространствами линейного оператора прямо определяется псевдообращением по Муру – Пенроузу [Moore,1920], [Penrose, 1955] (см. также, например, [Алберт, 1977]). Отметим, также, что важную роль в конструктивном решении прикладных задач с использованием линейных структур играет сингулярное представление (его называют также сингулярным разложением или SVD - представлением) матрицы в специфической записи в виде взвешенной суммы тензорных произведений специального набора пар векторов. Ниже рассматриваются основные свойства линейных структур, основные особенности и возможности их конструктивного описания, а также – использования для конструктивного решения важнейших прикладных задач прогноза, кластеризации и классификации, в других областях.

В заключение отметим, что основные идеи, дух и результаты предлагаемой работы восходят и используют результаты развития теории псевдообращения в работах нашего безвременно ушедшего коллеги и друга, профессора Н.Ф. Кириченко.

Евклидовы пространства: базовые линейные структуры и связи между ними

В дальнейшем, говоря о евклидовом пространстве R^n, будем иметь в виду множество конечных числовых последовательностей одной и той же длины n, записанных в столбик с покоординатными операциями сложения и умножения на скаляр и суммой покоординатных произведений в качестве скалярного произведения. Именно такой вариант евклидового пространства будем стандартным образом обозначать через R^n, а его элементы – через $a = \begin{pmatrix} a_1 \\ \cdots \\ a_n \end{pmatrix}$. Стандартные ортонормированные базисы, составленные из векторов с единственной единичной компонентой (остальные – нули) на месте с соответствующим номером будут обозначаться для R^m и R^n соответственно через $e(j) \in R^m, j = \overline{1,m}, e_{(i)} \in R^n, i = \overline{i,n}$. Оператор A из R^n, в R^m: $A: R^n \to R^m$, в ортонормированных базисах $e(j) \in R^m, j = \overline{1,m}, e_{(i)} \in R^n, i = \overline{i,n}$ будем отождествлять с $m \times n$ - матрицей $A = (a_{ij})$ этого оператора. Для матрицы $A = (a_{ij})$ будем использовать также блочное представление по столбцам (столбцовое) и строкам (строчное):

$$A = \begin{pmatrix} a_{(1)}^T \\ \cdots \\ a_{(m)}^T \end{pmatrix} = (a(1)\vdots\ldots\vdots a(n)), a_{(i)}^T \in R^n, i = \overline{1,m}, a(j) \in R^m, j = \overline{1,n}.$$

Линейное пространство всех $m \times n$ матриц будем обозначать $R^{m\times n}$.

Линейное подпространство, порождённое системой векторов, $c_k \in R^p, k = \overline{1,K}$ будет обозначаться через $L(c_k, k = \overline{1,K}) \equiv L(c_1 ..., c_K)$, а линейное подпространство значений линейного оператора $A: R^n \to R^m$ – через L_A.

Первым из набора базовых свойств является утверждение о том, что

1.

$$L_A = L(a(1),...,a(n)).$$

Таким образом, линейное подпространство, порождённое набором векторов, совпадает с подпространством значений матрицы, составленной из векторов набора, как из столбцов.

2.“Для элементов столбцового и строчного представления матрицы $A \in R^{m\times n}$ справедливы соотношения

$$a(j) = Ae_{(j)}, j = \overline{1,n},$$

$$a_{(i)}^T = e^T(i)A, i = \overline{1,m}.$$

3.Для произведения произвольных матриц B, C со столбовым и строчным представлением

$$B = (b(1)\vdots\ldots\vdots b(r)), b(j) \in R^m, j = \overline{1,r}, C = \begin{pmatrix} c_{(1)}^T \\ \cdots \\ c_{(r)}^T \end{pmatrix}, c_{(r)}, c(i) \in R^n, i = \overline{1,r}$$

соответственно и диагональной матрицы $\Lambda = diag(\lambda_1,...,\lambda_r)$ справедливо соотношение

$$B\Lambda C = \sum_{i=1}^{r} \lambda_i b(i) c_{(i)}^T.$$

Важной составляющей аппарата конструктивного описания и использования линейных структур является понятие ортогонального проектора, которое полностью отвечает стандартному геометрическому представлению об ортогональном проектировании. Общей, основой эффективного использования ортогональных проекторов является наличие двух эквивалентных определений таких проекторов и возможности их конструктивного построения в связи с линейными подпространствами через псевдообращение.

4.«Геометрическое определение ортогонального проектора»: для разложения $R^p = L + L^{\perp}$ в прямую сумму ортогональных подпространств ортогональным проектором P_L на линейное подпространство $L \subseteq R^p$ называется оператор, определяемый соотношением

$$P_L x = P_L(x_L + x_{L^{\perp}}) = x_L \text{, где}$$
$$x = x_L + x_{L^{\perp}}, x_L \in L, x_{L^{\perp}} \in L^{\perp} \quad (1)$$

– однозначное разложение произвольного вектора $x \in R^p$ по двум составляющим ортогональной суммы. Очевидным образом оператор ортогонального проектирования является линейным оператором.

5.Разложение (1) произвольного вектора $x \in R^p$ в силу симметричности относительно ортогональных слагаемых определяет одновременно два ортогональных проектора: $P_L, P_{L^{\perp}}$ с очевидным соотношением

$$P_L + P_{L^{\perp}} = E_p$$

где E_p – единичная матрица соответствующей размерности.

6.Для ортогонального проектора P_L на подпространство L оператор $Z_L \equiv E_p - P_L$ является ортогональным проектором на ортогональное дополнение $L^{\perp}$ к $L: Z_L \equiv E_p - P_L = P_{L^{\perp}}$.

7.Абстрактное определение ортогонального проектора: для того, чтобы линейный оператор $P: R^p \to R^p$, был оператором ортогонального проектирования необходимо и достаточно, чтобы он был идемпотентным симметричным оператором. Линейное пространство L_P, на которое совершается ортогональное проектирование в соответствии с «геометрическим определением» описывается одним из двух соотношений:

$$L_P = \{x : x = Pu, u \in R^p\} = \{x : x = Px, x \in R^p\}.$$

8.Сингулярное или SVD- представление произвольной $m \times n$ матрицы: для произвольной $A \in R^{m \times n}$ ранга $r \le \min(m,n)$ справедливо следующее представление матрицы в виде взвешенной суммы тензорных произведений

$$A = \sum_{i=1}^{r} \lambda_i u_i v_i^T \quad (2)$$

где

- $\lambda_1^2 \ge ... \ge \lambda_r^2 > 0$ общий набор ненулевых собственных чисел матриц $AA^T, A^T A$.
- $u_i \in R^m, i = \overline{j,r}$ - ортонормированный набор собственных векторов матрицы AA^T, отвечающих ненулевым собственным числам: $AA^T u_i = \lambda_i^2, \lambda_i^2 > 0, i = \overline{1,r}, u_i^T u_j = \delta_{ij}, i \ne j$;
- $v_i \in R^n, i = \overline{j,r}$ -- ортонормированный набор собственных векторов матрицы $A^T A$, отвечающих ненулевым собственным числам: $A^T A v_i = \lambda_i^2 v_i, \lambda_i^2 > 0,, i = \overline{1,r}, v_i^T v_j = \delta_{ij}, i \ne j$.

Принципиальную роль в описании базовых структур евклидовых пространств играет псевдообращение по Муру - Пенроузу: [Moore,1920], [Penrose, 1955] как одноместной операции A^+ над прямоугольными матрицами произвольной размерности: для произвольных $A \in R^{m \times n}$. В дальнейшем термин «псевдообращение» будем сокращать до ПдО.

Евклидовы пространства, базовые линейные структуры и ПдО

9.Определение псевдообращения через SVD - представление матрицы: для произвольной - матрицы A её ПдО A^+ определяется соотношением

$$A^+ = \sum_{i=1}^{r} \lambda_i^{-1} v_i u_i^T : R^m \to R^n , \tag{3}$$

где $u_i, v_i, \lambda_i, i = \overline{j,r}$ – элементы сингулярного разложения матрицы из соотношения (2).

Заметим, что SVD - определение ПдО (соотношение (3)) позволяет легко установить, что ПдО коммутирует с транспонированием, а также ряд других полезных соотношений, в частности, что

$$A^T (A^T)^+ = A^+ A .$$

10.Ортогональные проекторы на подпространства L_A, L_{A^T} в евклидовых пространства R^m, R^m соответственно, обозначим их $P(A^T), P(A)$ соответственно, определяются соотношениями

$$P(A^T) = AA^+ = \sum_{i=1}^{r} u_i u_i^T ,$$

$$P(A) = P\left((A^T)^T\right) = A^T (A^T)^+ = A^+ A = \sum_{i=1}^{r} v_i v_i^T .$$

11.“Операторы $Z(A), Z(A^T)$, определяемые соотношениями

$$Z(A) = E_n - P(A) = E_n - A^+ A ,$$

$$Z(A^T) = E_m - P(A^T) = E_m - A^{T+} A^T = E_m - AA^+ ,$$

являются операторами ортогонального проектирования на подпространства $L_A^{\perp}, L_{A^T}^{\perp}$

Важность последних соотношений определяются тем, что $L_{A^T}^{\perp}$ является множеством нулей оператора A.

12.Подпространство $L_{A^T}^{\perp}$ является ядром $KerA$ (множеством нулей) оператора A:

$$L_{A^T}^{\perp} = KerA = Z(A) R^n .$$

13.Для совместности СЛАУ $Ax = y$ необходимо и достаточно, чтобы $y^T Z(A^T) y = 0$. В этом случае $A^+ y$ является наименьшим по норме решением. Оно ортогонально к $KerA$, а множество всех решений Ω_y описывается соотношением

$$\Omega_y = A^+ y + Z(A) R^n = \{x : x = A^+ y + Z(A) v, v \in R^n\} . \tag{4}$$

14.Если СЛАУ $Ax = y$ несовместна, т. е $y^T Z(A^T) y > 0$ множество, определяемое соотношением (4) описывает совокупность всей наилучших квадратичных приближений правой части значениями левой:

$$\Omega_y = A^+ y + Z(A)R^n = \underset{x \in R^n}{Arg\min} \| Ax - y \|^2 . \qquad (5)$$

Значение невязки для любого наилучшего квадратичного приближения составляет $y^T Z(A^T)y$.

15.Для того, чтобы матричное уравнение $AX = Y, A \in R^{m\times n} X \in R^{n\times N}, Y \in R^{m\times N}$ имело корни необходимо и достаточно, чтобы $trY^T Z(A^T)Y = 0$. В этом случае множество Ω_Y определяется соотношением

$$\Omega_Y = \{X : X = A^+ Y + Z(A)V, V \in R^{n\times N}\} . \qquad (6)$$

16.Для линейной зависимости вектора $d \in R^m$ от столбцов матрицы $A \in R^{m\times n}$ необходимо и достаточно, чтобы выполнялось соотношение

$$d^T Z(A^T)d = 0. \qquad (7)$$

Ссылка на векторы—элементы столбцового представления матрицы A нисколько не ограничивают возможности применения утверждения этого пункта для определения линейной зависимости того или иного вектора от фиксированного набора векторов. Для использования результата в общем случае необходимо и достаточно составить матрицу, в которой векторы набора являются столбцами, и дополнительно использовать п.1 базовых свойств.

Отметим также, что условием линейной независимости строки $a^T, a \in R^n$ от строк матрицы $A \in R^{m\times n}$ является условие

$$a^T Z(A)a = 0. \qquad (8)$$

17. Прямые формулы Гревиля: ПдО произвольной матрицы $A \in R^{m\times n}$, дополненной строкой $a^T \in R^n$, определяется элементами $P \in R^{n\times m}, q \in R^n$ блочного представления ПдО $\begin{pmatrix} A \\ a^T \end{pmatrix}^+ = (P \vdots q)$ расширенной матрицы в соответствии с формулами

$$q = \begin{cases} \dfrac{Z(A)a}{a^T Z(A)a}, a^T Z(A)a > 0 (\text{нез.}) \\ \dfrac{R(A)a}{1 + a^T R(A)a}, a^T Z(A)a = 0 (\text{зав.}) \end{cases}, \qquad (9)$$

$$P = (E_m - qa^T)A^+ ,$$

где

$$R(A) = A^+ A^{+T} .$$

Первая строка в (9) отвечает случаю линейной независимости строки - расширения от строк матрицы A, второй – линейной зависимости.

С учётом коммутирования транспонирования с ПдО, прямые формулы Гревиля очевидным образом переписываются для варианта расширения матрицы столбцом.

18.Обратные формулы Гревиля: для блочного представления ПдО расширенной матрицы в виде $\begin{pmatrix} A \\ a^T \end{pmatrix}^+ = (P \vdots q)$ ПдО матрицы A определяется соотношениями

$$A^{+}=\begin{cases}(E\ -\dfrac{qq^{T}}{\|q\|^{2}})P, a^{T}q=1(\textit{незав.})\\(E\ +\dfrac{qa^{T}}{1-a^{T}q})P, a^{T}q<1(\textit{зав.}))\end{cases}. \tag{10}$$

Условие в первой строке в соотношениях (10) отвечает линейной независимости строки, которая удаляется, а второй – зависимости от остальных строк матрицы A. Справедливость этих условий непосредственно вытекает из п.17 “, и используется, когда дополнительно известна ПдО для расширенной матрицы.

19.Квадрат расстояния $\rho^{2}(a,L_{A})$ вектора $a\in R^{m}$ от подпространства L_{A} определяется соотношением

$$\rho^{2}(a,L_{A})=\min_{y\in R^{m}}\|a-y\|^{2}=a^{T}Z(A^{T})a\ .$$

20.Квадрат расстояния $\rho^{2}(a,\Gamma(b,L_{A}))$ вектора $a\in R^{m}$ от гиперплоскости $\Gamma(b,L_{A})=b+L_{A}$ определяется соотношением

$$\rho^{2}(a,\Gamma(b,L_{A}))=\min_{y\in\Gamma(b,L_{A})}\|a-y\|^{2}=(a-b)^{T}Z(A^{T})(a-b)\ .$$

Отметим, что привязка подпространств в пп.19, 20 к множеству значений оператора A не ограничивает сферу применимости результатов. С помощью п.1 они очевидным образом распространяются на ситуацию, когда подпространство порождается заданной конечной совокупностью векторов.

Формулы аналитического возмущения описывают ПдО изменённой матрицы, когда возмущение имеет вид аддитивной добавки $ab^{T}, a\in R^{m}, b\in R^{n}$. В работе [Кириченко,1997] исчерпывающим образом исследованы варианты представления ПдО возмущённой матрицы $(A+ab^{T})^{+}$. Как оказывается, вид соответствующих формул определяется линейной зависимостью или независимости векторов a,b^{T} от, соответственно, столбцов и строк матрицы A. Кроме того, на вид соответствующих формул влияет сохранение или падение ранга возмущенной матрицей. Последнее касается случая, когда одновременно оба вектора a,b^{T} линейно зависимы с, соответственно, столбцами и строками матрицы A. И условия линейной независимости и условие падения ранга носят аналитический характер. В п.16 представлены условия линейной зависимости. Условие сохранения ранга представлено следующим пунктом.

21.Ранги матриц A и $A+ab^{T}$ одинаковы (a,b^{T} одновременно зависимы от, соответственно, столбцов и строк матрицы A): $rank(A+ab^{T})=rankA$, тогда и только тогда, когда $b^{T}A^{+}a\neq-1$. Ранг возмущённой матрицы падает, когда $b^{T}A^{+}a=-1$.

Принимая во внимание громоздкость соответствующих формулировок, ниже приведен один из вариантов утверждения о виде ПдО для возмущённой матрицы. С полным вариантом утверждения можно ознакомиться в уже упомянутой работе [Кириченко,1997].

22.Аналитические формулы возмущения ПдО матриц, фрагмент: если компоненты возмущения: a,b^{T} линейно не зависимы от, соответственно, столбцов и строк матрицы A, т.е. $a^{T}Z(A^{T})a>0, b^{T}Z(A)b>0$, то

$$(A+ab^{T})^{+}=A^{+}-\frac{A^{+}aa^{T}Z(A^{T})}{a^{T}Z(A^{T})a}-\frac{Z(A)bb^{T}A^{+}}{b^{T}Z(A)b}+Z(A)ba^{T}Z(A^{T})\frac{1+b^{T}A^{+}a}{a^{T}Z(A^{T})ab^{T}Z(A)b}.$$

Евклидовы пространства, базовая нелинейная структура: группирующие операторы

Важнейшими нелинейными структурами евклидова пространства являются квадратичные формы, точнее: неотрицательно определённые квадратичные формы, – и отвечающие им эллипсы или эллипсоидальные цилиндры. Среди таких нелинейных структур принципиальными являются матрицы квадратичных форм, которые, как, оказывается, естественным образом связаны с групповыми свойствами набора векторов и которые поэтому естественно называть группирующими операторами. Группирующие операторы возникают в связи с набором векторов $a_j \in R^m, j = \overline{1,n}$ и отвечающей ему матрицей $A = (a_1 \vdots ... \vdots a_n), a_j \in R^m, j = \overline{1,n}$. В паре набор-матрица первый будем называть столбцовым представлением матрицы, вторую – матричным представлением набора. Как и ортогональные проекторы, группирующие операторы являются парными. Будем обозначать их, соответственно $R(A), R(A^T)$.

22.Определение группирующих операторов:

$$R(A) = A^+ A^{+T}$$

$$R(A^T) = A^{+T} A^+ .$$

23.Проектирование на нормированный вектор $u \in R^m : \| u \| = 1$ элементов набора векторов $a_j \in R^m, j = \overline{1,n}$. Основной результат этого пункта представлен леммой 1 ниже.

Лемма 1. Для произвольного набора $a_i \in R^m, i = \overline{1,n}$ с матричным представлением $A = (a_1,...,a_n)$ и произвольного нормированного вектора $u \in R^m : \| u \| = 1$, справедливо равенство:

$$\sum_{j=1}^{n} a_j{}^T u u^T a_j = u^T A A^T u \qquad (11)$$

Доказательство. Действительно, принимая во внимание связь векторов набора $a_i \in R^m, i = \overline{1,n}$ со своим матричным представлением в виде

$$a_j = Ae_{(j)}, j = \overline{1,n} ,$$

где $e_{(j)} \in R^n, j = \overline{1,n}$ - стандартный ортонормированный базис в $R^n : e^T{}_{(j)} = (0,...,\underset{j}{1},0...0), j = \overline{1,n}$,имеем:

$$\sum_{j=1}^{n} a_j{}^T u u^T a_j = \sum_{j=1}^{n} e^T{}_{(j)} A^T u u^T A e_{(j)} = \sum_{j=1}^{n} u^T A e_{(j)} e^T{}_{(j)} A^T u = u^T A \left[\sum_{j=1}^{n} e_{(j)} e^T{}_{(j)} \right] A^T u .$$

Остаётся только заметить, что

$$\sum_{j=1}^{n} e_{(j)} e^T{}_{(j)} = E_n ,$$

где E_n - единичная матрица (оператор) в R^n , и доказательство леммы завершено.

Замечание 1. Левая часть соотношение (11) леммы 1, собственно, представляет собою сумму квадратов проекций векторов набора $a_i \in R^m, i = \overline{1,n}$ на нормированный вектор $u \in R^m : \| u \| = 1$.

24.Проектирование на элементы $u_i \in R^m, i=\overline{1,r}$ SVD –разложения матричного представления A набора векторов $a_j \in R^m, j=\overline{1,n}$. Основной результат этого пункта представлен леммой 2, приведённой ниже.

Лемма 2. Для произвольного набора $a_i \in R^m, i=\overline{1,n}$ с матричным представлением $A=(a_1,\ldots,a_n)$ имеет место соотношение:

$$\sum_{j=1}^{n} a_j^T u_i u_i^T a_j = u_i^T A A^T u_i = \lambda_i^2, i=\overline{1,r}$$

Доказательство вытекает из леммы предыдущего пункта и из п.8, в котором в рамках сингулярного представления матрицы наборы $u_i \in R^m, v_i \in R^n, r = rankA$ определяются как ортонормированные наборы собственных векторов матриц AA^T, $A^T A$, отвечающих общему набору ненулевых собственных чисел $\lambda_i^2 > 0, i=\overline{1,r}$.

25. Группирующие операторы: эллипсы группировки набора векторов $a_i \in R^m, i=\overline{1,n}$. Основное отверждение пункта – теорема 1 ниже.

Теорема 1. Пусть $a_i \in R^m, i=\overline{1,n}$ произвольный набор векторов из R^n с матричным представлением $A=\left(a_1 \vdots \ldots \vdots a_n\right)$. Тогда все векторы набора принадлежат внутренности эллипса, точнее: эллипсоидального цилиндра, определяемого уравнением

$$x^T R(A^T) x = r, x \in R^m,$$

где, $R(A^T)$, «группирующий» оператор, определяемый стандартным образом

$$R(A^T) = A^{+^T} A^+ .$$

Доказательство. Рассмотрим квадраты проекций векторов набора $a_j, j=\overline{1,n}$ на каждый из векторов $u_i, i=\overline{1,r}$ SVD-представления (2) матрицы A. Принимая во внимание ортонормированность набора $u_i, i=\overline{1,r}$: $u_i^T u_j = \delta_{ij}, i \neq j, i, j=\overline{1,r}$, а также сингулярное представление (3), и обозначая квадраты проекций через $\left\|\Pr_{u_i} a_j\right\|^2, i, j=\overline{1,r}$, очевидным образом имеем

$$\left\|\Pr_{u_i} a_j\right\|^2 = a_j^T u_i u_i^T a_j, j=\overline{1,n}, i=\overline{1,r} .$$

Суммирование по всем векторам набора $a_j, j=\overline{1,n}$ и применение леммы 2 даёт

$$\sum_{j=1}^{n} a_j^T u_i u_i^T a_j = u_i^T A A^T u_i = \lambda_i^2, i=\overline{1,r},$$

т.е.

$$\sum_{j=1}^{n} a_j^T u_i u_i^T a_j = \lambda_i^2, i=\overline{1,r} .$$

Таким образом, после деления обеих частей последнего соотношения соответственно на $\lambda_i^2, i=\overline{1,r}$, имеем

$$\sum_{j=1}^{n}\frac{a_j{}^T u_i u_i^T a_j}{\lambda_i^2}=\frac{u_i^T AA^T u_i}{\lambda_i^2}=1, i=\overline{1,r},$$

т.е.

$$\sum_{j=1}^{n}\frac{a_j{}^T u_i u_i^T a_j}{\lambda_i^2}=1, i=\overline{1,r}.$$

Свернув (просуммировав) последнее равенство по $i=\overline{1,r}$, получаем

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{a_j{}^T u_i u_i^T a_j}{\lambda_i^2}=\sum_{i=1}^{m}\frac{u_i^T AA^T u_i}{\lambda_i^2}=r.$$

Поменяв порядок суммирования в двойной сумме, получаем

$$\sum_{j=1}^{n}\sum_{i=1}^{m}\frac{a_j{}^T u_i u_i^T a_j}{\lambda_i^2}=\sum_{j=1}^{n}a_j{}^T\sum_{i=1}^{m}\frac{u_i u_i^T}{\lambda_i^2}a_j=r.$$

Приняв во внимание, что

$$\sum_{i=1}^{r}\frac{u_i u_i^T}{\lambda_i^2}=A^{+T}A^{+}=R(A^T),$$

получаем окончательно

$$\sum_{j=1}^{n}\frac{a_j{}^T u_i u_i^T a_j}{\lambda_i^2}=\sum_{j=1}^{n}a_j{}^T\sum_{i=1}^{m}\frac{u_i u_i^T}{\lambda_i^2}a_j=\sum_{j=1}^{n}a_j{}^T A^{+T}A^{+}a_j=\sum_{j=1}^{n}a_j{}^T=\sum_{j=1}^{n}a_j{}^T R(A^T)a_j=r,$$

т.е.

$$\sum_{j=1}^{n}a_j{}^T R(A^T)a_j=r. \tag{12}$$

Поскольку $R(A^T)$ – симметричная, неотрицательно определённая матрица, то следствием соотношения (12) является одновременное выполнение неравенств

$$a_j{}^T R(A^T)a_j\le r, j=\overline{1,n}. \tag{13}$$

Принимая во внимание уже упомянутую выше симметричность и неотрицательную определённость квадратичной формы $x^T R(A^T)x, x\in R^m$, уравнение

$$x^T R(A^T)x=r, x\in R^m, \tag{14}$$

определяет эллипс, точнее: эллипсоидальный цилиндр, в R^m с длинами $\frac{1}{\lambda_i\sqrt{r}}, i=\overline{1,r}$ нетривиальных полуосей. Напомним, что $r=rankA\le\min(m,n)$.

Таким образом, выполнение неравенства (13) для всех векторов набора $a_i\in R^m, i=\overline{1,n}$ означает их одновременную принадлежность внутренности эллипсоидального цилиндра с уравнением (14), и доказательство теоремы завершено.

Замечание 2. В действительности неравенство (13) может давать существенное загрубление «радиуса» эллипса. Так, при векторах $a_i\in R^m, i=\overline{1,n}$, близких к ортогональным, очевидным образом, константа в правой части в правой части (13) можно выбрать близкой к 1. Так что справедлив более жёсткий вариант теоремы 1, являющийся предметом следующего утверждения.

26. Усиление результата об эллипсах группировки.

Теорема 2. Векторы произвольного набора $a_i \in R^m, i = \overline{1,n}$ с матричным представлением $A = (a_1 \vdots \ldots \vdots a_n)$ принадлежат внутренности эллипсоидального цилиндра

$$x^T R(A^T)x = r_{\max}, r_{\max}^2 \le r = rankA, x \in R^m,$$

$$r_{\max}^2 = \max_{j=\overline{1,n}} a_j^T R(A^T)a_j .$$

Применения: линейная регрессия, скалярные наблюдения

Применение ПдО в линейной регрессии определяется тем, что МНК – оценка $\hat{\beta}$ (оценка метода наименьших квадратов) неизвестного параметра $\beta \in R^p$ линейной регрессии $y = \sum_{j=0}^{p-1} \beta_j f_j(x) + \varepsilon$ на основе наблюдений $(x_i, y_i), x_i \in R^m, y_i \in R^1, i = \overline{1,n}$ определяется решением оптимизационной задачи

$$\hat{\beta} \in \underset{\beta \in R^p}{Arg \min} \| X\beta - Y \|^2 , \tag{15}$$

в которой матрица X – матрица плана, а Y – вектор - столбец с компонентами $y_i \in R^1, i = \overline{1,n}$ – вектор наблюдений регрессанта. В соответствии с соотношением (4) п.13 общее решении задачи (15) определяется соотношением

$$\hat{\beta} \in X^+ Y + Z(X)v, v \in R^p \tag{16}$$

со свободным параметром $v \in R^p$.

Решение задачи МНК - оценивания в виде (16) полностью согласуется с классическим решением уравнения Гаусса – Маркова в виде

$$\hat{\beta} = (X^T X)^{-1} X^T Y ,$$

поскольку в случае полного столбцового ранга матрицы плана, её ПдО X^+ и ортогональный проектор $Z(X)$ определяются соотношениями

$$X^+ = (X^T X)^{-1} X^T , Z(X) = 0 .$$

В этом случае классическом случае множество МНК – оценок в (16) является одноэлементным.

Применения: задача терминального управления

Под задачей терминального управления для линейной динамической системы с дискретным временем

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{b}(k)u(k),$$

$$\mathbf{x}(0) = \mathbf{x}_{(0)},$$

где

$$\mathbf{x}(k) \in \mathrm{R}^n, \; \mathrm{u(k)} \in \mathrm{R}^1, A(k) \in \mathrm{R}^{n \times n}, \;\; b(k) \in \mathrm{R}^n, \;\; k = \overline{0,N} ,$$

имеют в виду задачу выбора такого управления $u(k)$, $k=\overline{0,N}$, которое позволяет вывести фазовую траекторию в момент $N+1$ на уровень $\mathbf{x}_{(1)}$ или, если это невозможно, выбором того же управления минимизировать отклонение $\left\|\mathbf{x}(N+1)-\mathbf{x}_{(1)}\right\|^2$.

Принципиальным результатом для исследования задачи терминального управления является теорема редукции, позволяющая свести задачу терминального управления к СЛАУ.

Теорема 3(теорема редукции). Задача терминального управления является эквивалентной СЛАУ

$$\mathbf{W}(N+1)\mathbf{u}=\mathbf{x}_{(1)}-\mathbf{A}(N-1)\mathbf{A}(N-2)\ldots\mathbf{A}(0)x_{(0)},$$

в которой вектор $u\in R^{N+1}$ – объединенный вектор управления, а матрица $\mathbf{W}(N+1)$ является блочной:

$$\mathbf{W}(N+1)=\left(\mathbf{W}(N+1,0)\vdots\mathbf{W}(N+1),1)\vdots\ldots\vdots\mathbf{W}(N+1,N)\right),$$

с блоками $\mathbf{W}(N+1),k),k=\overline{0,N}$, определяемыми соотношениями

$$\mathbf{W}(N+1,k)=\mathbf{A}(N)\mathbf{A}(N-1)\ldots\mathbf{A}(k+1)b(k),k=\overline{0,N-1},$$

$$\mathbf{W}(N+1,N)=b(N).$$

Теорема редукции позволяет исчерпывающим образом исследовать задачу терминального управления с помощью пп.13,14 . Следует добавить, что аналогичным образом, с помощью ПдО удаётся исчерпывающим образом исследовать задачу терминального наблюдения, в том числе в случае ошибок и шумов (см., например, [Кириченко, Донченко, 2005])

Применения: кластеризация

ПдО расширяет возможности кластеризации, позволяя эффективно погружать классифицируемые объекты в подходящие подпространства или гиперплоскости. П.1 дает возможность связывать подпространство, порождённое набором векторов, с подходящей матрицей. Если объект связывается с гиперплоскостью, то её смещение – это, как правило, среднее по векторам порождающей совокупности, а подпространство – это подпространство значений матрицы, построенной из центрированных средним векторов порождающей совокупности, как из столбцов. Результаты пп..19, 20 обеспечивают возможность конструктивного вычисления расстояний от объектов (подпространств или гиперплоскостей), ассоциируемых с порождающей совокупностью. Применение стандартных рекуррентных последовательно уточняемых разбиений с расстояниями соответствия из п.19 или п. 20 придаёт процедуре кластеризации необходимой завершённости. С подробностями можно ознакомиться, например, в работах: [Кириченко, Донченко, 2007], [Кириченко, Донченко, 2008]).

Группирующие операторы дают возможность построить расстояния соответствия в связи с использованием базой нелинейной структуры: эллипса группирования и соответствующей квадратичной формы, описываемой группирующим оператором. В сущности, речь идёт о том, что результат теоремы 2 можно использовать для определения расстояния соответствия, точнее его квадрата, обозначим его $\rho^2(x,Kl)$, между вектором $x\in R^m$ и кластером Kl, порожденным обучающей выборкой $a_j\in R^m, j=\overline{1,n}$.

Обозначим через $\bar{a}$ среднее обучающей выборки $a_j\in R^m, j=\overline{1,n}$, а через $\tilde{A}$ матричное представление для векторов $\tilde{a}_j\in R^m, j=\overline{1,n}$ первоначального набора, центрированных средним:

$$\bar{a} = \frac{1}{n}\sum_{j=1}^{n} a_j ,$$

$$\tilde{a}_j = a_j - \bar{a} ,$$

$$\tilde{A} = (\tilde{a}_1 \vdots \ldots \vdots \tilde{a}_n) .$$

Тогда квадрат расстояния, определяемый соотношением

$$\rho^2(x, Kl) = \frac{1}{r_{\max}^2}(x - \bar{a})^T R(\tilde{A}^T)(x - \bar{a}) ,$$

является расстоянием, определяемой стандартной поверхностью уровня, т. е поверхностью, определяемой уравнением

$$(x - \bar{a})^T R(\tilde{A}^T)(x - \bar{a}) = r_{\max}^2 .$$

Такое определение расстояние определяется, собственно минимальным эллипсоидом группировки векторов $\tilde{a}_j \in R^m, j = \overline{1,n}$, тогда как среднее $\bar{a}$ по элементам обучающей выборки задаёт центр соответствующего эллипса.

При наличии набора кластеров $Kl_l, l = \overline{1,L}$, расстояния соответствия определяются, соответственно, соотношениями

$$\rho^2(x, Kl_l) = \frac{1}{r_{l,\max}^2}(x - \bar{a}_l)^T R(\tilde{A}_l^T)(x - \bar{a}_l), l = \overline{1,L} , \tag{17}$$

с очевидной детализацией обозначений в связи с употреблением соответствующего индекса.

Вычисление расстояний до кластеров в соответствии с (17) отвечает расстояниям до представителей кластеров $\bar{a}_l, l = 1, L$ в соответствии с минимальными эллипсами группировки центрированных элементов обучающих выборок.

Примеры эффективного применения расстояний соответствия вида (17) можно найти, например, в работах [Кириченко, Донченко, 2008], [Кириченко, Донченко, 2007], [Донченко, Омардибирова 2005].

Применения базовых свойств линейных структур: RFT- функциональные сети

Искусственные нейронные сети являются стандартным технологическим инструментом исследования в задачах прогноза, классификации и кластеризации. В сущности, такие сети представляют собой графические изображения: графы суперпозиций, стандартизованных функциональных элементов. В этом смысле их с полным правом можно назвать функциональными сетями. Стандартизация функциональных элементов (нейронов) проявляется в том, что они реализуют скалярную функцию векторного аргумента как суперпозицию линейного функционала и скалярной функции скалярного аргумента:

$$y = F(w^T x), w, x \in R^m, F : R^1 \to R^1 .$$

Упомянутая стандартизация - унификация может проявляться и в выборе внешней функции: F, которая называется функцией инициализации нейрона. Она может быть фиксированной или принимать значения из конечного набора функций.

Заметим, что, как правило, фиксированной является и структура сети: количество стандартных функциональных элементов нейронов и способ их соединения: топология сети.

Возможности функциональных сетей можно значительно расширить, если 1) придать большую функциональную универсальность и обеспечить адаптивность в построении каждого из стандартизованных элементов; 2) обеспечить более гибкие возможности в соединении элементов: большую свободу в формировании топологии сети; 3) гарантировать адаптивное построение структуры всей сети в целом. Последнее может быть конструктивно реализовано в ходе выполнения последовательных шагов наращивания сети, имеющих рекуррентный характер.

Реализация такой программы для задачи прогноза - восстановления функции, представленной своими значениями $(x_i, y_i), x_i \in R^n, y_i \in R^m, i = \overline{1,n}$, состоит в построении стандартного функционального элемента(RFT - преобразователя) в виде

$$y = A_+ \Psi(Cx) \ , \tag{18}$$

в котором матрица C – матрица предварительного преобразования вектора признаков x, а Ψ – нелинейное покоординатное преобразование измененного вектора признаков, A_+ - матрица МНК оценивания на выборке $(x_i, y_i), x_i \in R^n, y_i \in R^m, i = \overline{1,n}$.

Эффективность преобразователя (18) в прогнозе зависимости может контролироваться по невязке и по тестовой выборке.

В общем, адаптивная процедура построения функциональной сети развивает идею МГУА А.Г. Ивахненко

Заключение

В работе предложена и обоснована концепция базовых структур евклидового пространства, включая линейные: подпространства, гиперплоскости, линейные операторы, – и нелинейные: неотрицательно определённые квадратичные формы и отвечающие им эллипсоиды группировки. Изложены конструктивные способы описания и взаимного перехода от одних типов структур к другим: от линейных подпространств и гиперплоскостей к матрицам и наоборот, а также от набора векторов к матрицам квадратичных форм и эллипсам группировки и наоборот. В числе других рассмотрены конструктивные способы порождения подпространств и гиперплоскостей, а также ортогональных проекторов, связанных с указанными объектами. В том же русле лежит рассмотрение группирующих операторов. Упомянутая конструктивность обеспечивается применением псевдообращения по Муру – Пенроузу (ПдО), а также новыми результатами в этой области. Важность и эффективность использования приведённых результатов проиллюстрирована на широком спектре задач, включающем линейную регрессию, в том числе векторную, теорию оптимального управления, кластеризацию, прогноз и функциональные сети, являющиеся обобщение искусственных нейронных сетей.

Литература

[Донченко. 2009] Донченко В.С. Неопределённость и математические структуры в прикладных исследованиях/ Human aspects of Artificial Intelligence Inernational Book Series Information science & Computing.– Number 12. – Supplement to International Journal “Information technologies and Knowledge”. –Volume 3.–2009. – P. 9-18.

[Донченко, Омардибирова 2005] Донченко В.С., Омардибирова В.Н. Технология классификации электронных документов с использованием теории возмущения псевдообратных матриц// Proceedings of the XI-th International Conference “Knowledge-Dialogue-Solution”. – June 20-30, Varna, 2005. – Volume 1. – C.223-226.

[Кириченко, Донченко, 2008] В.С Кириченко Н.Ф. Донченко В.С. Гиперплоскости в «множествах и расстояниях соответствия»: кластеризация / Artificial Intelligence and Decision Making.– International book series “INFORMATION SCIENCE&COMPUTING”, Number 7.–Supplement to the International Journal “INFORMATION TECHNOLOGES&COMPUTING” V.2/2008 FOI ITHEA, Sofia 2008.– P. 25-36.

[Moore, 1920] Moore E.H. On the reciprocal of the general algebraic matrix // Bulletin of the American Mathematical Society. – 26, 1920. – P.394 -395.

[Penrose, 1955] Penrose R. A generalized inverse for matrices // Proceedings of the Cambridge Philosophical Society 51, 1955. – P.406-413.

[Алберт, 1977] Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание. – М.: Наука. – 1977.– 305 с.

[Кириченко, 1997] Кириченко Н.Ф. Аналитическое представление псевдообратных матриц //Киб. и СА.- №2. –1997.– С.98-122.

[Кириченко, Донченко,2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв’язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.

[Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации// Киб. и СА.- №4, 2007– С.98-122.

[Кириченко, Донченко, 2008] Кириченко Н.Ф. Донченко В.С. Гиперплоскости в «множествах и расстояниях соответствия»: кластеризация / Artificial Intelligence and Decision Making.– International book series “INFORMATION SCIENCE&COMPUTING”, Number 7.–Supplement to the International Journal “INFORMATION TECHNOLOGES&COMPUTING” V.2/2008 FOI ITHEA, Sofia 2008. – P. 25-36.

Информация об авторах

Владимир С. Донченко *– профессор; Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, e-mail: voldon@unicyb.kiev.ua*

Юрий Г. Кривонос *– академик НАНУ; зам. директора Института кибернетики НАНУ, Украина*

Виктория Н.Омардибирова *– аспирантка; Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина*