Математика для Data Science. Линейная алгебра. Шпаргалка

Содержание

Первая неделя. Линейные отображения и матрицы

Векторное пространство

Вектор это упорядоченный набор действительных чисел. В машинном обучении объекты часто заменяют на векторы признаков объектов.

$\mathbb{R}^n$ – это множество всех упорядоченных наборов вида $(x_1, x_2, \ldots, x_n)$, таких что $\forall i : x_i \in \mathbb{R}$. Каждый такой набор называется *вектором*. Множество $\mathbb{R}^n$ называется *векторным пространством*.

Обозначение. Мы будем обозначать вектор буквой со стрелочкой, а все его координаты – обычными буквами с индексами. Вот так: $\vec{x} = (x_1, x_2, \ldots, x_n)$.

Иногда нам будет удобно представлять некоторые векторы не как строку $\vec{y} = (y_1, \ldots, y_m)$, а как столбец. Вот так:

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Сложение векторов из $\mathbf{R^n}$. Сложить можно любые два вектора из одного и того же векторного пространства $\mathbb{R}^n$. Результатом сложения также будет вектор из $\mathbb{R}^n$. Сумма векторов $(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ определяется так:

$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \in \mathbb{R}^n$.

Умножение вектора на число. На $\mathbb{R}^n$ определена операция умножения вектора на число. Любой вектор из $\mathbb{R}^n$ можно умножить на любое число, и результат умножения тоже будет вектором из $\mathbb{R}^n$. Число в этом случае называют *скаляром*. Умножение вектора $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ на число $c \in \mathbb{R}$ определяется так:

$c(x_1, x_2, \ldots, x_n) = (cx_1, cx_2, \ldots, cx_n) \in \mathbb{R}^n$.

Линейные отображения

Линейным отображением векторного пространства V в векторное пространство W называется функция $f : V \to W$ удовлетворяющая следующим двум условиям:

1. $f(\vec{x}) + f(\vec{y}) = f(\vec{x} + \vec{y})$ для любых $\vec{x}, \vec{y} \in V$,
2. $f(c\vec{x}) = cf(\vec{x})$ для любого $\vec{x} \in V$ и любого числа $c \in \mathbb{R}$.

Общий вид некоторых линейных отображений

1. Любое линейное отображение $f : \mathbb{R} \to \mathbb{R}$ задаётся одним числом $f(1)$: а именно, $f(\vec{x}) = x \cdot f(1)$.

2. Любое линейное отображение $f : \mathbb{R} \to \mathbb{R}^m$ задаётся столбцом высоты m, состоящим из чисел. Этот столбец равен $f(1)$. При этом $f(\vec{x}) = x \cdot f(1)$.

3. Любое линейное отображение $f : \mathbb{R}^n \to \mathbb{R}$ задаётся строкой длины n, состоящей из чисел.

Обозначение.

- $\vec{e}_1 := (1, 0, 0, \ldots, 0)$
- $\vec{e}_2 := (0, 1, 0, \ldots, 0)$
- $\vdots$
- $\vec{e}_n := (0, 0, 0, \ldots, 1)$

Другими словами, у вектора $\vec{e}_i$ равна единице i-ая координата, а все остальные координаты равны нулю.
Рассмотрим любой вектор $\vec{x} = (x_1, \ldots, x_n)$. По определению векторного пространства $\mathbb{R}^n$, выполнено:
$(x_1, \ldots, x_n) = (x_1, 0, \ldots, 0) + (0, x_2, \ldots, 0) + \cdots + (0, 0, \ldots, x_n) = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n$.
Обозначение. В таких случаях говорят, что вектор $\vec{x}$ *выражается* через векторы $\vec{e}_1, \ldots, \vec{e}_n$.

Пусть дано линейное отображение $f : \mathbb{R}^n \to \mathbb{R}$ и для каждого i $f(\vec{e}_i) = a_i$, где $a_i \in \mathbb{R}$. Тогда f определяется набором чисел $a_1, \ldots, a_n$: $f(\vec{x}) = f(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) = f(x_1\vec{e}_1) + \cdots + f(x_n\vec{e}_n) = x_1 f(\vec{e}_1) + \cdots + x_n f(\vec{e}_n) = x_1 a_1 + \cdots + x_n a_n = a_1 x_1 + \cdots + a_n x_n$.

Вычисление значения функции f мы будем записывать как *действие* строки $(a_1, \ldots, a_n)$ на столбец $\vec{x}$:

$$f(\vec{x}) = (a_1, \ldots, a_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := a_1 x_1 + \cdots + a_n x_n = \sum_{i=1}^{n} a_i x_i.$$

Можно говорить, что мы *умножаем* строку $(a_1, \ldots, a_n)$ на столбец $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Матрицы

Любое линейное отображение $f : \mathbb{R}^n \to \mathbb{R}^m$ задаётся матрицей с m строками и n столбцами.

$$A := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

где $a_{ij} \in \mathbb{R}$ для всех i, j. Числа a_{ij} называют *элементами* матрицы.

Построим по матрице A линейное отображение:

i-ой строке матрицы $(a_{i1}, a_{i2}, \ldots, a_{in})$ соответствует линейное отображение $f_i : \mathbb{R}^n \to \mathbb{R}$:

$$f_i(\vec{x}) = (a_{i1}, a_{i2} \ldots, a_{in})\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j \in \mathbb{R}.$$

Пользуясь отображениями $f_1, \ldots, f_m$, мы так определяем действие отображения f на векторе $\vec{x}$:

$$f(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix} \in \mathbb{R}^m.$$

Обозначение. Полученный вектор $f(\vec{x})$ называется результатом *действия* матрицы A на векторе $\vec{x}$, и обозначается $A\vec{x}$.

Записывают действие матрицы A на векторе $\vec{x}$ так:

$$A\vec{x} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix} \in \mathbb{R}^m$$

Также можно говорить, что мы *умножили* A на $\vec{x}$ и получили $A\vec{x}$.

Как воспринимать столбцы матрицы: для любого j вектор $f(\vec{e}_j) = A\vec{e}_j$ совпадает с j-ым столбцом матрицы A. Таким образом, столбцы матрицы — это те векторы, в которые f переводит единичные векторы, направленные вдоль осей координат. $A = \begin{pmatrix} f(\vec{e}_1) & f(\vec{e}_2) & \ldots & f(\vec{e}_n) \end{pmatrix}$

Итак, любая матрица задаёт линейное отображение. При этом любое линейное отображение задаётся какой-то матрицей. Кроме того, разные матрицы задают разные линейные отображения. Значит, есть *взаимно однозначное соответствие* между матрицами и линейными отображениями. То есть каждой матрице соответствует ровно одно линейное отображение, и каждому линейному отображению поставлена в соответствие ровно одна матрица.

Умножение матриц

Пусть определены два линейных отображения: $f : \mathbb{R}^n \to \mathbb{R}^m$ и $g : \mathbb{R}^m \to \mathbb{R}^k$.

Их *композицией* называется отображение из $\mathbb{R}^n$ в $\mathbb{R}^k$ и обозначается $g \circ f$. Действует $g \circ f$ на любом векторе $\vec{x} \in \mathbb{R}^n$ так:

1. применяя f, получаем вектор $f(\vec{x}) \in \mathbb{R}^m$
2. применяя к полученному вектору g, получаем вектор $g(f(\vec{x})) \in \mathbb{R}^k$

Вектор $g(f(\vec{x})) \in \mathbb{R}^k$ и называется результатом действия $g \circ f$ на $\vec{x}$.

Композиция двух линейных преобразований тоже будет линейным преобразованием.

Если отображение $f : \mathbb{R}^n \to \mathbb{R}^m$ задано матрицей A, а отображение $g : \mathbb{R}^m \to \mathbb{R}^k$ — матрицей B, то $(g \circ f)(\vec{x}) = g(f(\vec{x})) = g(A(\vec{x})) = B(A(\vec{x}))$. Матрицу преобразования $g \circ f$ обозначают BA и называют *произведением* матрицы B и матрицы A.

Правило умножения матриц

На пересечении i-ой строки и j-ого столбца произведения двух матриц стоит произведение i-ой строки первой матрицы и j-ого столбца второй матрицы.

При этом произведение BA определено, только если число столбцов матрицы B совпадает с числом строк матрицы A.

Умножение называется *коммутативным*, если $B \cdot A = A \cdot B$ для любых A и B. Умножение матриц <u>не является</u> коммутативным: то есть, бывают матрицы A и B размера n на n, что $AB \neq BA$.

Примеры некоторых типов матриц

1. Матрица, на *главной диагонали* которой стоят единицы, а на всех остальных местах стоят нули, называется *единичной матрицей*. Чаще всего её обозначают буквами E или I. Если важно подчеркнуть, что матрица именно размера n на n, то добавляют нижний индекс: E_n, I_n.

$$E_n = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

2. Иногда матрицы удобнее записывать в виде нескольких блоков, каждый из которых соответствует меньшей матрице. Например, пусть

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{12} \end{pmatrix}, Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{12} \end{pmatrix}.$$

Тогда мы можем так определить *блочную* матрицу $\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}$:

$$\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} := \begin{pmatrix} x_{11} & x_{12} & 0 & 0 \\ x_{21} & x_{12} & 0 & 0 \\ 0 & 0 & y_{11} & y_{12} \\ 0 & 0 & y_{21} & y_{12} \end{pmatrix}.$$

При этом два нуля в этой блочной матрице обозначают заполненные нулями матрицы размера 2 на 2.