ЕВКЛИДОВЫ ПРОСТРАНСТВА ЧИСЛОВЫХ ВЕКТОРОВ И МАТРИЦ: КОНСТРУКТИВНЫЕ МЕТОДЫ ОПИСАНИЯ БАЗОВЫХ СТРУКТУР И ИХ ИСПОЛЬЗОВАНИЕ

Владимир Донченко

Аннотация: *Определены и рассмотрены базовые линейные структуры матричных евклидовых пространств, которые естественным образом расширяют возможности представления моделируемого объекта до «матрицы признаков» и позволяют непосредственным образом учесть динамику исследуемого объекта. «Базовость» структур: линейных и нелинейных, для матричного случая, как и для евклидовых пространств числовых векторов, означает их принципиальное значение для использования в прикладных задачах. Как и в векторном случае, фундаментальную роль в конструктивном построении матричных базовых структур: их описания, взаимного перехода и использования, является аппарат сингулярного представления для линейных операторов между евклидовыми пространствами. Ту же роль в матричном случае сохраняет и псевдообращение по Муру - Пенроузу для линейных операторов на матричных аргументах. В работе представлены результаты, касающиеся сингулярного разложения и псевдообращения по Муру-Пенроузу для матрицы линейных операторов между матричными евклидовыми пространствами. Введены в рассмотрение кортежные объекты и операции над ними, обобщающие понятия классических числовых векторов и блочных матриц ленточного характера. Развита теория группирующих операторов, важная в прикладном отношении как инструмент выявления и использования групповых свойств объектов в евклидовых пространствах, как для пространств числовых векторов, так и для матричных евклидовых пространств. Кроме того, группирующие операторы позволяют, в частности, прояснить алгебраическую суть расстояния Махаланобиса.*

Ключевые слова*: Псевдообращение по Муру - Пенроузу, сингулярное представление матрицы, кластеризация, расстояние Махаланобиса.*

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 развита концепция базовых структур матричных евклидовых пространства, которая переносит представление о базовых структурах евклидовых пространств числовых векторов на матричный случай. Как и в векторном случае, к «базовым структурам» предлагается отнести объекты двух основных видов: линейные и нелинейные, которые могут носить множественный или «сингловый», единичный характер. К множественным структурам следует отнести подпространства и гиперплоскости (линейные структуры), а также эллипсоиды (нелинейные структуры квадратического характера). К «сингловым» структурам, относятся линейные операторы и квадратические формы, а также матрицы, которые их представляют. Отметим, что к базовым нелинейным структурам отнесены структуры, носящие квадратичный характер: связанные и определяемые подходящими неотрицательно определёнными (симметричными) квадратичными формами. Важность множественных вариантов структур евклидовых пространств определяется естественным характером их использования. Так, к примеру, подпространства и гиперплоскости могут использоваться как для представления классов группирования объектов, так и для их дискриминации. Что касается эллипсоидов, то они, к примеру, естественным образом появляются, в описании неопределённостей в задачах минимаксного характера. Кроме того, они естественным образом могут представлять классы группируемых объектов через минимальные эллипсоиды группировки. Отметим, что примечательной особенностью эллипсоидов группировки [Донченко, 2010] является то, что они содержат все векторы набора векторов, и являются в определённом смысле «минимальными» для них. В работе на случай матричных евклидовы пространств перенесены конструктивные методы генерации базовых структур в связи с порождающей их совокупностью объектов. Кроме того, получены и приведены формулы взаимного перехода от одних типов структур к другим: от линейных подпространств и гиперплоскостей к подходящим матрицам и наоборот, а также от набора векторов к матрицам квадратичных форм и эллипсоидов группировки. В числе других рассмотрены конструктивные способы порождения подпространств и гиперплоскостей, а также ортогональных проекторов, связанных с указанными объектами. В том же русле конструктивности лежит рассмотрение концепции группирующих операторов и минимальных эллипсоидов группировки. Концепция группирующих операторов позволяет использовать в прикладных исследованиях связь того или иного набора векторов с важным видом нелинейных структур евклидовых пространств: минимальными эллипсоидами группировки. И в линейном, и в нелинейном случае 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]. Для расширения возможности аппарата, предложено развитие теории псевдообращения на матричные отображения между евклидовыми пространствами матриц фиксированных размерностей со следовыми скалярными произведениями. Кроме того, на случай отображений между евклидовыми матричными пространствами перенесена теорема о сингулярном разложении матрицы оператора, доказана теорема свёртки о соотношении между сингулярными разложениями одной и той же матрицы в векторном и матричном случаях.

Следует отметить, также, что важную роль в применении аппарата псевдообращения играет сингулярное представление (его называют также SVD - представлением) произвольной $m \times n$ матрицы в специфической форме записи в виде взвешенной суммы тензорных произведений специального набора пар векторов с прозрачным геометрическим содержанием. При этом матрица рассматривается как матрица линейного оператора между двумя пространствами числовых векторов: R^n и R^m. Это определяет специфику сингулярного разложения и интерпретацию его составляющих.

Заметим, однако, что одна и та же $m \times n$ матрица A задаёт как линейный оператор $A : R^n \to R^m$ между евклидовыми пространствами числовых векторов, так и линейный оператор между матричными евклидовыми пространствами $A : R^{n\times p} \to R^{m\times p}$ (p -произвольное натуральное). Переход к евклидовым пространствам $R^{n\times p}, R^{m\times p}$ матриц со следовыми скалярными произведениями требует учёта специфики этих евклидовых пространств для передачи принципиальных особенностей SVD-разложения для такого рода объектов. В работе приведено SVD-разложение для матриц, рассматриваемых как линейные операторы между пространствами матриц, доказана теорема свёртки, связывающие два варианта SVD-представления матрицы: векторного и матричного.

Важным в описании нелинейных структур являются «минимальные» эллипсоиды группировки и, как и в линейном случае, конструктивные способы их описания. Говоря о «минимальных» эллипсоидах группировки, будем иметь в виду «минимальный» в определённом смысле эллипсоид, включающий заданный набор векторов. Как оказывается, матрицей квадратичной формы для такого эллипсоида является группирующий оператор, конструктивно описываемый средствами псевдообращения. Квадратичная форма, отвечающая группирующему оператору, в искажающем суть такого оператора виде, фигурирует в «статистическом» варианте расстояния Махаланобиса [Mahalanobis,1936] (см. также, например, [McLachlan, Geoffry, 1992]), когда вместо ковариационной матрицы многомерного нормального распределения используется её стандартная оценка на основе выборки векторов $a(1),...,a(n)$.

В заключение отметим, что основные идеи, дух и результаты предлагаемой работы восходят к работам и используют результаты развитой в них теории псевдообращения нашего безвременно ушедшего коллеги, друга и учителя, профессора Н.Ф. Кириченко.

Базовые линейные структуры и связи между ними

В дальнейшем, говоря о евклидовом пространстве R^p, будем иметь в виду множество конечных числовых последовательностей одной и той же длины p, 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^p, а его элементы – через $a, a^T = (a_1, ..., a_p)$. 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 $e(j) \in R^m, j = \overline{1,m}, e_{(i)} \in R^n, i = \overline{i,n}$, а для общего случая - $e_k \in R^p, k = \overline{1,p}$. Оператор A из R^n, в R^m: $A : R^n \to R^m$, задаваемый 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 в этих базисах. Для матрицы $A = (a_{ij})$ будем использовать также блочное представление по столбцам (столбцовое) и строкам (строчное):

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

Пространство всех $p \times q$ матриц с покоординатным умножением и сложением, а также «следовым» скалярным произведением будем обозначать $R^{p \times q}$. «Следовое» скалярное произведение $(\cdot,\cdot)_{tr}$ двух матриц определяется соотношением

$$(C\cdot, D\cdot)_{tr} = trC^T D = \sum_{i=1}^{q} \sum_{k=1}^{p} c_{ki} d_{ki}, C = (c_{ij}), D = (d_{ij}) \in R^{p \times q}.$$

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, соответственно, для A^T - через L_{A^T}.

То же самое будет относиться к линейным подпространствам матричных евклидовых пространств: через $L(C_k, k = \overline{1,K}) \equiv L(C_1 ..., C_K) \subseteq R^{p \times q}, C_k \in R^{p \times q}, k = \overline{1,K}$ будем обозначать линейное подпространство в $R^{p \times q}$, порождённое набором матриц $C_k \in R^{p \times q}, k = \overline{1,K}$.

Соответственно, через $\Gamma(c, L) = c + L, \Gamma(C, L) = C + L$ в вектором или матричном случае будем обозначать гиперплоскость в соответствующем евклидовом пространстве.

В работе [Донченко, 2010] представлены основные формулы, устанавливающие связь между множественными и сингловыми линейными структурами для евклидовы пространств числовых векторов. Отметим, что аналоги этих формул справедливы и для матричных объектов с достаточно очевидной заменой стандартных единичных ортов в R^n или R^m на аналоги, каковыми являются блочные матрицы, составленные из нулевых и единичных матриц подходящей размерности. Заметим также, что блочные матрицы: строчные или столбцовые, - составленные из блоков одинаковой размерности представляют собой, собственно, строчные или столбцовые кортежи с матричными элементами. Явно вводя длину кортежа в его наименование, будем говорить об n - кортежах строках или n -кортежах столбцах с компонентами из соответствующего множества. Так числовой вектор-столбец из R^n является -кортежем с числовыми элементами, а упорядоченный набор $(C_1...,C_n), C_i \in R^{p\times q}, i=\overline{1,n}$ является n -кортежем-строкой с элементами из матричного евклидова пространства $R^{p\times q}$. Возможны операции с разнородными кортежными конструкциями, обобщающие матричные операции. Так, например можно говорить о произведении двух n - кортежей: строчного и стобцового, определяемого через сумму произведений компонент. Смысл операции произведения в каждом конкретном случае определяется природой компонент кортежей. Так для строчного - кортежа $(C_1...,C_n), C_i \in R^{p\times q}, i=\overline{1,n}$ с матричными компонентами и числового столбцового - кортежа $b: b^T=(b_1...,b_n), b_i \in R^1, i=\overline{1,n}$ (числового вектора $b \in R^n$) произведение $(C_1...,C_n)b$ определяется как

$$(C_1...,C_n)b=\sum_{i=1}^{n} C_i b_i \ .$$

Если же числовые компонент у столбцового n -кортежа b заменить матричными:

$$b=\begin{pmatrix} B_1 \\ \cdots \\ B_n \end{pmatrix}, B_i \in R^{q\times p}, i=\overline{1,n},$$

То: $$(C_1...,C_n)b=\sum_{i=1}^{n} C_i B_i \ .$$

Точно таким же образом можно говорить об блочных матрицах с элементами из подходящих евклидовых пространств и умножении слева или справа таких матриц на строчные или столбцовые кортежи.

Заметим, что вместе с кортежными операциями могут быть определены и обычные матричные. Как правило, результат выполнения кортежных и матричных операций в случае их одновременного рассмотрения, является одинаковыми.

В рамках определённых таким образом операций, обобщающих матричные на кортежные, можно, например, утверждать, что справедливо следующее соотношение, являющееся обобщением теоремы о представлении произведения матриц суммой тензорных произведений компонент сомножителе [Донченко, 2010].

1. *«Тензорное произведение» строчного и столбцового кортежей*. Для произведения матричных - кортежей B, C : строчного и столбцового

$$B=(B(1)\vdots...\vdots B(r)), B(j) \in R^{p\times q}, j=\overline{1,r}, C^T=\begin{pmatrix} C_{(1)}^T \\ \cdots \\ C_{(r)}^T \end{pmatrix}, C_{(i)} \in R^{q\times k}, i=\overline{1,r} \qquad (1)$$

а также для диагональной матрицы $\Lambda = diag(\lambda_1,...,\lambda_r), \lambda_i \in R^1, i=\overline{1,r}$ справедливо соотношение

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

В упомянутой выше работе [Донченко, 2010] приведен вариант SVD-представления матрицы оператора между двумя евклидовыми пространствами числовых векторов в форме, допускающей его прямое перенесение на матрицы произвольной размерности, рассматриваемые как матрицы линейных отображений между матричными евклидовыми пространствами. Этот вариант представления приведен ниже под номером 2.

2. *Вариант SVD- представление произвольной $m \times n$ матрицы.* Для произвольной $A \in R^{m \times n}$ ранга $r \le \min(m,n)$, рассматриваемой как матрица линейного отображения между евклидовыми пространствами числовых векторов: $A : R^n \to R^m$, справедливо следующее представление матрицы в виде взвешенной суммы тензорных произведений векторов

$$Ax = \sum_{i=1}^{r} \lambda_i u_i (v_i, x)_{R^n}, x \in R^n \qquad (2)$$

где

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

3. *Комментарий к п.2.* Собственно, действие оператора A на вектор $x \in R^n$ в варианте (2) SVD-разложения, представляет собой перенос разложения вектора по ортогональному базису или его части $v_i \in R^n, i = \overline{1,r}$ с координатами $x_i = v_i^T x = (v_i, x)_{R^n}, i = \overline{1,r}$, в одном пространстве в разложение по ортогональному базису $u_i \in R^m, i = \overline{1,r}$, или - его части, в другом пространстве с умножением координат на положительные числа, соответственно, $\lambda_i \in R^m, i = \overline{1,r}$.

Ортогональные проекторы (в дальнейшем ОП) являются важной составляющей аппарата конструктивного описания и использования линейных структур в матричном варианте евклидовых пространств, как в случае евклидовых пространств числовых векторов. И в этом, более общем варианте евклидовых пространств, общей, основой их эффективного использования, является псевдообращение по Муру Пенроузу. Напомним, что в определении по Муру [Moore, 1920] прямо связывается с двумя основными линейными подпространствами линейного оператора: его множеством значений и его ядром. Ортогональных проекторов, как и возможности их конструктивного построения в связи с линейными подпространствами, является наличие двух эквивалентных определений таких проекторов, а также возможность их описание через псевдообращение. Для матрицы, рассматриваемой как линейный оператор между матричными евклидовыми пространствами, приведённое выше соображение остаётся в полной мере справедливым.

Матрицы как линейные операторы между матричными евклидовыми пространствами

4. *Матрица $A \in R^{m \times n}$ как линейный оператор из матричного евклидового пространства $R^{n \times p}$ в $R^{m \times p}$.* Произвольная $m \times n$ матрица A может рассматриваться как матрица линейного оператора между двумя евклидовыми пространствами матриц $R^{n \times p}$ и $R^{m \times p}$ со следовым скалярным произведением. Этот оператор описывается стандартным образом: для произвольной матрицы $X \in R^{n \times p}$ результатом действия оператора является матрица $AX \in R^{m \times p}$.

5. *Матричное SVD-разложение $m \times n$ матрицы $A \in R^{m \times n}$ (M-SVD).*

Теорема 1. Справедливо следующее представление матрицы в виде суммы, реализующий перенос разложения по ортонормированным собственным элементам (собственным матрицам) матрицы $A^T A$ на ортонормированные собственные элементы матрицы $A^T A$

$$A \Rightarrow X \sum_{i=1}^{r} \sum_{k=1}^{p} \lambda_i u_i e_k^T (v_i e_k^T, X)_{tr} \tag{3}$$

где

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

6. Совпадение векторного и матричного варианта SVD: теорема свёртки.

Теорема 2. Векторный и матричный вариант SVD-разложения одной и той же $m \times n$ матрицы $A \in R^{m\times n}$ совпадают между собой: для произвольной матрицы $X \in R^{n\times p}$

$$\left(\sum_{i=1}^{r} \lambda_i u_i v_i^T\right) X = \sum_{i=1}^{r} \sum_{k=1}^{p} \lambda_i u_i e_k^T (v_i e_k^T, X)_{t\,r} \tag{4}$$

Как отмечалось выше, принципиальную роль в описании базовых структур евклидовых пространств играет псевдообращение по Муру – Пенроузу [Moore,1920], [Penrose, 1955] (в дальнейшем ПдО) как одноместной операции A^+ над матрицами произвольной размерности

С учётом п.5 необходимо различать ПдО для матрицы, как векторного отображения, и ПдО для той же матрицы, как матричного отображения. Первое, собственно, классическое по Муру-Пенроузу, будем обозначать как ВПдО, второе – как МПдО.

Приведём для сравнения вариант классического SVD-разложения в тензорной записи. В обозначениях, используемых выше, оно имеет вид

$$A = \sum_{i=1}^{r} \lambda_i u_i v_i^T \tag{5}$$

Псевдообращение: векторный и матричный вариант отображения

7.Определение МПдО через матричное SVD - представление матрицы.

Теорема 3. Для произвольной $m \times n$ - матрицы A её МПдО A_M^+ определяется соотношением

$$A_M^+ X == \sum_{i=1}^{r} \sum_{k=1}^{p} \lambda_i^{-1} v_i e_k^T (u_i e_k^T, X)_{tr} : R^{m\times p} \to R^{n\times p}. \tag{6}$$

8.*Совпадение векторного (классического) ПдО и ПдО матричного*.

Теорема 4. Матричный и векторный вариант ПдО для произвольной $m \times n$ - матрицы A совпадают между собой: т.е. совпадают операторы A^+, A_M^+, определяемые соотношениями (5), (6) соответственно: $A_M^+ = A^+$

.

Доказательство. Результат очевидным образом вытекает из теоремы свёртки п.12.

Евклидовы пространства, базовые линейные структуры и ПдО

9. *Основные ортогональные проекторы (ОП), связанные с матрицей линейного оператора*. Основными ОП, связанными с матрицей A являются ОП $P(A^T), P(A)$ на подпространства L_A, L_{A^T} соответственно, а также $Z(A^T), Z(A)$ соответственно - на их ортогональные дополнения в R^m, R^n соответственно. Они определяются соотношениями:

$$P(A^T) = AA^+ = \sum_{i=1}^{r} u_i u_i^T, P(A) = P\left((A^T)^T\right) = A^T (A^T)^+ = A^+ A = \sum_{i=1}^{r} v_i v_i^T , \quad (7)$$

$$Z(A) = E_n - P(A) = E_n - A^+ A , Z(A^T) = E_m - P(A^T) = E_m - A^{T^+} A^T = E_m - AA^+ . \quad (8)$$

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

Теорема 5. Основными ОП, связанными с матрицей A, как матрицей линейного оператора над матричными евклидовыми пространствами, являются ОП, определяемые соотношениями (7),(8) и являющимися ОП соответственно на $L_A, KerA$.

10. Исследование матричного линейного алгебраического уравнения (МЛаУ).

Теорема 6. Для того чтобы МЛаУ $AX = Y, A \in R^{m \times n} X \in R^{n \times p}, Y \in R^{m \times p}$ было разрешимо необходимо и достаточно, чтобы $trY^T Z(A^T)Y = 0$. В этом случае множество решений Ω_Y определяется соотношением

$$\Omega_Y = \{X : X = A^+ Y + Z(A)V, V \in R^{n \times p}\} . \quad (9)$$

В случае, когда МЛаУ несовместно, т.е. когда $trY^T Z(A^T)Y > 0$,множество (9) описывает решение оптимизационной задачи наилучшего квадратического приближения правой части Y значениями левой части AX того же уравнения:

$$\underset{X \in R^{n \times p}}{Arg} \min \| AX - Y \|_{tr}^2 = \Omega_Y = \{X : X = A^+ Y + Z(A)V, V \in R^{n \times p}\} . \quad (10)$$

Величина невязки для каждого решения оптимизационной задачи составляет $trY^T Z(A^T)Y$.

*Замечание 1.*Обратим внимание, что множество решений МЛаУ, и множество псевдорешений, когда МЛаУ несовместно, описываются одним и тем же соотношением (9), (10).

Замечание 2. Поскольку при совместности МЛаУ $\min\limits_{X \in R^{n \times p}} \| AX - Y \|_{tr}^2 = 0$ и этот минимум достигается на решениях МЛаУ, то множество Ω_Y из (9) описывает и множество решений, и множество псевдорешений как решение одной и той же оптимизационной задачи

$$\Omega_Y = \{X : X = A^+ Y + Z(A)V, V \in R^{n \times p}\} = \underset{X \in R^{n \times p}}{Arg} \min \| AX - Y \|_{tr}^2 \quad (11)$$

Замечание 3. Обратим внимание, что матрица $A^+ Y$ является решением, точным или псевдо -, МЛаУ. Это решение будем называть базовым.

11. Оптимизационное свойство Пенроуза для матричного ПдО.

Теорема 7. Базовое решение МЛаУ как в точном, так и в псевдоварианте, является единственным наименьшим по норме решением оптимизационной задачи поиска наилучшего квадратичного приближения правой части МЛаУ значениями левой части того же уравнения:

$$A^+ Y = \underset{X \in \underset{X \in R^{n \times p}}{Arg} \min \| AX - Y \|_{tr}^2}{Arg} \min \| X \|_{tr} . \quad (12)$$

Доказательство. В силу замечания 2 и множество решений, и множество псевдорешений являются решением одной и той же оптимизационной задачи. Поэтому для любого $X \in \underset{x \in R^n}{Arg} \min \| AX - Y \|_{tr}^2$ в соответствии с (10) справедливо представление

$$X = A^+Y + Z(A)V, V \in R^{n \times N}. \quad (13)$$

Поскольку $Z(A)V \in Z(A)R^{n \times p}$ и, как нетрудно убедиться, для любого $V \in R^{n \times p}$

$$A^+Y \perp_{tr} Z(A)V,$$

то в соответствии с многомерным вариантом теоремы Пифагора

$$\| X \|_{tr}^2 = \| A^+Y + Z(A)V \|_{tr}^2 = \| A^+Y \|_{tr}^2 + \| Z(A)V \|_{tr}^2 \geq \| A^+Y \|_{tr}^2.$$

Таким образом, для произвольного $X \in \underset{X \in R^{n \times p}}{Arg} \min \| AX - Y \|_{tr}$

$$\| X \|_{tr}^2 \geq \| A^+Y \|_{tr}^2,$$

и базовое решение A^+Y, псевдо - или точное, является наименьшим по норме решением оптимизационной задачи, определяемой правой частью соотношения (12). Единственность решения оптимизационной задачи вытекает из того, что $\| Z(A)V \|^2 = 0 \Leftrightarrow Z(A)V = 0$.

12. В связи с ограниченностью работы только упомянем другие важные в применениях результаты, которые приведены в работе [Донченко et al, 2010]. Это прямые и обратные формулы Гревиля, а также формулы аналитического возмущения ПдО Н.Ф.Кириченко [Кириченко,1997]. В сущности, в этих результатах речь идёт о формулах, связывающих ПдО изменённой (возмущённой) матрицы с ПдО исходной, а также характеристиками возмущения. В формулах Гревиля таким возмущением является добавление (прямые) или вычёркивание (обратные) строки или столбца матрицы. Собственно, Гревилю принадлежит формула, касающаяся добавления строки, да ещё и только для случая независимости добавляемой строки от строк исходной матрицы. Результаты, связанные с добавлением зависимой строки, как и с вычёркиванием произвольного типа строки, принадлежат Н.Ф.Кириченко [Кириченко,1997].

В аналитических формулах возмущения [Кириченко,1997] изменение исходной матрицы происходит аддитивно: через добавление «простейшей» матрицы. В качестве «простейшей» матрицы выступает тензорное произведение ab^T двух векторов $a \in R^m, b \in R^n$. Как и в случае формул Гревиля, вид ПдО возмущённой матрицы определяется тем, являются ли компоненты возмущения зависимыми или независимыми от, соответственно, столбцов и строк возмущаемой матрицы. Кроме того, результат зависит также и от того, падает или сохраняется ранг возмущённой матрицы, когда составляющие элементы возмущения зависят от соответствующих составляющих возмущаемой матрицы. Обратим внимание, что все формулы или условия описываются явными аналитическими выражениями. Условия линейной зависимости или независимости для столбцов уже нашли своё отражение в п.22. Для описания зависимости от строк в соответствующем условии необходимо только заменить A^T на A.

Что же касается условия падения или сохранения ранга возмущённой матрицы, когда вектор a зависит от столбцов, а вектор b^T - от строк A, то они имеют вид:

$$b^T A^+ a \neq -1 \text{ сохраняется, } b^T A^+ a = -1 \text{ падает.}$$

13. Квадрат расстояния матрицы до гиперплоскости.

Теорема 8. Квадрат расстояния $\rho^2(C, \Gamma(B, L_A))$ матрицы $C \in R^{m \times p}$ от гиперплоскости $\Gamma(B, L_A) = B + L_A$ определяется соотношением

$$\rho^2(C,\Gamma(B,L_A)) = \min_{Y\in\Gamma(B,L_A)} \| C-Y \|_{tr}^2 = tr(C-B)^T Z(A^T)(C-B).$$

Базовые нелинейные структуры: группирующие операторы

Важнейшими нелинейными структурами евклидова пространства являются квадратичные формы (в работе - неотрицательно определённые) и отвечающие им эллипсоиды или эллипсоидальные цилиндры. Среди таких нелинейных структур принципиальными являются матрицы так называемых «группирующих операторов», которые естественным образом связаны с групповыми свойствами набора векторов. Группирующие операторы возникают в связи с набором векторов $a(j)\in R^m, j=\overline{1,n}$ и отвечающей ему матрицей $A=(a(1)\vdots\ldots\vdots a(n))$. Как и ортогональные проекторы, группирующие операторы являются парными. Будем обозначать их, соответственно $R(A), R(A^T)$. В дальнейшем рассмотрение свойств группирующих операторов будет проводиться в связи с евклидовыми пространствами числовых векторов и переноситься на случай отображений между матричными евклидовыми пространствами

14.*Определение группирующих операторов*. Группирующие операторы, обозначаемые $R(A), R(A^T)$, и для векторного и для матричного варианта действия матрицы A определяются соотношениями

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

Их важность для практики и область применения раскрываются в свойствах, приведённых ниже.

15.*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 $a(j)\in R^m, j=\overline{1,n}$ с матричным представлением $A=(a(1)\vdots\ldots\vdots a(n))$ и произвольного нормированного вектора $u\in R^m : \| u \| = 1$, справедливо равенство:

$$\sum_{j=1}^{n} a^T(j)uu^T a(j) = u^T AA^T u \tag{14}$$

Доказательство. Действительно, принимая во внимание связь п.2 векторов набора $a(j)\in R^m, j=\overline{1,n}$ со своим матричным представлением: $a(j)=Ae_{(j)}, j=\overline{1,n}$, имеем:

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

Остаётся только заметить, что $\sum_{j=1}^{n} e_{(j)}e_{(j)}^T = E_n$, где E_n - единичная матрица в R^n.

Замечание 4. Левая часть соотношения (14) леммы 1 представляет собою сумму квадратов проекций векторов набора $a(j)\in R^m, j=\overline{1,n}$ на нормированный вектор $u\in R^m : \| u \| = 1$.

16.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 $a(j)\in R^m, j=\overline{1,n}$, с матричным представлением $A=(a(1)\vdots\ldots\vdots a(n))$ имеет место соотношение:

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

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

17. *Группирующие операторы: эллипсоиды группировки набора векторов* $a(j) \in R^m, j=\overline{1,n}$. Основное утверждение пункта – теорема 6 ниже.

Теорема 9. Пусть $a(j) \in R^m, i=\overline{1,n}$ произвольный набор векторов из R^n с матричным представлением $A = (a(1) \vdots \ldots \vdots a(n)), rankA = r \le \min(m,n)$. Тогда все векторы набора принадлежат внутренности эллипсоида, точнее: эллипсоидального цилиндра, определяемого уравнением

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

где, $R(A^T)$, группирующий оператор: $R(A^T) = A^{+^T} A^+$.

Доказательство. Рассмотрим квадраты проекций векторов набора $a(j), j=\overline{1,n}$, на каждый из векторов $u_i, i=\overline{1,r}$, SVD-представления (5) матрицы A. Принимая во внимание ортонормированность набора $u_i, i=\overline{1,r}$, и обозначая квадраты проекций через $\left\|\Pr_{u_i} a(j)\right\|^2, i=\overline{1,r},\ j=\overline{1,n}$, очевидным образом имеем

$$\left\|\Pr_{u_i} a(j)\right\|^2 = a^T(j)u_iu_i^Ta(j), j=\overline{1,n}, i=\overline{1,r} .$$

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

$$\sum_{j=1}^{n} a^T(j)u_iu_i^Ta(j) = u_i^T AA^T u_i = \lambda_i^2, i=\overline{1,r} .$$

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

$$\sum_{j=1}^{n} \frac{a^T(j)u_iu_i^Ta(j)}{\lambda_i^2} = \frac{u_i^T AA^T u_i}{\lambda_i^2} = 1, i=\overline{1,r} ,$$

т.е.

$$\sum_{j=1}^{n} \frac{a^T(j)u_iu_i^Ta(j)}{\lambda_i^2} = 1, i=\overline{1,r} .$$

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

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{a^T(j)u_iu_i^Ta(j)}{\lambda_i^2} = \sum_{i=1}^{m} \frac{u_i^T AA^T u_i}{\lambda_i^2} = r .$$

Поменяв порядок суммирования в двойной сумме, получаем

$$\sum_{j=1}^{n}\sum_{i=1}^{m} \frac{a^T(j)u_iu_i^Ta(j)}{\lambda_i^2} = \sum_{j=1}^{n} a^T(j) \sum_{i=1}^{m} \frac{u_iu_i^T}{\lambda_i^2} a(j) = r .$$

Далее, приняв во внимание, что

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

получаем окончательно

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

т.е.

$$\sum_{j=1}^{n}a^T(j)R(A^T)a(j)=r\,. \quad (15)$$

Поскольку $R(A^T)$ – симметричная, неотрицательно определённая матрица, то следствием соотношения (15) является одновременное выполнение неравенств

$$a^T(j)R(A^T)a(j)\le r,\, j=\overline{1,n}\,. \quad (16)$$

Та же симметричность и неотрицательная определённость позволяет сделать вывод, что уравнение

$$x^TR(A^T)x=r,x\in R^m \quad (17)$$

определяет эллипсоид, 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 (16) для всех векторов набора $a(j)\in R^m, j=\overline{1,n}$, означает их одновременную принадлежность внутренности эллипсоидального цилиндра с уравнением (17), и доказательство теоремы завершено.

Замечание 5. В действительности неравенство (16) может давать существенное загрубление «радиуса» эллипсоида. Так, при векторах $a(j)\in R^m, j=\overline{1,n}$, близких к ортогональным, очевидным образом, константу в правой части (16) можно выбрать близкой к 1.

18. *Группировка матричных объектов*.

Теорема 10. Пусть $A(j)\in R^{m\times p},i=\overline{1,n}$ произвольный набор общим рангом r матриц из $R^{m\times p}$, и $\Re$ строчный - кортеж из указанных элементов: $\Re=(A(1),...,A(n))$ Тогда все векторы набора принадлежат внутренности эллипсоида, точнее: эллипсоидального цилиндра, определяемого уравнением

$$trX^TR(\Re^T)X=r,X\in R^{m\times p}\,,$$

где, $R(A^T)$, группирующий оператор: $R(\Re^T)=\Re^{+T}\Re^{+}$.

19. *Усиление результата об эллипсоидах группировки для числовых векторов.*

Теорема 11. Все векторы набора $a(j)\in R^m,i=\overline{1,n}$, с матричным представлением $A=\left(a(1)\vdots\ldots\vdots a(n)\right)$ принадлежат внутренности эллипсоидального цилиндра

$$x^TR(A^T)x=r_{\max}^2, r_{\max}^2\le r=rankA, x\in R^m\,, r_{\max}^2=\max_{j=\overline{1,n}}a^T(j)R(A^T)a(j)\,,$$

который будем называть минимальным эллипсоидом группировки для рассматриваемого набора.

Группирующие операторы и расстояние Махаланобиса

Группирующие операторы с точностью до скалярного множителя неожиданным образом оказываются связанными с расстоянием Махаланобиса [Mahalanobis,1936] (см. также, например, [McLachlan, Geoffry, 1992]), и проясняют его суть. Напомним, что расстояние Махаланобиса (в дальнейшем РаМ) является одним из способов определения степени принадлежности $\rho^2(x,N)$ неслучайного вектора $x \in R^m$ многомерному $N(a,B)$ нормальному распределению, и определяется соотношением

$$\rho^2(x,N) = (x-a)^T B^{-1}(x-a). \tag{18}$$

Напомним также, что $a \in R^m$ является математическим ожиданием, а симметричная неотрицательно определённая матрица $B : B \in R^{m \times m}, B^T = B, B \geq 0$, является матрицей ковариаций распределения. Неслучайный вектор $x \in R^m$ интерпретируется как возможный элемент выборки.

Как правило, при использовании РаМ предполагается, что наблюдаемое значение может относиться к одному из нескольких $N_k, k = \overline{1,K}$ нормальных распределений $N(a_k, B_k), k = \overline{1,K}$. В этом случае для каждого из возможных распределений определяется своё РаМ. В этом случае рассматривается набор расстояний вида (18):

$$\rho^2(x,N_k) = (x-a_k)^T B_k^{-1}(x-a_k)\,, \tag{19}$$

а отнесение вектора $x \in R^m$ к одному из K возможных нормальных распределений осуществляется по миниммуму РаМ'ов из (19).

Отметим для сравнения, что средний квадрат расстояния от неслучайного вектора $x \in R^m$ до $N(a,B)$ распределённой случайной величины ξ определяется выражением:

$$\mathrm{M}\,\|\xi - x\|^2 = trB + \|x-a\|^2\,,$$

а квадратичная форма $(x-a)^T B(x-a)$ описывает дисперсию случайной величины $(x-a)^T \xi$.

Таким образом, квадратичная форма (18) не является адекватной с точки зрения описания естественных квадратических характеристик нормально распределённого случайного вектора на основе среднего. Единственным обоснованием использования расстояния Махаланобиса яаляется присутствие квадратичной формы из (18) в плотности распределения $f(x), x \in R^m$, определяемой соотношением

$$f(x) = (2\pi)^{-\frac{m}{2}}\,|B|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x-a)^T B^{-1}(x-a)\}, x \in R^m\,.$$

Заканчивая знакомство с РаМ, отметим, что при его практическом использовании вместо параметров a, B используются их оценки $\hat{a}, \hat{B}$ на основе выборки $a(1),...,a(n) \in R^m$, определяемые соотношениями

$$\hat{a} = \bar{a} = \frac{1}{n}\sum_{i=1}^{n} a(i)\,, \hat{B} = \frac{1}{n}\sum_{i=1}^{n}(a(i)-\bar{a})^T(a(i)-\bar{a})\,. \tag{20}$$

В этом случае РаМ приобретает «статистическое» представление

$$\hat{\rho}^2(x,N) = (x-\hat{a})^T \hat{B}^{-1}(x-\hat{a}). \tag{21}$$

Теорема 12. Статистический вид РаМ может быть представлен в виде

$$\hat{\rho}^2(x,N) = n(x-\bar{a})^T R(\tilde{A}^T)(x-\bar{a})\,,$$

где $\tilde{A} = (a(1)-\bar{a} \vdots ... \vdots a(n)-\bar{a})$.

Доказательство. Действительно, из (20) непосредственно следует, что $\hat{a} = \bar{a}$. Из тех же соотношений с использованием результата п.3 вытекает, что

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n}(a(i)-\bar{a})^T(a(i)-\bar{a}) = \frac{1}{n}\tilde{A}\tilde{A}^T \tag{22}$$

Если матрица $\hat{B}$ невырожденная и определяется соотношением (22), то $\hat{B}^{-1} = n(\tilde{A}\tilde{A}^T)^{-1}$. Из соотношения (5) для векторного SVD-разложения для матрицы $\tilde{A}$ вытекает, что

$$\tilde{A}\tilde{A}^T = \sum_{i=1}^{n} \lambda_i^2 u_i^T u_i \ . \tag{23}$$

Из (23) в свою очередь следует, что матрица $(AA^T)^{-1}$, если она существует, определяется соотношением

$$(\tilde{A}\tilde{A}^T)^{-1} = \sum_{i=1}^{m} \frac{u_i^T u_i}{\lambda_i^2} . \tag{24}$$

Как нетрудно убедиться, правая часть в (24) совпадает с $R(\tilde{A}^T)$. Таким образом, окончательно имеем:

$$\hat{B}^{-1} = nR(\tilde{A}^T) \ ,$$

что и завершает доказательство теоремы.

Очевидна, таким образом, эвристическая основа расстояния Махаланобиса и его жёсткая теоретико-вероятностная привязка. Последнее означает, что вне теории вероятностей и математической статистики и вне связи с многомерным нормальным распределением говорить о расстоянии Махаланобиса не имеет смысла.

В то же время группирующие операторы предоставляют возможности выявления групповых свойств векторов на основе минимальных эллипсоид группировки в любой ситуации.

Применения: кластеризация

Примеры применения базовых линейных структур в линейной регрессии, в линейных системах управления с дискретным временем, а также для специального класса функциональных сетей, обобщающих искусственные нейронные сети, можно найти в работах [Кириченко, Донченко,2005] , [Донченко et al 2010].

Отметим дополнительные возможности, появляющиеся в связи с использованием базовых линейных и нелинейных структур для отображений между матричными евклидовыми пространствами. Отметим, что, как и в случае евклидовых пространств числовых векторов, для представлений классов матричных объектов могут использоваться как гиперплоскости, так и эллипсоиды группировки. Заметим. Что использование матричных объектов как «векторов признаков» существенно расширяет границы применимости алгоритмов кластеризации на основе гиперплоскостей или эллипсоидов группировки. Действительно, использование матричных объектов позволяет непосредственно учесть динамику развития объекта через объединение в одну матрицу классических векторов признаков для некоторого дискретизированного интервала времени. Классическим примером представления исследуемого объекта «матричным вектором признаков» является спектрограмма. Напомним, что спектрограмма является матрицей (или её изображением) последовательных наборов энергетических спектров частей звукового сигнала в окне фиксированной длины. Набор энергетических спектров соответствует последовательности окон, которые получаются последовательным и равномерным сдвигом от начала звукового сигнала к его концу. Заметим, что классическое цифровое изображение также является матрицей.

Заключение

В работе предложена и обоснована концепция базовых структур евклидового пространства, как линейных, так и нелинейных. Изложены конструктивные способы описания и взаимного перехода от одних типов структур к другим. В числе других рассмотрены конструктивные способы порождения, описания и использования базовых структур. Упомянутая конструктивность обеспечивается применением как классических результатов ПдО, так и новыми результатами в этой области. В частности, в работе приведена теорема о сингулярном разложении матрицы как линейного оператора между матричными пространствами, доказана теорема свёртки, обеспечивающая эквивалентность векторного и матричного сингулярных и ПдО-матриц на основе соответствующих сингулярных разложений. Рассмотрены применения полученных результатов для построения кластеризации с разными вариантами расстояний соответствия. Введены в рассмотрение кортежные объекты и операции над ними, обобщающие понятия классических числовых векторов и блочных матриц ленточного характера.

Литература

[Донченко, 2010] Донченко В.С. Евклидовы пространства: конструктивные методы описания базовых структур и их использование / Information Models of Knowledge.- Editors: Krassimir Markov, Vitalii Velichko, Oleksy Voloshin - ITHEA.- Kiev, Ukraine – Sofia, Bulgaria. Number 19.– 2010.- P. 362-376.ISBN 978-954-16-0048-1.

[Донченко et al, 2010] Донченко В.,Кривонос Ю., Омардибирова В. Базовые структуры евклидовых пространств: конструктвные методы описания и использования/ New Trends in Classification and Data Mining. – ITHEA, Sofia, Bulgaria. -2010.- ISBN 978-954-16-0042-9.- P. 155-170.

[Донченко. 2009] Донченко В.С. Неопределённость и математические структуры в прикладных исследованиях/ Human aspects of Artificial Intelligence International 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.– 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.– Sofia 2008. – P. 25-36.

[Mahalanobis,1936] Mahalanobis, P C. On the generalized distance in statistics.//Proceedings of the National Institute of Sciences of India.- 1936.-2 (1).-P. 49–55.

[McLachlan, Geoffry, 1992] McLachlan, Geoffry J. Discriminant Analysis and Statistical Pattern Recognition. Wiley Nescience.-1992.- ISBN 0471691151.

Информация об авторе

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