Матрицы и основные матричные операции

1. Определение матриц

Матрицей размера $n \times m$ называется прямоугольная таблица специального вида, состоящая из n строк и m столбцов, заполненная числами. Матрица обычно обозначаются заглавными буквами латинского алфавита, например A. Элементы матрицы A обозначаются a_{ij}, где i и j — номер строки и столбца, где расположен этот элемент, соответственно. Пространство матриц $n \times m$ обозначается $\mathbb{R}^{n \times m}$.

Матрицы можно использовать для работы с системами линейных алгебраических уравнений. Например, в задачах линейной классификации решается система линейных алгебраических уравнений с матрицей объект-признак относительно вектора неизвестных параметров. При этом получившаяся система уравнений может как иметь бесконечное число решении, так и не иметь решений вовсе. Подробнее эти вопросы будут обсуждаться в курсе по машинному обучению.

Для матриц размера $m \times n$ результат операции умножения на вектор-столбец размера n (т.е. на матрицу размера $n \times 1$) есть новый вектор-столбец размера m:

$$Aw = \begin{pmatrix} \sum_{i=1}^n a_{1i} w_i \\ \sum_{i=1}^n a_{2i} w_i \\ \dots \\ \sum_{i=1}^n a_{mi} w_i \end{pmatrix}$$

Другими словами, можно сказать, что матрица задает линейное отображение.

Системы линейных уравнений лаконично записываются с помощью матриц, например:

$$\begin{cases} 12w_1 + 7w_2 + 21w_3 + 31w_4 + 11w_5 = 1 \\ 45w_1 - 2w_2 + 14w_3 + 27w_4 + 19w_5 = 0 \\ -3w_1 + 15w_2 + 36w_3 + 71w_4 + 21w_5 = 0 \\ 4w_1 - 13w_2 + 55w_3 + 34w_4 + 15w_5 = 1 \end{cases}$$

в матричной записи имеет вид:

$$\begin{pmatrix} 12 & 7 & 21 & 31 & 11 \\ 45 & -2 & 14 & 27 & 19 \\ -3 & 15 & 36 & 71 & 21 \\ 4 & -13 & 55 & 34 & 15 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

2. Матричные операции

Произведение матриц

Ранее уже было сказано, как производится умножение матрицы размера $m \times n$ на вектор-столбец длины n:

$$A \cdot w = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n a_{1i} w_i \\ \sum_{i=1}^n a_{2i} w_i \\ \dots \\ \sum_{i=1}^n a_{mi} w_i \end{pmatrix}.$$

При этом следует отметить, что такая операция возможна только тогда, когда число столбцов матрицы совпадает с длиной вектора. Иначе операция не определена. Можно определить более общую операцию умножения матриц следующим образом.

Пусть $A \in \mathbb{R}^{m \times n}$ и $B \in \mathbb{R}^{n \times k}$ — две матрицы соответствующих размеров (число столбцов в A совпадает с числом строк во B). Тогда произведением матрицы A на матрицу B называется матрица C размера $m \times k$:

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mk} \end{pmatrix}, \qquad c_{ij} = \sum_{p=1}^n a_{ip} b_{pj}.$$

Пример: произведение матриц

Заданы две матрицы согласованного размера:

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 10 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

тогда их произведение:

$$A \cdot B = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 0 & 1 \cdot 0 + 2 \cdot 0 & 1 \cdot 0 + 2 \cdot 2 \\ 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 0 + 1 \cdot 0 & 0 \cdot 0 + 1 \cdot 2 \\ 10 \cdot 1 + 0 \cdot 0 & 10 \cdot 0 + 0 \cdot 0 & 10 \cdot 0 + 0 \cdot 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 \\ 0 & 0 & 2 \\ 10 & 0 & 0 \end{pmatrix}$$

Произведение матриц и линейные отображения

Ранее было сказано, что матрицы задают линейные отображения. Матрица $A \in \mathbb{R}^{m \times n}$ размера $m \times n$ будет задавать отображение из пространства векторов длины n в пространство векторов размера m. Аналогично матрица $B \in \mathbb{R}^{n \times k}$ задает отображение из пространства векторов длины k в пространство векторов длины n.

Но тогда можно построить отображение из пространства векторов длины n в пространство векторов длины k последовательным применением отображений, задаваемых матрицами A и B:

$$y_j = \sum_{j=1}^{n} a_{ij} \left(\sum_{p=1}^{k} b_{jp} w_p \right) = \sum_{p=1}^{k} \left(\sum_{j=1}^{n} a_{ij} b_{jp} \right) w_p = \sum_{p=1}^{k} c_{ip} w_p.$$

Матрица получившегося преобразования в точности равна произведению матриц A и B:

$$c_{ip} = \sum_{j=1}^{n} a_{ij} b_{jp}.$$

Определённое таким образом произведение матриц оказывается совершенно естественным.

Например, при решении системы линейных уравнений $Ax = b$ (например при решении задачи линейной классификации) на x могут дополнительно быть наложены определенные ограничения. Если такие ограничения могут быть введены путем замены $x = Bz$, то вектор z может быть найдет из системы уравнений:

$$ABz = b,$$

матрица которой в точности равна произведению матриц A и B.

Сложение и умножение на число

Для матриц также определены операции сложения и умножения на число. **Складывать можно только матрицы одинакового размера.** Сложение матриц происходит поэлементно, то есть $c_{ij} = a_{ij} + b_{ij}$, если $C = A + B$, где A и B — матрицы одного размера. Умножение матрицы на число заключается в умножении каждого элемента матрицы на это число, то есть $d_{ij} = \lambda a_{ij}$, если $D = \lambda A$.

Другими словами, матрицы одного размера образуют векторное пространство.

Транспонирование матриц

Для матриц также определяют операцию транспонирования. Матрица $B = A^T$, полученная транспонированием матрицы A размера $m \times n$, будет иметь размер $n \times m$, а её элементы будут связаны с элементами исходной матрицы следующим образом: $b_{ij} = a_{ji}$. Говоря нестрого, столбцы исходной матрицы стали строками новой матрицы, а строки — столбцами.

Пример: транспонирование матриц

$$A = \begin{pmatrix} 1 & 4 & 1 \\ 2 & 3 & 3 \\ 2 & 1 & 0 \\ 5 & 1 & 1 \end{pmatrix}, \qquad A^T = \begin{pmatrix} 1 & 2 & 2 & 5 \\ 4 & 3 & 1 & 1 \\ 1 & 3 & 0 & 1 \end{pmatrix}.$$

3. Ранг и определитель

Определитель матрицы 2×2: геометрический смысл

Важным понятием линейной алгебры является определитель (детерминант) матрицы. Он имеет множество приложений, в том числе и геометрических. Например, с помощью него можно вычислить площадь построенного на двух векторах параллелограмма на плоскости.

Для этого каждый из векторов следует записать как отдельный столбец матрицы A размера 2×2. Тогда преобразование, задаваемое матрицей A, будет отображать единичный квадрат в требуемый параллелограмм. Определитель $\det A$ с точностью до знака будет равен площади параллелограмма:

$$\det A = a_{11}a_{22} - a_{12}a_{21}, \qquad S = |\det A|.$$

т. Гаусса

Таким образом, абсолютное значение определителя отражает коэффициент изменения площади при преобразовании, а знак показывает, выполняет ли преобразование A отражение. Вообще говоря, свойства преобразования можно изучать на основе того, в какой параллелограмм оно переводит единичный квадрат.

Пример: нахождение площади параллелограмма

вектора II
линейно зависимы
Sпарал = 0

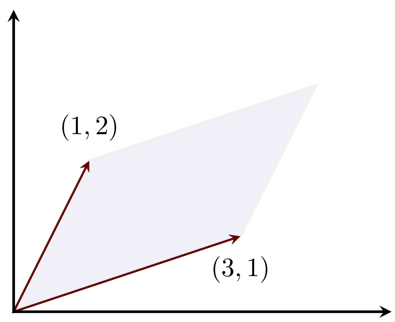

Рис. 1: Площадь параллелограмма.

Рис. 2: Вырожденный случай.

Для того, чтобы вычислить площадь параллелограмма, построенного на векторах $(1, 2)^T$ и $(3, 1)^T$, вычислим определитель матрицы, составленной из этих векторов:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, \qquad S = |\det A| = |3 \cdot 2 - 1 \cdot 1| = 5.$$

Определитель матрицы

Понятие определителя матрицы имеет смысл только для квадратных матриц. Дать определение определителю на случай пространств больших размерностей можно несколькими способами: существуют определение через перестановки, определение через формулу разложения по строке и аксиоматическое определение. Геометрический смысл определителя некоторой матрицы — ориентированный объем многомерного параллелепипеда, построенного на векторах, которые являются столбцами этой матрицы.

Можно дать определение используя следующую рекурсивную формулу (разложение по строке):

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \bar{M}_j^1,$$

где $\bar{M}_j^1$ — определитель матрицы, полученной из исходной вычеркиванием 1-строки и j-го столбца.

Свойства определителя

Определитель широко используется в линейной алгебре, так как обладает целым рядом важных свойств:

- Определитель матрицы, содержащей линейно зависимые строки или столбцы, равен нулю.
- Определитель не меняется при транспонировании.
- Если $A, B \in \mathbb{R}^{n \times n}$ — квадратные матрицы одного размера, то $\det(AB) = \det A \cdot \det B$.

Ранг матрицы

Ранг матрицы $\operatorname{rg} A$ — другое важное понятие из линейной алгебры. Сначала дают два определения для ранга матрицы:

- **Строчный ранг:** называется максимальное число линейно независимых строк.
- **Столбцовый ранг:** называется максимальное число линейно независимых столбцов.

Фундаментальная теорема «о ранге матрицы» говорит, что строчный ранг равен столбцовому, и поэтому говорят просто о ранге матрицы. Чем меньше ранг, тем меньше можно сжать матрицу без потерь.

4. Системы линейных уравнений (СЛАУ)

Систему из m линейных уравнений относительно n неизвестных x_1,..., x_n можно записать в общем виде:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

Но удобно использовать матричную форму СЛАУ. Система линейных алгебраических уравнений может быть представлена в матричной форме как:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

или более кратко $Ax = b$, где A — это матрица коэффициентов, b — столбец свободных членов, а x — вектор из неизвестных. Причем возможны три случая:

1. Система вовсе не имеет решения. Такая система называется несовместной.
2. Система имеет единственное решение.
3. Система имеет бесконечно много решений.
   Это связано с тем, что если система имеет два решения $\vec{x}_1$ и $\vec{x}_2$, то любой вектор вида $t\vec{x}_1 + (1-t)\vec{x}_2$, где $t \in \mathbb{R}$, также будет решением.

Теорема Кронекера-Капелли дает критерий совместности системы. Система линейных алгебраических уравнений $Ax = b$ совместна тогда и только тогда, когда ранг её основной матрицы A равен рангу её расширенной матрицы $(A|b)$, причём система имеет единственное решение, если ранг равен числу неизвестных, и бесконечное множество решений, если ранг меньше числа неизвестных. Расширенная матрица получается из матрицы A приписыванием к ней справа столбца b.

Решение системы уравнений с квадратной матрицей можно записать с привлечением так называемой обратной матрицы. Матрица называется обратной к матрице A и обозначается A^{-1}, если выполнено:

$$AA^{-1} = A^{-1}A = I,$$

где I — единичная матрица нужного размера. **Обращение возможно только в случае квадратных невырожденных матриц.** Действительно, если предположить, что у вырожденной матрицы A ($\det A = 0$) существовала бы обратная матрица, легко можно прийти к противоречию:

$$1 = \det I = \det(A^{-1}A) = \det A^{-1} \det A = \det A^{-1} \cdot 0 = 0.$$

Решение уравнения $Ax = b$, если матрица A — невырожденная, единственно и запишется в виде:

$$x = A^{-1}b.$$

Задача вычисления обратной матрицы и решения СЛАУ, строго говоря, сравнимы по сложности.

5. Типы матриц

Неоднократно было сказано, что матрица задаёт линейное отображение из одного векторного пространства в другое. Особый интерес представляет случай, когда строится отображение из некоторого линейного пространства в его само же. Такой класс линейных отображений называется линейными преобразованиями. Матрицы линейных преобразований всегда квадратные.

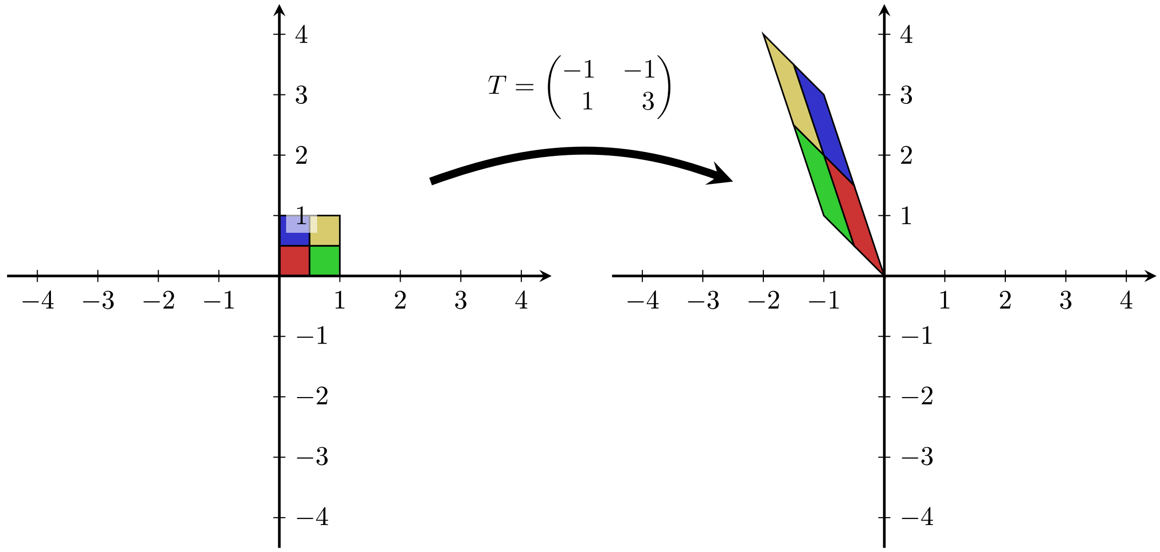

Рис. 3: Действие линейного преобразования, задаваемого матрицей T, на единичный квадрат.

Квадратная матрица называется диагональной в том случае, если все её элементы, стоящие вне главной диагонали, равны нулю. Частным случаем диагональной матрицы является единичная матрица I, на главной диагонали которой стоят единицы. Если матрица преобразования диагональная, то она задает растяжение или сжатие по осям системы координат.

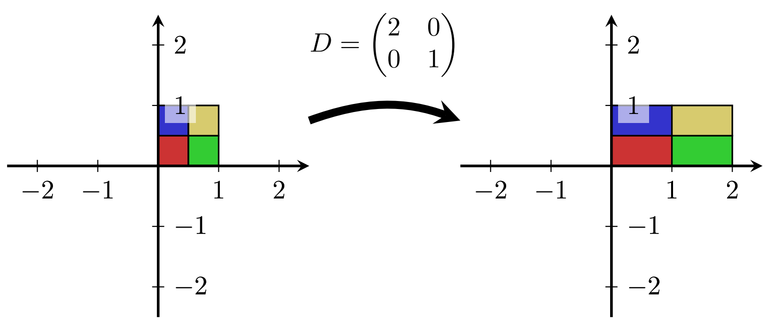

Рис. 4: Действие линейного преобразования, задаваемого диагональной матрицей D.

Важный класс квадратных матриц — ортогональные матрицы. Матрица называется ортогональной, если:

$$A^T A = AA^T = I.$$

Непосредственно из определения следуют следующие важные свойства:

- Ортогональная матрица обратима, причем $A^{-1} = A^T$.
- Ортогональные преобразования сохраняют скалярное произведение:

$$\langle Ax, Az\rangle = (Ax)^T(Az) = x^T A^T Az = x^T z = \langle x, z\rangle.$$

- Ортогональные преобразования сохраняют длины векторов $\|Ax\| = \|x\|$ и углы между векторами.

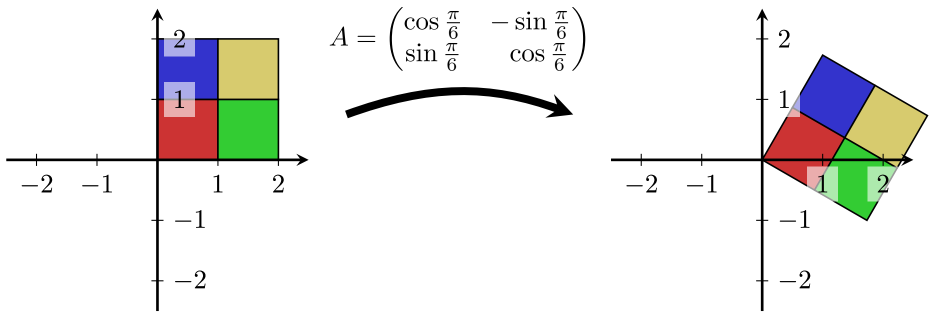

Рис. 5: Действие линейного преобразования, задаваемого ортогональной матрицей A.

В случае евклидовой плоскости всякое ортогональное преобразование является поворотом или поворотом с инверсией, и его матрица в любом ортонормированном базисе имеет вид

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

det A = +-1

Другой важнейший класс матриц — симметричные. Матрица A называется симметричной, если $A = A^T$.

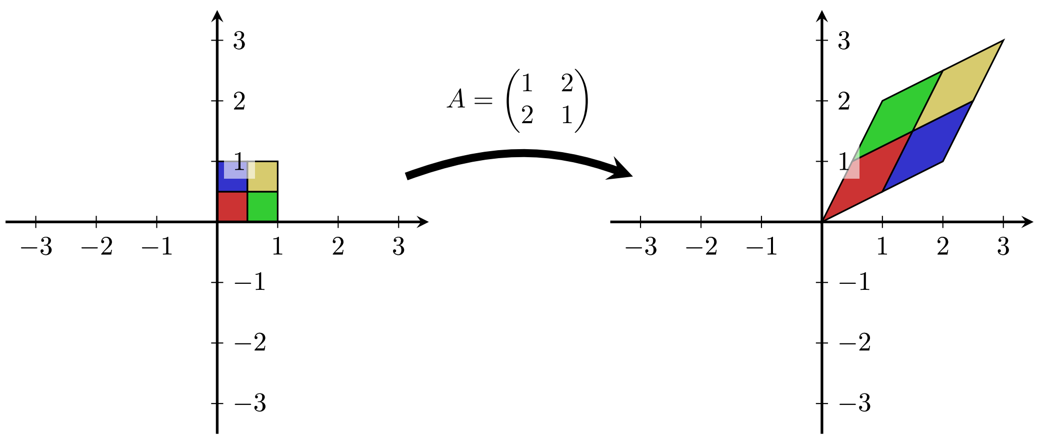

Рис. 6: Действие линейного преобразования, задаваемого симметричной матрицей A.

Важным результатом линейной алгебры является теорема о приведении симметричных матриц к диагональному виду с помощью ортогонального преобразования. А именно всякую симметричную матрицу $A = A^T$ можно записать как $A = QDQ^T$, где Q — некоторая ортогональная матрица, а D — диагональная матрица.

6. Понятие собственного вектора

Важной характеристикой матрицы, а также линейного преобразования, заданного этой матрицей, является спектр — набор собственных векторов и соответствующих собственных значений.

Собственным вектором линейного преобразования A называется такой ненулевой вектор $x \in V$, что для некоторого $\lambda \in \mathbb{R}$ выполняется $Ax = \lambda x$.

Линейное преобразование может как не иметь собственных векторов вообще, например поворот в двумерном пространстве (кроме нескольких исключительных случаев), или иметь n собственных векторов с различными собственными значениями. Вопросы существования собственных векторов преобразования разбираются в курсе линейной алгебры.

Понятие собственного вектора используется в Методе Главных Компонент, который предназначен для уменьшения размерности данных с потерей наименьшего количества информации.