Математика для Data Science. Линейная алгебра. Шпаргалка

Содержание

Четвёртая неделя. Матричные разложения

Комплексные числа

Комплексное число z — это упорядоченная пара действительных чисел (a, b). Запись: $z = a + bi$.

Число a называют *действительной частью* числа $a + bi$ и обозначают $\mathrm{Re}(z)$. Число b называют *мнимой частью* числа $a + bi$ и обозначают $\mathrm{Im}(z)$.

Множество всех комплексных чисел обозначают буквой $\mathbb{C}$.

Складываются и вычитаются комплексные числа покоординатно: $(a + bi) + (c + di) := (a + c) + (b + d)i$

Умножение комплексных чисел определяется так: $(a + bi)(c + di) := (ac - bd) + (ad + bc)i$

Комплексное число $z = a + bi$ можно представить как вектор из действительных чисел $(a, b) \in \mathbb{R}^2$. Длина вектора z называется *модулем* числа z и обозначается $|z|$. Угол между вектором z и осью OX называется *аргументом* числа z и обозначается $\arg(z)$. У числа 0 длина определена и равна 0, но аргумент у него не определён.

Теорема. При умножении комплексных чисел модули перемножаются, а аргументы складываются. То есть для любых не равных нулю $z_1, z_2 \in \mathbb{C}$ выполнено:

1. $|z_1 z_2| = |z_1| \cdot |z_2|$
2. $\arg(z_1 z_2) = \arg(z_1) + \arg(z_2)$

Многочлен — это функция вида: $P(x) := a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$.

Корнем многочлена P называется число, при подстановке которого вместо x в выражение $P(x)$ получается 0. Другими словами, число c называется коренем многочлена P, если и только если $P(c) = 0$.

Утверждение. $\mathbb{C}$ алгебраически замкнуто, то есть любой многочлен с комплексными коэффициентами имеет корень в $\mathbb{C}$.

Собственные векторы

Собственным вектором преобразования $A : \mathbb{R}^n \to \mathbb{R}^n$ называется ненулевой вектор $\vec{v} \in \mathbb{R}^n$, такой что $A\vec{v} = \lambda\vec{v}$ для какого-нибудь числа $\lambda \in \mathbb{R}$.

Если $\det(A - \lambda E) = 0$, то число λ называется *собственным числом* преобразования A. Собственный вектор затем ищется из уравнения $A\vec{v} = \lambda\vec{v}$.

Свойства собственных векторов.

1. В базисе из собственных векторов матрица имеет диагональный вид
2. Если $\vec{v}$ это собственный вектор A, то и $c\vec{v}$ это собственный вектор A, где c – любое ненулевое число
3. Каждому собственному числу соответствует хотя бы один собственный вектор
4. Пусть $\vec{v}_1, \dots, \vec{v}_k$ это собственные векторы с различными собственными числами $\lambda_1, \dots, \lambda_k$. Тогда $\vec{v}_1, \dots, \vec{v}_k$ линейно независимы.

Спектральное разложение

Утверждение. Выражение $\det(A - \lambda E)$ это многочлен степени n от λ. То есть

$$\det(A - \lambda E) = b_n \lambda^n + b_{n-1} \lambda^{n-1} + \dots + b_1 \lambda + b_0,$$

где $b_n, b_{n-1}, \dots, b_0$ – какие-то действительные числа, определяемые по коэффициентам матрицы A, $b_n = (-1)^n$.

Пусть у многочлена $\det(A - \lambda E)$ существует n различных корней $\lambda_1, \dots, \lambda_n$. Найдём соответствующие числам $\lambda_1, \dots, \lambda_n$ собственные векторы $\vec{v}_1, \dots, \vec{v}_n$.

В базисе $\vec{v}_1, \dots, \vec{v}_n$ преобразование A имеет диагональную матрицу:

$$\Lambda := \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ & & & & \lambda_n \end{pmatrix}.$$

Пусть V — матрица перехода из базиса $\vec{v}_1, \dots, \vec{v}_n$ в стандартный базис, то есть её столбцы равны $\vec{v}_1, \dots, \vec{v}_n$. Значит, выполнено: $A = V \cdot \Lambda \cdot V^{-1}$.

Множество собственных чисел матрицы называется *спектром* матрицы. Построенное выше разложение $A = V \cdot \Lambda \cdot V^{-1}$ называется *спектральным разложением* матрицы A.

Когда действительных корней меньше, чем n, то надо переходить к комплексным числам. Почти все конструкции из курса переносятся с $\mathbb{R}$ на $\mathbb{C}$, кроме:

- Для скалярного произведения, углов и ортогональных преобразований требуется операция *сопряжения*, которая меняет знак у мнимой части: если $z = a + bi$, то сопряжённое число это $\bar{z} := a - bi$.
- Определителем диагональной комплексной матрицы называют произведение элементов на диагонали. А определитель любой недиагональной матрицы вычисляют через метод Гаусса.

Если некоторые корни $\det(A - \lambda E)$ всё же совпадают, то спектрального разложения может не существовать. Обобщение спектрального разложения матрицы это *Жорданово разложение* матрицы, которое существует всегда.

Низкоранговое приближение матрицы

Рекомендательная система

Строки матрицы A размера $n \times m$ отвечают за пользователей, а столбцы — за фильмы. В каждой клетке матрицы A стоит оценка от -2 до 2, либо знак звёздочка $*$, если оценки нет.

Цель — уметь предсказывать, какую оценку пользователь даст фильму, который он ещё не смотрел, чтобы затем решать, какие фильмы ему рекомендовать.

Есть два подхода к решению задачи рекомендаций: Content based filtering (основан на знании контента) и Collaborative filtering (основан только на оценках).

Из векторов-строк $\vec{u}_i$, отвечающих за предпочтения i-ого пользователя, составим матрицу U размера $n \times k$. А из векторов-столбцов $\vec{m}_j$, описывающих j-ый фильм, составим матрицу M размера $k \times m$. Мы пытаемся приблизить A произведением UM, а это то же самое, что приблизить матрицу A матрицей ранга не больше k.

Функция потерь в данном случае равна $L(U, M) = \sum\limits_{a_{ij} \neq *} (a_{ij} - \langle \vec{u}_i, \vec{m}_j \rangle)^2$.

Минимум функции потерь ищется либо градиентным спуском, либо с помощью EM-алгоритма (находим минимум $L(U, M)$ как многочлена второй степени от M при фиксированной U, а затем наоборот фиксируем M и т.д.)

Сингулярное разложение — SVD

У любой матрицы A размера m на n (у которой известны все элементы) существует такое разложение в произведение трёх матриц $A = U \Sigma V^T$, где

- U — ортогональная матрица размера m на m
- V^T — ортогональная матрица размера n на n
- Σ — диагональная матрица размера m на n, при этом диагональные элементы $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq 0$ неотрицательны и упорядочены по убыванию.

Числа σ_i называются *сингулярными числами* матрицы A. Поэтому такое разложение называется *сингулярным разложением*. На английский это переводится как *singular value decomposition*, или коротко *SVD*.

Пусть дана диагональная матрица Σ с диагональными элементами $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$. Сохраним первые k диагональных элементов, а остальные заменим на нули. Полученную матрицу обозначим Σ_k.

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & \sigma_{k+1} & & \\ & & & & \sigma_{k+2} & \\ & & & & & \ddots \end{pmatrix}, \Sigma_k := \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix}$$

Обозначим $A_k := U\Sigma_k V^T$.

Для двух матриц X и Y одинакового размера *норма Фробениуса* равна $||X - Y||_F := \sqrt{\sum_{i,j}(x_{ij} - y_{ij})^2}$.

Теорема. Дана матрица A. Минимум $||A - B||_F^2$ при ограничении $\mathrm{rank}(B) \leq k$ достигается при $B = A_k$.

Доля *объяснённой дисперсии* при этом равна $\frac{||A_k||_F^2}{||A||_F^2} = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{l} \sigma_i^2}$.

Другие матричные разложения

Главная диагональ матрицы A размера m на n — это все элементы a_{ii}, где $i \leq \min(n, m)$.

Матрица, у которой все элементы ниже главной диагонали равны нулю, называется *верхнетреугольной*:

$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}$$

Аналогично, матрица, у которой все элементы выше главной диагонали равны нулю, называется *нижнетреугольной*.

LU-разложение

Если A это квадратная обратимая матрица (с некоторым незначительным условием), то существует разложение $A = LU$, где

- L это нижнетреугольная матрица (L от "Lower triangular")
- U это верхнетреугольная матрица (U от "Upper triangular")
- размеры L и U совпадают с размером A

Иногда используют модификации LU, называемые LDP и LUP.

QR-разложение

Если A имеет размер m на n, где $m \geq n$, то существует разложение $A = QR$, где

- Q это ортогональная матрица размера m на m
- R это верхнетреугольная матрица размера m на n

Заметьте, что из-за верхнетреугольности последние $(m - n)$ строк матрицы R всегда оказываются заполнены нулями:

$$\begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$