Учебник по Эконометрике

Лекция 2: Линейная Алгебра и Матрицы

Джон Стачурски

Лекции: Акшай Шенкер

Перевел: Алексей Кедо

28 сентября 2020 г.

Матрицы

Обычная $N \times K$ **матрица**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix}$$

Символ a_{nk} означает элемент, стоящий в n-ой строке k-ого столбца

$N \times K$ матрица также называется

- **вектором строки**, если $N = 1$
- **вектором столбца**, если $K = 1$

Если $N = K$, то матрицу $\mathbf{A}$ называют **квадратной**

Если $\mathbf{A}$ квадратная и $a_{nk} = a_{kn}$ для любых k и n, то $\mathbf{A}$ называют **симметричной**

Часто элементы матрицы $\mathbf{A}$ представляют собой коэффициенты в системе линейных уравнений

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1K}x_K &= b_1 \\
&\vdots \\
a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NK}x_K &= b_N
\end{aligned}
$$

Для матрицы $\mathbf{A}$ применяются обозначения:

- $\mathrm{row}_n\,\mathbf{A}$ означает n-ую строку $\mathbf{A}$
- $\mathrm{col}_k\,\mathbf{A}$ означает k-ый столбец $\mathbf{A}$

Символы $\mathbf{0}$ и $\mathbf{1}$ представляют собой матрицы, все элементы которых равны нулю и единице соответственно

Для квадратной матрицы $\mathbf{A}$, элементы a_{nn} называют **главной диагональю**:

$$\begin{pmatrix} \color{red}{a_{11}} & a_{12} & \cdots & a_{1N} \\ a_{21} & \color{red}{a_{22}} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & \color{red}{a_{NN}} \end{pmatrix}$$

Единичная матрица:

$$\mathbf{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Заметим, что $\operatorname{col}_n \mathbf{I} = \mathbf{e}_n$ — n-ый канонический базис $\mathbb{R}^N$

Алгебраические операции над матрицами

Операции над матрицами:

- Умножение на скаляр
- Сложение
- Умножение матриц

Умножение на скаляр выполняется поэлементно, как с векторами:

$$\gamma \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} := \begin{pmatrix} \gamma a_{11} & \gamma a_{12} & \cdots & \gamma a_{1K} \\ \gamma a_{21} & \gamma a_{22} & \cdots & \gamma a_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma a_{N1} & \gamma a_{N2} & \cdots & \gamma a_{NK} \end{pmatrix}$$

Сложение тоже выполняется поэлементно:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1K} \\ b_{21} & \cdots & b_{2K} \\ \vdots & \vdots & \vdots \\ b_{N1} & \cdots & b_{NK} \end{pmatrix}$$

$$:= \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1K}+b_{1K} \\ a_{21}+b_{21} & \cdots & a_{2K}+b_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1}+b_{N1} & \cdots & a_{NK}+b_{NK} \end{pmatrix}$$

Заметим, что матрицы должны быть одинакового размера

Умножение матриц:

Произведение $\mathbf{AB}$: i, j-ый элемент — скалярное умножение i-ой строки $\mathbf{A}$ и j-ого столбца $\mathbf{B}$

$$c_{ij} = \langle \mathrm{row}_i\, \mathbf{A}, \mathrm{col}_j\, \mathbf{B} \rangle = \sum_{k=1}^{K} a_{ik} b_{kj}$$

Для $i = j = 1$:

$$
\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix}
\begin{pmatrix} b_{11} & \cdots & b_{1J} \\ b_{21} & \cdots & b_{2J} \\ \vdots & \vdots & \vdots \\ b_{K1} & \cdots & b_{KJ} \end{pmatrix}
=
\begin{pmatrix} c_{11} & \cdots & c_{1J} \\ c_{21} & \cdots & c_{2J} \\ \vdots & \vdots & \vdots \\ c_{N1} & \cdots & c_{NJ} \end{pmatrix}
$$

Здесь

$$
c_{11} = \langle \mathrm{row}_1(\mathbf{A}), \mathrm{col}_1(\mathbf{B}) \rangle = \sum_{k=1}^{K} a_{1k} b_{k1}
$$

Предположим, что размер $\mathbf{A}$ — $N \times K$, $\mathbf{B}$ — $J \times M$

- $\mathbf{AB}$ определена, только если $K = J$
- Размер $\mathbf{AB}$ — $N \times M$

Запомните правило:

Произведение $N \times K$ и $K \times M$ равно $N \times M$

Умножение не коммутативно: $\mathbf{AB} \neq \mathbf{BA}$

Заметим, что произведение $\mathbf{BA}$ определено, только если $N = M$ соблюдается

Факт. (??) Для согласованных матриц $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ и скаляра α, верно

1. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$,
2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$,
3. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$,
4. $\mathbf{A}\alpha\mathbf{B} = \alpha\mathbf{AB}$, и
5. $\mathbf{AI} = \mathbf{A}$ и $\mathbf{IA} = \mathbf{A}$, где $\mathbf{I}$ — единичная матрица.

Здесь “согласованость” значит, что операция определена при заданных размерностях матриц

k-**ая мощность** квадратной матрицы $\mathbf{A}$ определяется как

$$\mathbf{A}^k := \underbrace{\mathbf{A}\cdots\mathbf{A}}_{k \text{ terms}}$$

Если матрица $\mathbf{B}$ такая, что $\mathbf{B}^2 = \mathbf{A}$, то $\mathbf{B}$ называется **квадратным корнем** матрицы $\mathbf{A}$ и записывается как $\sqrt{\mathbf{A}}$

Возьмем матрицу $\mathbf{A}$ размерности $N \times K$ и $K \times 1$ вектор столбца $\mathbf{x}$, перемножение $\mathbf{Ax}$:

$$\mathbf{Ax} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}$$

$$= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{N2} \end{pmatrix} + \cdots + x_K \begin{pmatrix} a_{1K} \\ a_{2K} \\ \vdots \\ a_{NK} \end{pmatrix}$$

$$= \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Матрицы как отображения

Можно размышлять о матрице $\mathbf{A}$ размерна $N \times K$ как об отображении из $\mathbb{R}^K$ в $\mathbb{R}^N$:

$$\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$$

Такое отображение линейно

Как насчет примеров линейных функций, не использующих матрицы?

...на самом деле, таких не существует!

Множеству линейных функций из $\mathbb{R}^K$ в $\mathbb{R}^N$ инъективно соответствует множество матриц $N \times K$:

Теорема. (**??**) Пусть T является функцией из $\mathbb{R}^K$ в $\mathbb{R}^N$. Следующее эквивалентно:

1. T линейна.
2. Существует матрица $\mathbf{A}$ размера $N \times K$, такая что $T\mathbf{x} = \mathbf{A}\mathbf{x}$ для всех $\mathbf{x} \in \mathbb{R}^K$.

Доказательство. (1. $\implies$ 2.)

Пусть $T\colon \mathbb{R}^K \to \mathbb{R}^N$ линейна

Мы собираемся построить матрицу $\mathbf{A}$ размерности $N \times K$, такую что

$$T\mathbf{x} = \mathbf{A}\mathbf{x}, \qquad \forall\, \mathbf{x} \in \mathbb{R}^K$$

Как обычно, пусть $\mathbf{e}_k$ — k-ый канонический базисный вектор в $\mathbb{R}^K$

Определим матрицу $\mathbf{A}$ как $\operatorname{col}_k(\mathbf{A}) = T\mathbf{e}_k$. Возьмем любой $\mathbf{x} = (x_1, \ldots, x_K) \in \mathbb{R}^K$. По линейности, получается

$$T\mathbf{x} = T\left[\sum_{k=1}^{K} x_k \mathbf{e}_k\right] = \sum_{k=1}^{K} x_k T\mathbf{e}_k = \sum_{k=1}^{K} x_k \operatorname{col}_k(\mathbf{A}) = \mathbf{A}\mathbf{x}$$

Доказательство. (2 $\implies$ 1) Возьмем матрицу $\mathbf{A}$ размера $N \times K$ и пусть T определяется

$$T \colon \mathbb{R}^K \to \mathbb{R}^N, \qquad T\mathbf{x} = \mathbf{A}\mathbf{x}$$

Возьмем любые $\mathbf{x}$, $\mathbf{y}$ в $\mathbb{R}^K$, и любые скаляры α и β

Правила матричной арифметики говорят нам, что

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) := \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} =: \alpha T\mathbf{x} + \beta T\mathbf{y}$$

Рассмотрим возможность решения системы линейных уравнений, например, $\mathbf{Ax} = \mathbf{b}$

Существование: можем ли мы найти $\mathbf{x}$, удовлетворяющий уравнению при любых заданных $\mathbf{b}$?

- является ли линейное отображение $T\mathbf{x} = \mathbf{Ax}$ сюръекцией?
- эквивалентно, равен ли $\operatorname{rng} T$ всему $\mathbb{R}^N$?

Пространство столбцов

Диапазон T — все вектора вида $T\mathbf{x} = \mathbf{A}\mathbf{x}$, где $\mathbf{x}$ варьируется в $\mathbb{R}^K$

Для $\mathbf{x} \in \mathbb{R}^K$ получается

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Таким образом, $\operatorname{rng} T$ равен **пространству столбцов** $\mathbf{A}$ — линейная оболочка столбцов $\mathbf{A}$

$$\operatorname{colspace} \mathbf{A} := \operatorname{span}\{\operatorname{col}_1 \mathbf{A}, \ldots, \operatorname{col}_K \mathbf{A}\}$$

В итоге,

$$\operatorname{colspace} \mathbf{A} = \operatorname{rng} T = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^K\}$$

Ранг

Эквивалентные вопросы

- Насколько велик диапазон линейного отображения $T\mathbf{x} = \mathbf{A}\mathbf{x}$?
- Насколько велико пространство столбцов $\mathbf{A}$?

Очевидной мерой размера линейного подпространства является его размерность

Размерность colspace $\mathbf{A}$ известна как **ранг** $\mathbf{A}$

$$\operatorname{rank} \mathbf{A} := \dim \operatorname{colspace} \mathbf{A}$$

Так как colspace $\mathbf{A}$ — линейная оболочка K векторов, получается

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{colspace} \mathbf{A} \leq K$$

$\mathbf{A}$ имеет **полный ранг системы столбцов**, если

$$\operatorname{rank} \mathbf{A} = \text{ количество столбцов } \mathbf{A}$$

Факт. (??) Для любой матрицы $\mathbf{A}$, следующие утверждения эквивалентны:

1. $\mathbf{A}$ является полным рангом системы столбцов
2. Столбцы $\mathbf{A}$ линейно независимы
3. Если $\mathbf{Ax} = \mathbf{0}$, то $\mathbf{x} = \mathbf{0}$

Квадратные матрицы и обратимость

Рассмотрим случай с матрицей $\mathbf{A}$ размера $N \times N$

Мы ищем условия, при которых для каждого $\mathbf{b} \in \mathbb{R}^N$, существует только один $\mathbf{x} \in \mathbb{R}^N$, такой что $\mathbf{Ax} = \mathbf{b}$

Пусть T является линейным отображением $T\mathbf{x} = \mathbf{Ax}$

- Когда каждая точка $\mathbf{b} \in \mathbb{R}^N$ имеет только один прообраз в T?
- Эквивалентно, когда T — биекция?

Напомним, линейные биекции называются несингулярными функциями.

Факт. (??) Для матрицы $\mathbf{A}$ размера $N \times N$ следующее эквивалентно:

1. Столбцы $\mathbf{A}$ линейно независимы.
2. Столбцы $\mathbf{A}$ формируют базис $\mathbb{R}^N$.
3. $\operatorname{rank} \mathbf{A} = N$.
4. $\operatorname{colspace} \mathbf{A} = \mathbb{R}^N$.
5. $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{y} \implies \mathbf{x} = \mathbf{y}$.
6. $\mathbf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$.
7. Для каждого $\mathbf{b} \in \mathbb{R}^N$, уравнение $\mathbf{A}\mathbf{x} = \mathbf{b}$ имеется решение.
8. Для каждого $\mathbf{b} \in \mathbb{R}^N$, уравнение $\mathbf{A}\mathbf{x} = \mathbf{b}$ имеет единственное решение.

Если любое из эквивалентных условий факта **??** верное, мы назовем **несингулярной** не только отображение T, но и основную матрицу $\mathbf{A}$

Если хоть одно — и, следовательно, все — из этих условий не выполняются, то $\mathbf{A}$ называется **сингулярной**

Любая биекция имеет обратную функцию (смотрите §?? в ET)

Любое несингулярное отображение T имеет несингулярное обратное T^{-1} (факт **??** на странице **??**)

- если T создано матрицей $\mathbf{A}$, обратное T^{-1} также связано с матрицей, называемой обратной $\mathbf{A}$

Теорема. (??) Для несингулярной $\mathbf{A}$ следующие утверждения верны:

1. Существует квадратная матрица $\mathbf{B}$, такая что $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, где $\mathbf{I}$ — единичная матрица. Матрица $\mathbf{B}$ называется **обратной** $\mathbf{A}$, и записывается как $\mathbf{A}^{-1}$.
2. Для каждого $\mathbf{b} \in \mathbb{R}^N$, единственное решение задачи $\mathbf{Ax} = \mathbf{b}$ записывается как

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Пример. Расстмотрим линейный спрос для N товаров

$$q_n = \sum_{k=1}^{N} a_{nk} p_k + b_n, \quad n = 1, \dots, N$$

где q_n и p_n — количество и цена n-ого товара

Мы хотим вычислить обратную функцию спроса, которая показывает зависимость цены от количества

Записшем систему в матричной форме: $\mathbf{q} = \mathbf{A}\mathbf{p} + \mathbf{b}$

Если столбцы $\mathbf{A}$ линейно независимы, то система обратима — единственное решение существует для любых фиксированных $\mathbf{q}$ и $\mathbf{b}$:

$$\mathbf{p} = \mathbf{A}^{-1}(\mathbf{q} - \mathbf{b})$$

Для матриц $\mathbf{A}$ и $\mathbf{B}$ размера $N \times N$, если

- $\mathbf{B}$ является **левой обратной**, в том смысле, что $\mathbf{BA} = \mathbf{I}$
- или $\mathbf{B}$ является **правой обратной**, в том смысле, что $\mathbf{AB} = \mathbf{I}$

Тогда $\mathbf{A}$ обратима и $\mathbf{B}$ является обратной матрицей $\mathbf{A}$

Факт. (??) Пусть $\mathbf{A}$ и $\mathbf{B}$ — квадратные матрицы размера $N \times N$. Если $\mathbf{B}$ является левой или правой обратной матрицы $\mathbf{A}$, то $\mathbf{A}$ несингулярна и $\mathbf{B}$ — ее обратная матрица.

Факт. (??) Если $\mathbf{A}$ и $\mathbf{B}$ несингулярны и $\alpha \neq 0$, то

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
2. $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$, и
3. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Определитель

Определитель — уникальное число для любой квадратной матрицы $\mathbf{A}$

Пусть $S(N)$ является множеством всех биекций из $\{1, \ldots, N\}$ в саму себя

Для $\pi \in S(N)$ определим **функцию знака** π

$$\operatorname{sgn}(\pi) := \prod_{m<n} \frac{\pi(m) - \pi(n)}{m - n}$$

Определитель матрицы $\mathbf{A}$ размера $N \times N$ — это

$$\det \mathbf{A} := \sum_{\pi \in S(N)} \operatorname{sgn}(\pi) \prod_{n=1}^{N} a_{\pi(n)n}$$

Факт. (??) Если $\mathbf{I}$ — единичная матрица размера $N \times N$, $\mathbf{A}$ и $\mathbf{B}$ являются марицами размера $N \times N$ и $\alpha \in \mathbb{R}$, то

1. $\det \mathbf{I} = 1$,
2. $\mathbf{A}$ несингулярна тогда и только тогда, когда $\det \mathbf{A} \neq 0$,
3. $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$,
4. $\det(\alpha \mathbf{A}) = \alpha^N \det(\mathbf{A})$, и
5. $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$.

В случае матрицы размера 2×2 можно показать, что

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Доказательство. (Факт ??)

Возьмем кважратную матрицу $\mathbf{A}$ и предположим, что правая обратная матрица $\mathbf{B}$ существует:

$$\mathbf{AB} = \mathbf{I}$$

Тогда обе матрицы $\mathbf{A}$ и $\mathbf{B}$ несингулярны, так как по правилам факта ??:

$$\det(\mathbf{A})\det(\mathbf{B}) = 1$$

Оба определителя $\det \mathbf{A}$ и $\det \mathbf{B}$ ненулевые и, следовательно, обе матрицы несингулярные. Далее, $\mathbf{AB} = \mathbf{I}$, значит

$$\mathbf{BAB} = \mathbf{B}$$

Умножим на $\mathbf{B}^{-1}$, получается $\mathbf{BA} = \mathbf{I}$

Диагональные и треугольные матрицы

Квадратная матрица назвается **нижней треугольной**, если каждый элемент строго выше главной диагонали равен нулю

Пример.

$$\mathbf{L} := \begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix}$$

Квадратная матрица назвается **верхней треугольной**, если каждый элемент строго ниже главной диагонали равен нулю

Пример.

$$\mathbf{U} := \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

Квадратная матрица назвается **треугольной**, если она верхняя или нижняя треугольная

Факт. (??) Если $\mathbf{A} = (a_{mn})$ треугольная, то $\det \mathbf{A} = \prod_{n=1}^{N} a_{nn}$.

Связанные линейные уравнения просто решить

Пример.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

становится

$$\begin{gathered} x_1 = b_1 \\ 2x_1 + 5x_2 = b_2 \\ 3x_1 + 6x_2 + x_3 = b_3 \end{gathered}$$

Верхнее уравнение включает только x_1, так что его можно решить напрямую

Подставьте это значение во второе уравнение и решите для x_2 и т.д.

Рассмотрим квадратную матрицу $\mathbf{A}$ размера $N \times N$

N элементов a_{nn} называются **главной диагональю**

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

Квадратная матрица $\mathbf{D}$ называется **диагональной**, если все значения вне главной диагонали равны нулю

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}$$

Часто записывается как

$$\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_N)$$

Диагональные системы очень легко решить

Пример.

$$\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

эквивалентно

$$\begin{aligned} d_1 x_1 &= b_1 \\ d_2 x_2 &= b_2 \\ d_3 x_3 &= b_3 \end{aligned}$$

Факт. (??) Если $\mathbf{C} = \operatorname{diag}(c_1, \ldots, c_N)$ и $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_N)$, то

1. $\mathbf{C} + \mathbf{D} = \operatorname{diag}(c_1 + d_1, \ldots, c_N + d_N)$
2. $\mathbf{CD} = \operatorname{diag}(c_1 d_1, \ldots, c_N d_N)$
3. $\mathbf{D}^k = \operatorname{diag}(d_1^k, \ldots, d_N^k)$ для любых $k \in \mathbb{N}$
4. $d_n \geq 0$ для всех $n \implies \mathbf{D}^{1/2}$ существует и равняется
$$\operatorname{diag}(\sqrt{d_1}, \ldots, \sqrt{d_N})$$
5. $d_n \neq 0$ для всех $n \implies \mathbf{D}$ несингулярна и
$$\mathbf{D}^{-1} = \operatorname{diag}(d_1^{-1}, \ldots, d_N^{-1})$$

Доказательства: Проверьте 1 и 2 напрямую. Части 3-5 следуют из 1 и 2.

След, транспонирование, симметрия

След матрицы определяется как

$$\operatorname{trace}\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} = \sum_{n=1}^{N} a_{nn}$$

Факт. (??) Если $\mathbf{A}$ и $\mathbf{B}$ являются квадратными матрицами и $\alpha, \beta \in \mathbb{R}$, то

$$\operatorname{trace}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \operatorname{trace}(\mathbf{A}) + \beta \operatorname{trace}(\mathbf{B})$$

Если $\mathbf{A}$ размера $N \times M$ и $\mathbf{B}$ размера $M \times N$, то $\operatorname{trace}(\mathbf{AB}) = \operatorname{trace}(\mathbf{BA})$

Транспонирование превращает матрицу $\mathbf{A}$ размера $N \times K$ в матрицу $\mathbf{A}^{\mathsf{T}}$ размера $K \times N$, определяемую как

$$\mathrm{col}_n(\mathbf{A}') = \mathrm{row}_n(\mathbf{A})$$

Пример. Если

$$\mathbf{A} := \begin{pmatrix} 10 & 40 \\ 20 & 50 \\ 30 & 60 \end{pmatrix} \quad \text{, то} \quad \mathbf{A}' = \begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix}$$

Если

$$\mathbf{B} := \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \quad \text{, то} \quad \mathbf{B}' := \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Квадратная матрица $\mathbf{A}$ называется **симметричной**, если $\mathbf{A}^\top = \mathbf{A}$

- эквивалентно, $a_{nk} = a_{kn}$ для всех k и n

Матрицы $\mathbf{A}^\top \mathbf{A}$ и $\mathbf{A}\mathbf{A}^\top$ всегда корректно определены и симметричны

Факт. (??) Для согласующихся матриц $\mathbf{A}$ и $\mathbf{B}$ транспонирование удовлетворяет

1. $(\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A}$,
2. $(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$,
3. $(\mathbf{A} + \mathbf{B})^\mathsf{T} = \mathbf{A}^\mathsf{T} + \mathbf{B}^\mathsf{T}$, и
4. $(c\mathbf{A})^\mathsf{T} = c\mathbf{A}^\mathsf{T}$ для любых постоянных c.

Факт. (??) Для любой квадратной матрицы $\mathbf{A}$, верно

1. $\mathrm{trace}(\mathbf{A}) = \mathrm{trace}(\mathbf{A}^\intercal)$ и
2. $\det(\mathbf{A}^\intercal) = \det(\mathbf{A})$.
3. Если $\mathbf{A}$ несингулярна, то $\mathbf{A}^\intercal$ тоже несингулярна, и $(\mathbf{A}^\intercal)^{-1} = (\mathbf{A}^{-1})^\intercal$.

Если $\mathbf{a}$ и $\mathbf{b}$ — векторы размера $N \times 1$, умножение матриц $\mathbf{a}^\intercal\mathbf{b} = \mathbf{b}^\intercal\mathbf{a}$ равняется $\sum_{n=1}^{N} a_n b_n$,

- так же, как скалярное произведение $\langle \mathbf{a}, \mathbf{b} \rangle$

Собственные значения и собственные векторы

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Обычно $\mathbf{A}$ отображает $\mathbf{x}$ в какое-то произвольное новое место $\mathbf{Ax}$

Но иногда $\mathbf{x}$ будет только масштабироваться:

$$\mathbf{Ax} = \lambda \mathbf{x} \quad \text{для некоторого скаляра } \lambda \tag{1}$$

Если (1) выполняется и $\mathbf{x}$ is ненулевой, то

1. $\mathbf{x}$ называется **собственным вектором** $\mathbf{A}$ и λ называется **собственным значением**
2. $(\mathbf{x}, \lambda)$ называется **собственной парой**

Ясно, что $(\mathbf{x}, \lambda)$ — собственная пара $\mathbf{A} \implies (\alpha \mathbf{x}, \lambda)$ — собственная пара $\mathbf{A}$ для любых ненулевых α

Пример. Пусть

$$\mathbf{A} := \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}$$

Тогда

$$\lambda = 2 \quad \text{и} \quad \mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

формируют собственную пару, потому что $\mathbf{x} \neq \mathbf{0}$ и

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lambda \mathbf{x}$$

Пример. Рассмотрим матрицу

$$\mathbf{R} := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Вызывает вращение против часовой стрелки в любой точке на 90°

Значит ни одна точка $\mathbf{x}$ не масштабируется

Значит не существует пары $\lambda \in \mathbb{R}$ и $\mathbf{x} \neq \mathbf{0}$, такой что

$$\mathbf{R}\mathbf{x} = \lambda \mathbf{x}$$

- Другими словами, не существует вещественных собственных пар

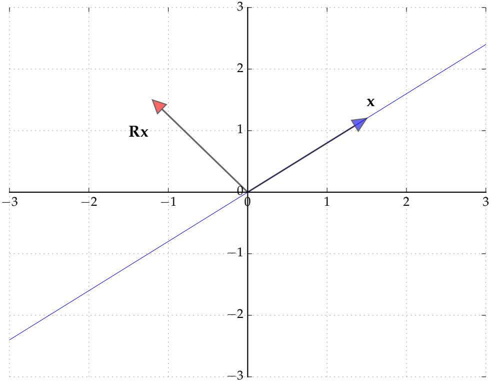

Рис.: Матрица $\mathbf{R}$ поворачивает точки на 90°

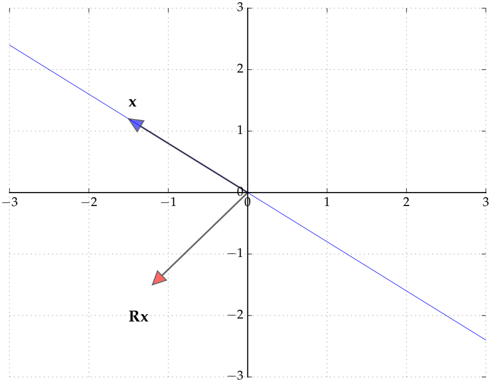

Рис.: Матрица $\mathbf{R}$ поворачивает точки на 90°

Но $\mathbf{Rx} = \lambda\mathbf{x}$ может выполняться, если мы допускаем комплексные числа

Пример.

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} i \\ 1 \end{pmatrix} = i \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

То есть

$$\mathbf{Rx} = \lambda\mathbf{x} \quad \text{для} \quad \lambda := i \quad \text{и} \quad \mathbf{x} := \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Тогда $(\mathbf{x}, \lambda)$ является собственной парой при условии, что мы допускаем комплексные числа

Факт. (??) для любой квадратной матрицы $\mathbf{A}$

$$\lambda \text{ является собственным значением } \mathbf{A} \iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

Доказательство. Пусть $\mathbf{A}$ — матрица размера $N \times N$ и $\mathbf{I}$ — единичная матрица размера $N \times N$

Получается

$$\begin{aligned}
\det(\mathbf{A} - \lambda\mathbf{I}) = 0 &\iff \mathbf{A} - \lambda\mathbf{I} \text{ сингулярно} \\
&\iff \exists \mathbf{x} \neq \mathbf{0} \text{ s.t. } (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \\
&\iff \exists \mathbf{x} \neq \mathbf{0} \text{ s.t. } \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \\
&\iff \lambda \text{ является собственным значением } \mathbf{A}
\end{aligned}$$

Пример. В случае матрицы 2×2,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad \mathbf{A} - \lambda \mathbf{I} = \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$$

$$\begin{aligned} \therefore \quad \det(\mathbf{A} - \lambda \mathbf{I}) &= (a - \lambda)(d - \lambda) - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) \end{aligned}$$

Значит собственные значения $\mathbf{A}$ являются двумя корнями уравнения

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0$$

Эквивалентно,

$$\lambda^2 - \mathrm{trace}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0$$

Существование собственных значений

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Факт. Существуют комплексные числа $\lambda_1, \ldots, \lambda_N$, такие что

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda)$$

Каждое такое λ_i является собственным значением $\mathbf{A}$, потому что

$$\det(\mathbf{A} - \lambda_i \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda_i) = 0$$

Важно: не все собственные значения обязательно различны — могут быть повторы

Факт. (??) Возьмем матрицу $\mathbf{A}$ размера $N \times N$ с собственными значениями $\lambda_1, \ldots, \lambda_N$. Получается

1. $\det(\mathbf{A}) = \prod_{n=1}^{N} \lambda_n$
2. $\operatorname{trace}(\mathbf{A}) = \sum_{n=1}^{N} \lambda_n$
3. если $\mathbf{A}$ симметрична, то $\lambda_n \in \mathbb{R}$ для всех n
4. если $\mathbf{A}$ несингулярна, то
   eigenvalues of $\mathbf{A}^{-1} = 1/\lambda_1, \ldots, 1/\lambda_N$
5. если $\mathbf{A} = \operatorname{diag}(d_1, \ldots, d_N)$, то $\lambda_n = d_n$ для всех n

Значит матрица $\mathbf{A}$ несингулярна $\iff$ все ее собственные значения ненулевые

Квадратичные формы

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Квадратичная функция или **квадратичная форма** в $\mathbb{R}^N$, связанная с матрицей $\mathbf{A}$, — это отображение Q, определяемое как

$$Q(\mathbf{x}) := \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{j=1}^{N} \sum_{i=1}^{N} a_{ij} x_i x_j$$

Пример. Пусть $N = 2$ и $\mathbf{A}$ — единичная матрица $\mathbf{I}$. Тогда

$$Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$$

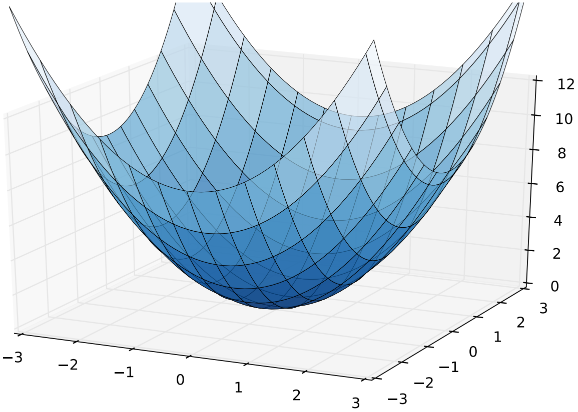

Рис.: Квадратичная функция $Q(\mathbf{x}) = x_1^2 + x_2^2$

Внимание:

- График для $Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$ лежит всюду выше нуля

Матрица $\mathbf{A}$ с квадратичной формой с указанным выше свойством $Q(\mathbf{x}) \geq 0$ называется *положительно определенной*

В более общем смысле, симметричная матрица $\mathbf{A}$ размера $N \times N$ называется

- **неотрицательно определенной**, если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \geq 0$ для всех $\mathbf{x} \in \mathbb{R}^N$,
- **положительно определенной**, если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} > 0$ для всех $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$,
- **неположительно определенной**, если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \leq 0$ для всех $\mathbf{x} \in \mathbb{R}^N$, и
- **отрицательно определенной**, если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} < 0$ для всех $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$.

Если $\mathbf{A}$ не подходит ни к одной из этих категорий, то $\mathbf{A}$ называется **неопределенной**

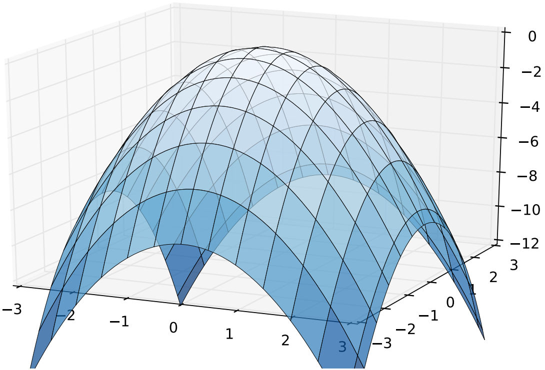

Рис.: Квадратичная функция $Q(\mathbf{x}) = -x_1^2 - x_2^2$

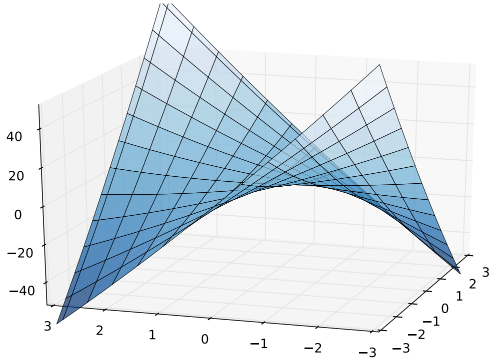

Рис.: Квадратичная функция $Q(\mathbf{x}) = x_1^2/2 + 8x_1x_2 + x_2^2/2$

Когда матрица $\mathbf{A}$ диагональная:

$$\mathbf{A} = \mathrm{diag}(d_1, \ldots, d_N) \quad \text{подразумевает} \quad Q(\mathbf{x}) = d_1 x_1^2 + \cdots + d_N x_N^2$$

Диагольная матрица положительно определена, тогда и только тогда, когда все диагональные элементы положительны

Факт. (??) Пусть $\mathbf{A}$ — некоторая симметричная матрица. Матрица $\mathbf{A}$

1. положительна определена тогда и только тогда, когда все ее собственные значения положительны
2. отрицательно определена тогда и только тогда, когда все ее собственные значения отрицательны

...аналогично для неположительно и неотрицательно определенных

Факт. (??) Если $\mathbf{A}$ положительно определена, то $\mathbf{A}$ несингулярна, $\det \mathbf{A} > 0$

Необходимое (но не достаточное) условие для каждого вида определенности:

Факт. (??) Если $\mathbf{A}$ положительна определена, то каждый элемент a_{nn} на главной диагонали положительный, то же самое для неотрицательной, неположительной и отрицательной.

Матрицы проекции

Напомним, что для любого подпространства $\mathbb{R}^N$, S, соответствующая проекция $\mathbf{P} = \operatorname{proj} S$ является линейным отображением из $\mathbb{R}^N$ в $\mathbb{R}^N$

Вспомним теорему **??**: существует матрица $\hat{\mathbf{P}}$ размера $N \times N$, такая что $\mathbf{P}\mathbf{x} = \hat{\mathbf{P}}\mathbf{x}$ для всех $\mathbf{x} \in \mathbb{R}^N$

- с этого момента $\mathbf{P}$ также будет означать соответствующую матрицу

Как выглядит эта матрица?

Теорема. (??) Пусть S — подпространство в $\mathbb{R}^N$. Если $\mathbf{P} = \operatorname{proj} S$, то

$$\mathbf{P} = \mathbf{B}(\mathbf{B}^\intercal\mathbf{B})^{-1}\mathbf{B}^\intercal \tag{2}$$

для каждой матрицы $\mathbf{B}$, такой что столбцы $\mathbf{B}$ формируют базис S

Смотрите упражнение **??** для доказательства

- Матрица $\mathbf{M} = \mathbf{I} - \mathbf{P}$ обозначает остаточную проекцию (смотрите страницу **??**)

Пример. Вспомним пример ?? на странице ??

Мы выяснили, что проекция $\mathbf{y} \in \mathbb{R}^N$ на $\mathrm{span}\{\mathbf{1}\}$ — это $\bar{y}\mathbf{1}$

Тот же результат с помощью теоремы (??):

- Так как $\mathbf{1}$ является базисом $\mathrm{span}\{\mathbf{1}\}$:

$$\mathbf{P} = \mathrm{proj}\ \mathrm{span}\{\mathbf{1}\} \implies \mathbf{P} = \mathbf{1}(\mathbf{1}^\intercal \mathbf{1})^{-1}\mathbf{1}^\intercal = \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal$$

- Значит, $\mathbf{P}\mathbf{y} = \bar{y}\mathbf{1}$, как и ожидалось
- Соответствующая остаточная проекция

$$\mathbf{M}_c = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal$$

Факт. (??) По постановке теоремы **??**, получается

1. $\mathbf{MB} = \mathbf{0}$
2. $\mathbf{PB} = \mathbf{B}$

Докажите в качестве упражнения (упр. **??** в ET)

Легко заметить, что $\mathbf{M}_c$ в предыдущем примере отображает $\mathbf{1}$ в $\mathbf{0}$

Квадратная матрица $\mathbf{A}$ является **идемпотентной**, если $\mathbf{A}\mathbf{A} = \mathbf{A}$

Факт. (??) И $\mathbf{P}$, и $\mathbf{M}$ симметричны и идемпотентны

(Упражнение: проверьте прямыми вычислениями)

Интуиция: проецирование на подпространство дважды — это то же самое, что проецировать один раз – вспомните факт **??** на странице **??**

Факт. (??) Пусть S является линейным подпространством в $\mathbb{R}^N$. Если $\mathbf{P} = \operatorname{proj} S$ и $\mathbf{M}$ — остаточная проекция, то

1. $\operatorname{rank} \mathbf{P} = \operatorname{trace} \mathbf{P} = \dim S$ и
2. $\operatorname{rank} \mathbf{M} = \operatorname{trace} \mathbf{M} = N - \dim S$

Доказательство.

- Ранг линейного отображения — это размерность его диапазона. Когда $\mathbf{P} = \operatorname{proj} S$, Диапазон $\mathbf{P}$ равняется S
- Чтобы показать, что $\operatorname{trace} \mathbf{P} = \dim S$ также соблюдается, используем факт **??**– $\operatorname{trace} \mathbf{P} = \dim S$,
- Это следует из того, что $\operatorname{trace} \mathbf{M} = N - \dim S$, потому что

$$\operatorname{trace} \mathbf{M} = \operatorname{trace}(\mathbf{I} - \mathbf{P}) = \operatorname{trace} \mathbf{I} - \operatorname{trace} \mathbf{P} = N - \dim S$$

Переопределенные системы уравнений

Рассмотрим систему уравнений в виде $\mathbf{Ax} = \mathbf{b}$, когда:

- Матрица $\mathbf{A}$ размера $N \times K$ имеет полный ранг столбцов
- Вектор $\mathbf{x}$ имеет размер $K \times 1$
- Вектор $\mathbf{b}$ имеет размер $N \times 1$
- $K \leq N$

Принимая как данность $\mathbf{A}$ и $\mathbf{b}$, мы ищем $\mathbf{x} \in \mathbb{R}^K$, такой что $\mathbf{Ax} = \mathbf{b}$

Если $K = N$, то система имеет ровно одно решение

Когда $N > K$, система уравнений считается **переопределенной**:

- количество уравнений $>$ количества неизвестных
- количество ограничений $>$ количества степеней свободы

Возможно, не удастся найти $\mathbf{b}$, который удовлетворяет всем N уравнениям

Вспомним линейное отображение $T\colon \mathbb{R}^K \to \mathbb{R}^N$, где $T\mathbf{x} = \mathbf{A}\mathbf{x}$

Следующие утверждения эквивалентны:

1. существует $\mathbf{x} \in \mathbb{R}^K$, такой что $\mathbf{A}\mathbf{x} = \mathbf{b}$
2. вектор $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$
3. вектор $\mathbf{b} \in \operatorname{rng} T$

Теорема **??** на странице **??**: когда $K < N$, функция T не может быть сюръекцией – возможные $\mathbf{b}$ находятся вне диапазона T

Когда $K < N$, случай с $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$ является “очень редким“, потму что:

- точка $\mathbf{b}$ является произвольной точкой в $\mathbb{R}^N$
- пространство $\operatorname{colspace} \mathbf{A}$ имеет размерность K
- подпространста $\mathbb{R}^N$ с размерностями K имеют “нулевую меру Лебега” – “шанс” того, что $\mathbf{b}$ лежит в этом подпространстве, крошечный

Стандартный подход: признать, что точного решения может не существовать

Следует сосредоточиться на поиске $\mathbf{x} \in \mathbb{R}^K$, чтобы $\mathbf{A}\mathbf{x}$ оказалось настолько близко к $\mathbf{b}$, насколько возможно

- близки по обычной Евклидовой норме

Задача минимизации называется **методом наименьших квадратов**

$$\hat{\mathbf{x}} := \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^K} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \tag{3}$$

При условии, что $\mathbf{A}$ размера $N \times K$ с $K \leq N$ и $\mathbf{b}$ размера $N \times 1$, мы можем использовать теорему ортогональной проекции для решения (3)

Теорема. (??) Если $\mathbf{A}$ имеет полный ранг столбцов, то (3) имеет единственное решение

$$\hat{\mathbf{x}} := (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \tag{4}$$

Доказательство.

Пусть:

- $\mathbf{A}$ и $\mathbf{b}$ будут как в формулировке теоремы
- $\hat{\mathbf{x}}$ как в (4) и
- $S := \operatorname{colspace} \mathbf{A}$

При условии полного ранга столбцов, столбцы $\mathbf{A}$ формируют базис для S. Применяем теорему **??**, ортогональная проекция $\mathbf{b}$ на S:

$$\mathbf{Pb} := \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top \mathbf{b} = \mathbf{A}\hat{\mathbf{x}} \tag{5}$$

Поскольку теорема об ортогональной проекции дает единственный минимизатор в терминах ближайшей точки в S к $\mathbf{b}$,

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{y}\| \quad \text{для всех } \mathbf{y} \in S,\ \mathbf{y} \neq \mathbf{A}\hat{\mathbf{x}} \tag{6}$$

Доказательство. (прод.) Возьмем любой $\mathbf{x} \in \mathbb{R}^K$, такой что $\mathbf{x} \neq \hat{\mathbf{x}}$

Получается $\mathbf{A}\mathbf{x} \in S$

К тому же, так как $\mathbf{x} \neq \hat{\mathbf{x}}$ и $\mathbf{A}$ имеют полный ранг столбцов, должно быть $\mathbf{A}\mathbf{x} \neq \mathbf{A}\hat{\mathbf{x}}$ (упр. **??**)

Значит

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \quad \text{для всех } \mathbf{x} \in \mathbb{R}^K,\ \mathbf{x} \neq \hat{\mathbf{x}}$$

Другими словами, $\hat{\mathbf{x}}$ является единственным решением (3)

В (4) матрица $(\mathbf{A}^\intercal\mathbf{A})^{-1}\mathbf{A}^\intercal$ называется **псевдообратной** матрицы $\mathbf{A}$

Если $K = N$, то решение методом наименьших квадратов $\hat{\mathbf{x}}$ в (4) сокращается до:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Что произойдет, если столбцы $boldA$ не будут линейно независимыми?

- множество $\operatorname{colspace} \mathbf{A}$ все еще линейное подпространство и теорема ортогональной проекции все еще дает нам ближайшую точку $\mathbf{Pb}$ к $\mathbf{b}$ в $\operatorname{colspace} \mathbf{A}$
- так как $\mathbf{Pb} \in \operatorname{colspace} \mathbf{A}$, все еще существует вектор $\hat{\mathbf{x}}$, такой что $\mathbf{Pb} = \mathbf{A}\hat{\mathbf{x}}$
- но таких векторов существует бесконечное множество

Смотрите упражнение **??**

QR-разложение

QR-разложение данной матрицы $\mathbf{A}$ — это произведение вида$\mathbf{QR}$

- первая матрица имеет ортонормированные столбцы и
- вторая является верхней треугольной

Приложения включают задачи наименьших квадратов и вычисление собственных значений

Теорема. (??) Если матрица $\mathbf{A}$ размера $N \times K$ имеет полный ранг столбцов, то существует факторизация $\mathbf{A} = \mathbf{QR}$, где

1. $\mathbf{R}$ размера $K \times K$ является верхней треугольной и несингулярной, и
2. $\mathbf{Q}$ размера $N \times K$ имеет ортонормированные столбцы

Смотрите страницу **??** в ET для доказательства

Возьмем разложение $\mathbf{A} = \mathbf{QR}$, рещение методом наименьших квадратов $\hat{\mathbf{x}}$, определенное в (4), может также записываться как:

$$\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{b}$$

Смотрите упр. **??**

Умножение на $\mathbf{R}$:

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\mathsf{T}}\mathbf{b}$$

Диагонализация и спектральная теория

Если $f\colon A \to A$ и $g\colon B \to B$, то говорят, что g **топологически сопряжен** с f везде, где существует непрерывная биекция $\tau\colon B \to A$, такая что

$$f = \tau \circ g \circ \tau^{-1}$$

Может быть полезно, если g проще f

Говорят, что квадратная матрица $\mathbf{A}$ **подобна** матрице $\mathbf{B}$, если существует обратимая матрица $\mathbf{P}$, такая что $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$

$$\begin{tikzcd}
\bullet \arrow[r, "\mathbf{B}"] & \bullet \arrow[d, "\mathbf{P}"] \\
\mathbf{x} \arrow[u, "\mathbf{P}^{-1}"] \arrow[r, "\mathbf{A}"] & \mathbf{A}\mathbf{x}
\end{tikzcd}$$

Рис.: $\mathbf{A}$ подобна $\mathbf{B}$

Если $\mathbf{A}$ подобна диагональной матрице, то $\mathbf{A}$ называется **диагонализируемой**

Нас интересует подобие простым матрицам, диагональные матрицы — самый простой вид

Факт. (??) Если $\mathbf{A}$ подобна $\mathbf{B}$, то $\mathbf{A}^t$ подобна $\mathbf{B}^t$ для всех $t \in \mathbb{N}$

Пример. Мы хотим вычислить $\mathbf{A}^t$ для некоторых данных $t \in \mathbb{N}$

Если $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ для некоторых $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, то по факту **??** и факту **??**, получается

$$\mathbf{A}^t = \mathbf{P}\,\mathrm{diag}(\lambda_1^t, \ldots, \lambda_N^t)\mathbf{P}^{-1}$$

Диагонализация и собственные значения

Факт. (??) Если $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ для некоторых $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$, то $(\operatorname{col}_n \mathbf{P}, \lambda_n)$ является собственной парой $\mathbf{A}$ для любого n

Доказательство. Заметим, что $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ значит, что $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}$

Equating the nth column on each side gives Рассмотрим столбец n с каждой стороны:

$$\mathbf{A}\mathbf{p}_n = \lambda_n \mathbf{p}_n$$

Где $\mathbf{p}_n := \operatorname{col}_n \mathbf{P}$

Заметим, что $\mathbf{p}_n$ — ненулевой вектор, так как $\mathbf{P}$ обратима

Но когда $\mathbf{A}$ диагонализируема?

Факт. (3.3.7) Матрица $\mathbf{A}$ размера $N \times N$ диагонализируема тогда и только тогда, когда она имеет N линейно независимых собственных векторов

В некоторых случаях, мы можем получить еще более простое разложение матрицы, если матрица $\mathbf{P}$ имеет ортогональные столбцы

Такие матрицы называются **ортогональными матрицами**

Факт. (??) Если $\mathbf{Q}$ и $\mathbf{P}$ — ортогональные матрицы размера $N \times N$, то

1. $\mathbf{Q}^\mathsf{T}$ ортогональна и $\mathbf{Q}^{-1} = \mathbf{Q}^\mathsf{T}$,
2. $\mathbf{QP}$ ортогонально, и
3. $\det \mathbf{Q} \in \{-1, 1\}$.

Если $\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}$ и $\mathbf{Q}$ имеют ортогональные столбцы, то

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathsf{T}}$$

Ясно, что $\mathbf{A}$ должна быть симметричной. Следующая теорема показывает, что симметрия $\mathbf{A}$ также достаточна

Теорема. (??) Если $\mathbf{A}$ симметрична, то $\mathbf{A}$ может быть диагонализована как $\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathsf{T}}$, где $\mathbf{Q}$ — ортогональная матрица и $\boldsymbol{\Lambda}$ — диагональная матрица, сформированная из собственных значений $\mathbf{A}$

Вышеупомянутая теорема была версией **теоремы о спектральном разложении**

$\mathbf{Q\Lambda Q}^\intercal$ называется **симметричным разложением на собственные значения** матрицы $\mathbf{A}$ – воздействие $\mathbf{A}$ на вектор $\mathbf{x}$ размера $N \times 1$:

$$\mathbf{Ax} = \sum_{n=1}^{N} \lambda_n (\mathbf{u}_n^\intercal \mathbf{x}) \mathbf{u}_n$$

где λ_n является n-ым собственным значением $\mathbf{A}$ и $\mathbf{u}_n = \mathrm{col}_n\, \mathbf{Q}$

Сравните с $\mathbf{x} = \sum_{n=1}^{N} (\mathbf{u}_n^\intercal \mathbf{x}) \mathbf{u}_n$

Факт. (??) Если $\mathbf{A}$ неотрицательно определена, то $\sqrt{\mathbf{A}}$ существует и равняется $\mathbf{Q}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^{\mathsf{T}}$. Матрица $\sqrt{\mathbf{\Lambda}}$ равняется $\mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$

Факт. (??) Если $\mathbf{A}$ положительно определена, то существует несингулярная, верхняя треугольная матрица $\mathbf{R}$, такая что $\mathbf{A} = \mathbf{R}^{\mathsf{T}}\mathbf{R}$

Такое разложение называется разложением Холецкого

Доказательство. (Разложение Холецкого) Мы можем написать:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top = \mathbf{Q}\sqrt{\mathbf{\Lambda}}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\top = (\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\top)^\top\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\top$$

Затем применим QR-разложение к $\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\top$:

$$\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\top = \tilde{\mathbf{Q}}\mathbf{R}$$

где $\mathbf{R}$ несингулярная и верхняя треугольная, и $\tilde{\mathbf{Q}}$ имеет ортонормированные столбцы

Так как столбцы $\tilde{\mathbf{Q}}$ ортонормированные,

$$\mathbf{A} = (\tilde{\mathbf{Q}}\mathbf{R})^\top\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\top\tilde{\mathbf{Q}}^\top\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\top\mathbf{R}$$

Нормы и непрерывность

Возьмем векторную последовательность $\{\mathbf{x}_n\}$ в $\mathbb{R}^K$ и любую точку $\mathbf{x} \in \mathbb{R}^K$. Говорят, что $\{\mathbf{x}_n\}$ **сходится** к $\mathbf{x}$ и записывается $\mathbf{x}_n \to \mathbf{x}$, если для любых $\epsilon > 0$, существует $N \in \mathbb{N}$, такое что $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ при $n \geq N$

Эквивалентно, последовательность действительных чисел $z_n := \|\mathbf{x}_n - \mathbf{x}\|$ сходится к нулю ы $\mathbb{R}$ при $n \to \infty$

Факт. (3.3.11) Имеют место следующие результаты:

1. если $\mathbf{x}_n \to \mathbf{x}$ и $\mathbf{y}_n \to \mathbf{y}$, то $\mathbf{x}_n + \mathbf{y}_n \to \mathbf{x} + \mathbf{y}$.
2. если $\mathbf{x}_n \to \mathbf{x}$ и $\alpha \in \mathbb{R}$, то $\alpha\mathbf{x}_n \to \alpha\mathbf{x}$.
3. $\mathbf{x}_n \to \mathbf{x}$ тогда и только тогда, когда $\mathbf{a}^\top\mathbf{x}_n \to \mathbf{a}^\top\mathbf{x}$ для всех $\mathbf{a} \in \mathbb{R}^K$.

Мы хотим распространить понятие сходимости на матрицы

Норма матрицы $\mathbf{A}$ размера $N \times K$:

$$\|\mathbf{A}\| := \max \left\{ \|\mathbf{A}\mathbf{x}\| \ : \ \mathbf{x} \in \mathbb{R}^K, \ \|\mathbf{x}\| = 1 \right\} \tag{7}$$

Значение нормы матрицы в общем случае найти непросто

Однако, матричная норма ведет себя как векторная норма

Факт. (3.3.12) Для любых согласующихся матриц $\mathbf{A}$ и $\mathbf{B}$, матричная норма удовлетворяет

1. $\|\mathbf{A}\| \geq 0$ и $\|\mathbf{A}\| = 0$ тогда и только тогда, когда все записи $\mathbf{A}$ нулевые,
2. $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$ для любых скаляров α,
3. $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$, и
4. $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$.

Факт. (??) Для любых матриц $\mathbf{A}$ размера $J \times K$ с элементами a_{jk}, выходит

$$\|\mathbf{A}\| \leq \sqrt{JK} \max_{jk} |a_{jk}|$$

если каждый элемент $\mathbf{A}$ близок к нулю, то $\|\mathbf{A}\|$ тоже близка к нулю

Ряд Неймана

В дальнейшем мы изучаем динамические системы вида

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$$

Существует ли “стационарный“ вектор $\mathbf{x} \in \mathbb{R}^N$ в том смысле, что $\mathbf{x}_t = \mathbf{x}$ значит, что $\mathbf{x}_{t+1} = \mathbf{x}$?

Мы ищем $\mathbf{x} \in \mathbb{R}^N$, являющийся решением системы уравнений

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b} \qquad (\mathbf{A} \text{ is } N \times N \text{ и } \mathbf{b} \text{ is } N \times 1) \tag{8}$$

Рассмотрим случай со скалярами $x = ax + b$

Если $|a| < 1$, то существует единственное решение

$$\bar{x} = \frac{b}{1-a} = b \sum_{k=0}^{\infty} a^k$$

Лемма о рядах Неймана помогает обобщить это до $\mathbb{R}^N$

Теорема. (??) Если матрица $\mathbf{A}$ квадратная и $\|\mathbf{A}^j\| < 1$ для некоторых $j \in \mathbb{N}$, то $\mathbf{I} - \mathbf{A}$ обратима, и

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{i=0}^{\infty} \mathbf{A}^i$$

Когда выполняется условие леммы о рядах Неймана, (8) имеет единственное решение

$$\bar{\mathbf{x}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

Чтобы проверить это решение, используем **спектральный радиус** $\mathbf{A}$:

$$\varrho(\mathbf{A}) := \max\{|\lambda| : \lambda \text{ является собственным значением } \mathbf{A}\}$$

$|\lambda|$ — это **модуль** возможно комплексного числа λ

Факт. Если $\varrho(\mathbf{A}) < 1$, то $\|\mathbf{A}^j\| < 1$ для некоторых $j \in \mathbb{N}$

Почему достаточно $\varrho(\mathbf{A}) < 1$?

Нам нужно $\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A})$ было близко к $\mathbf{I}$ для больших t

Получается:

$$\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A}) = \sum_{i=0}^{t} \mathbf{A}^i - \sum_{i=0}^{t} \mathbf{A}^{i+1} = \mathbf{I} - \mathbf{A}^{t+1}$$

Упростим до случая, где $\mathbf{A}$ диагонализируется:

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$$

где $\mathbf{\Lambda}$ — диагональная матрица, содержащая собственные значения $\lambda_1, \ldots, \lambda_N$ матрицы $\mathbf{A}$ на главной диагонали.

Теперь используем факт **??**,

$$\mathbf{A}^t = \mathbf{P} \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_N^t \end{pmatrix} \mathbf{P}^{-1}$$

Если $\varrho(\mathbf{A}) < 1$, то $|\lambda_n| < 1$ для всех n, и значит $\lambda_n^t \to 0$ при $t \to \infty$. Из этого следует, что $\mathbf{A}^t \to \mathbf{0}$