Практическая линейная алгебра

Минкин Даниэль

25 февраля 2025 г.

Содержание

1 Введение

1.1 Темы

Вот ключевые темы курса:

- Псевдообратная матрица
- Линейная регрессия и МНК
- Полиномиальная интерполяция, сплайны, кривые Безье (возможно сплайны Безье)
- Аппроксимация (лишь часть связанная с лин. алгеброй) (Многочлены Чебышева)
- Метрики и нормы
- Оценка погрешности стандартных задач линейной алгебры
- Итерационные алгоритмы
- Проблемы собственных значений и собственных векторов
- Неотрицательные и положительные матрицы и их специальные свойства (Теорема Фробениуса — Перрона). Пример использования данной теоремы — RageRank
- Функции от матриц
- (Если успеем) Системы алгебраических уравнений
- (Если успеем) Линейные экономические модели

В процессе также будут затронуты:

- Разложения матриц
- Приближение малого ранга (a.k.a PCA в общем виде)

1.2 Организация курса

Будет 2 онлайн кр (как на выч. стате дают сложные задачи на подумать) в конце каждого из модулей. Формула оценки задается как:

$$0.5 \cdot \text{КР1} + 0.5 \cdot \text{КР2} + \text{Бонус} \tag{1}$$

Бонус будет формироваться в зависимсоти от активности на парах, НО основная его часть будет получена за формирование доклада по линейной алгебре, который будет презентован на одной из пар.

Чел упомянул интересную тему для доклада: существует теорема, по которой мы с помощью двух линейных слоев и функции активации можно аппроксимировать любую функцию с заданной точностью

2 Разложение полного ранга (скелетное разложение)

Пусть A матрица (m, n), такая что $\text{rang}(A) = r$. Разложение $A = FG$, где F имеет размерность (m, r), а G имеет размерность (r, n) называется разложением полного ранга.

Заметим, что $\text{rang}(F) <= r$ из-за размера, при этом $\text{rang}(F) >= r$ по свойству умножения матриц. Следовательно $\text{rang}(F) = r$, это матрица полного столбцового ранга, аналогично получаем, что G матрица полного строкового ранга. Таким образом матрица A представима, как произведение двух полнорангвых матриц

2.1 Теорема о существовани скелетеного разложения

Для любой матрицы существуют скелетное разложение

Доказательство

Покажем, что для любой матрицы существует скелетное разложение, у нас есть r независимых столбцов внутри матрицы. Выберем первые r столбцов (мы можем так сделать так, как нам не важно какие брать, остальные точно могут быть выражены через них), тогда матрица A представима как $A = (\bar{a}_1 \ ... \ \bar{a}_r \ \sum_{i=1}^{r} w_{1,i}\bar{a}_i \ ... \ \sum_{i=1}^{r} w_{n-r,i}\bar{a}_i)$. Пусть тогда матрица F равна

$(\bar{a}_1 \dots \bar{a}_r)$, она имеет размер (m, r). По условию отбора данных столбцов они все линейно независимы, таким образом $\mathrm{rang}(F) = r$, так как $\mathrm{rang}(F) <= r$ (по условию размера матрицы) и при этом $\mathrm{rang}(F) >= r$ (так как все столбцы ЛНЗ).

Рассмотрим матрицу G, которая выглядит так $G = (e(1) \dots e(r)\ w_1 \dots w_{n-r})$, где $e(i)$ — вектор с размерностью m где все координаты равны 0, а i - ая координата равна 1. Такая матрица имеет размерность (r, n) и ее первые r столбцов линейно независимы, а остальные выражаются через них, следовательно $\mathrm{rang}(G) = r$. При этом $FG = A$. Ч.Т.Д

2.2 Практический способ нахождения склетного разложения

Рассмотрим практический способ нахождения данного разложения. Приведем матрицу A с размерами (m, n) к каноническому виду (на диагонали 1, а под ними ничего). У нас получится r ступенек, вырежем верхний левый блок канонической матрицы, размером (r, n). Далее выберем столбцы у которых в каноническом виде в столбце всего лишь одна 1 и во всех остальных местах нули. Добавим эти столбцы в новую матрицу, тогда это будет матрица F

3 Псевдообратная матрица

Почему мы не можем всегда использовать метод Гаусса для решения СЛАУ?

Самый большой минус - сложность задаваемая как $O(n^2)$ для квадратной матрицы n на n.

Вторая проблема - аккумуляция погрешности при использовании данного метода, пример: при вычитании первой строчки из второй, ошибка первой строчки суммируется со второй. Плюс также погрешность создают float-ы (особенно с первой причиной вместе)

Вспомним, что такое обратная матрица. Это матрица A^{-1} такая, что

$$AA^{-1} = A^{-1}A = \mathrm{E} \qquad (2)$$

Пусть матрица A имеет размеры m на n, попробуем проверить условия, которыми задаются условия для обратной матрицы

Тогда $AA^{-1} = \mathrm{E}$, тут Е будет иметь размеры m на m, а A^{-1} иметь размеры n на m. В обратном же случае $A^{-1}A = \mathrm{E}$ матрица Е будет иметь размеры n на n.

Если мы будем использовать лишь одно из условий, тогда очень легко определить обратную матрицу

3.1 Немного о ранге

rang может быть определен как

- Наибольшее кол-во линейно независимых строк/столбцов в матрице
- $\dim(\mathrm{Im}(A))$, где A — наша матрица, а Im — образ нашей матрицы.

Вспоминаем условие про ранг суммы и произведения

Ранг суммы

Рассмотрим выражение $\mathrm{rang}(A+B)$ и попробуем его оценить. Очевидно, что $\mathrm{rang}(A+B) <= \mathrm{rang}(A)+\mathrm{rang}(B)$. Почему так? По свойству линейных операторов можно заметить, что $(A+B)v = Av+Bv$, следовательно, любой вектор из $\mathrm{Im}(A+B)$ является подпространством $\mathrm{Im}(A)+\mathrm{Im}(B)$. При этом $\mathrm{Im}(A)+\mathrm{Im}(B) = \{u+v|u \in \mathrm{Im}(A), v \in \mathrm{Im}(B)\}$. По свойству пространств $\dim(U+V) = \dim(U)+\dim(V)-\dim(U\cap V)$ (Это уже на 100% интуитивно). Следовательно $\dim(U+V) <= \dim(U)+\dim(V)$. Применяя данное уравнение к образам, получим $\dim(\mathrm{Im}(A+B)) <= \dim(\mathrm{Im}(A))+\dim(\mathrm{Im}(B))$, что исходя из определения rang приводит к $\mathrm{rang}(A+B) <= \mathrm{rang}(A)+\mathrm{rang}(B)$. Ч.Т.Д

Ранг произведения

$\mathrm{rang}(AB) <= \max(A, B)$. В данном случае посмотрим на суть произведения матриц — это композиция линейных операторов. Т.е функция от функции, сначала применяется линейный оператор B, а к результату применяется оператор A. Пусть A — матрица (m, n), а B — матрица (n, k). Рассмотрим процесс работы данной композиции, к нам приходит вектор с размерностью k, его отображают в пространство с размерностью $\mathrm{rang}(B)$, а после к нему применяется оператор A, заметим, что он не может отобразить вектор в пространство с размерностью больше чем $\mathrm{rang}(B) \Rightarrow \mathrm{rang}(AB) <= \mathrm{rang}(B)$, при этом выходная размерность также не может превысить $\mathrm{rang}(A)$. Таким образом $\mathrm{rang}(AB) <= \min(\mathrm{rang}(B), \mathrm{rang}(A)) <= \max(\mathrm{rang}(B), \mathrm{rang}(A))$

3.2 Снова возврат к псевдообратной матрице

$$\mathrm{rang}(AA^{-1}) = m \tag{3}$$

$$\mathrm{rang}(A^{-1}A) = n \tag{4}$$

При этом $\mathrm{rang}(A) <= \min(m, n)$. Исходя из неравенств выше получаем, что $\mathrm{rang}(A) >= m$ и $\mathrm{rang}(A) >= n$ (Так как максимальный элемент из $\mathrm{rang}(A)$ и $\mathrm{rang}(A^{-1})$ больше данных значений). **Однако в таком случае возникает противоречие, если** $m \neq n$. Это говорит о том, что мы не можем использовать такие аксиомы для нашей “обратной“ матрицы, которая может быть применена в общем случае

3.3 Рассмотрим виды СЛАУ

Какие виды систем уравнений бывают:

- **Определенная** — у нас есть матрица (m, n), где $m > n$. Система имеет единственное решение. Мы хотим представить решение как Jb, когда уравнение задается как $Ax = b$, где J — произвольная матрица с размерами (n, m).
- **Не определенная** — решений бесконечно много. У нас возникает вопрос как выразить хоть какое то решение
- **Не совместна** — решений нет. Мы хотим получить наиболее правдоподобное приближение к решению

3.4 Эрмитова матрица

Введем операцию комплексного (эрмитова) сопряжения для матрицы. Пусть дана матрица A, тогда эрмитово сопряженная матрица (которая обозначается как A^*) это транспонированная матрица A в которой каждый элемент заменили на комплексно сопряженный ему. Т.е для любого $a \in \mathbb{C} \backslash \mathbb{R}$ в матрице A $\overline{a + bi} = a - bi$. Новая матрица называется **эрмитово сопряженной**

Эрмитова матрица — квадратная матрица, которая при транспонировании равна совей эрмитово сопряженной матрице

$$A^T = A^* \tag{5}$$

NOTE: Любая симметричная матрица является эрмитовой

3.5 Наконец то опредлеим псевдообратную матрицу

Пусть A — некая прямоугольная матрица. Матрица C называется псевдообратной матрицей (а точнее псевдообратной Мура-Пенроуза), если выполняются следующие 4 аксиомы Пенроуза

- $ACA = A$
- $CAC = C$
- $(AC)^* = AC$ (i.e это эрмитова матрица)
- $(CA)^* = CA$ (тоже эрмитова)

Важные замечания которые следуют из аксиом

- Если A — матрица (m, n), то C имеет размер (n, m)

- Если A — невырожденная квадрантная матрица, то псевдообратная матрица совпадет с обратной матрицей

Покажем что данная матрица единственна

1. Пусть B и C — две псевдообратные матрицы для A. 2. Тогда:

$$B = BAB = B(ACA)B = (BA)(CA)B.$$

3. Поскольку BA и CA эрмитовы, их можно переставлять в сопряжённом виде:

$$\begin{aligned} BACAB &= (BA)^*(CA)^*B = A^*B^*A^*C^*B = (ABA)^*C^*B \\ &= A^*C^*B = (CA)^*BAB = CABAB \\ &= CAB. \end{aligned}$$

4. Аналогично:

$$C = CAC = C(ABA)C = (CA)(BA)C = CAB.$$

5. Из $B = CAB$ и $C = CAB$ следует $B = C$, что доказывает единственность.

Введем специальный символ

Псевдообратную матрицу принято обозначать, как A^+

3.6 Немного опредлений

Матрица A (m, n) называется **матрицей полного столбового ранга**, если $\text{rang}(A) = n$. При этом $m > n$. Т.е все ее столбцы Л.Н.З

Матрица A (m, n) называется **матрицей полного строкового ранга**, если $\text{rang}(A) = m$. При этом $m < n$. Т.е все ее строки Л.Н.З

Вставил не по порядку рассказа чисто по своему желанию:

Вспомним определение ядра: $\ker(A) = \{v | Av = \bar{0}\}$. При этом по свойству ядра $\dim(\text{Im}(A)) + \dim(\ker(A)) = n$, для матрицы (m, n)

3.7 Примеры работы

3.7.1 Частный случай

Рассмотрим матрицу $(1\ 1)$. Нам нужно найти псевдообратную к ней, мы точно значем, что псевдообратная матрица будет иметь размерность $(2, 1)$. Т.е псевдобратная матрица это некий вектор $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$. Запишем, все аксиомы:

- $(1\ 1)\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)(1\ 1) = (1\ 1)$
- $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)(1\ 1)\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$
- $((1\ 1)\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right))^* = (1\ 1)\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$
- $(\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)(1\ 1))^* = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)(1\ 1)$

Заметим, что третий пункт всегда выполняется, значит его можно исключить, выполним перемножение матриц

- $(a + b)(1\ 1) = (1\ 1)$
- $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)(a + b) = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$
- $\left(\begin{smallmatrix} a & a \\ b & b \end{smallmatrix}\right)^* = \left(\begin{smallmatrix} a & a \\ b & b \end{smallmatrix}\right)$

Следовательно $a = b$ и $a + b = 1$ таким образом псевдообратная матрица имеет вид $\left(\begin{smallmatrix} 0.5 \\ 0.5 \end{smallmatrix}\right)$

3.7.2 Общий случай

В общем случае для вектора $\bar{a}$ псевдообратная матрица будет иметь вид:

$$\bar{a}^+ = \frac{\bar{a}^*}{\|a\|^2} \tag{6}$$

Покажем, что все аксиомы выполняются

- $\bar{a}\frac{\bar{a}^*}{\|a\|^2}\bar{a} = \bar{a}$
- $\frac{\bar{a}^*}{\|a\|^2}\bar{a}\frac{\bar{a}^*}{\|a\|^2} = \frac{\bar{a}^*}{\|a\|^2}$
- $(\bar{a}\frac{\bar{a}^*}{\|a\|^2})^* = \bar{a}\frac{\bar{a}^*}{\|a\|^2}$
- $(\frac{\bar{a}^*}{\|a\|^2}\bar{a})^* = \frac{\bar{a}^*}{\|a\|^2}\bar{a}$

Это равносильно (Note: для расчета нормы комплексного вектора используется не транспонирование, а комплексное сопряжение)

- $\bar{a} \cdot 1 = \bar{a}$
- $1 \cdot \frac{\bar{a}^*}{\|a\|^2} = \frac{\bar{a}^*}{\|a\|^2}$

- $(\bar{a}\frac{\bar{a}^*}{\|a\|^2})^* = \bar{a}\frac{\bar{a}^*}{\|a\|^2}$
- $1^* = 1$

Таким образом нам нужно, чтобы выполнялось выражение

$$(\bar{a}\frac{\bar{a}^*}{\|a\|^2})^* = \bar{a}\frac{\bar{a}^*}{\|a\|^2} \tag{7}$$

Сократим константу

$$(\bar{a}\bar{a}^*)^* = \bar{a}\bar{a}^* \tag{8}$$

Раскроем эрмитово сопряжение

$$\bar{a}\bar{a}^* = \bar{a}\bar{a}^* \tag{9}$$

Следовательно, аксиома 3 тоже выполняется. А по условию единственности это и есть псевдообратная матрица, легко показать, что для вектора строки ситуация будет аналогичной

3.8 Лемма об обратимости A^*A

Пусть A — матрица полного столбцового ранга, то матрица A^*A — невырожденная квадрантная матрица. (Для матрицы полного строкового ранга все практически 1 в 1 сейм)

Доказательство

Пусть A — матрица (m, n), такая, что $n < m$ и $\mathrm{rang}(A) = n$. Заметим, что $\mathrm{rang}(A^*A) <= \mathrm{rang}(A)$ (по свойству произведения матриц). Тогда для док-ва нам достаточно показать, что $\mathrm{rang}(A^*A) >= \mathrm{rang}(A)$. По свойству ядра, если X имеет размерность (m, n), то $\dim(\ker(X)) + \mathrm{rang}(X) = n$. Пусть $x \in \ker(A^*A)$, тогда $A^*Ax = 0$, домножим на x^* справа, тогда выражение примет вид $(Ax)^*Ax = 0$. Это скалярное произведение вектора Ax на самого себя и по свойству скалярного произведения мы получаем, что $Ax = 0$, при этом $\dim(\ker(A)) = 0$, следовательно $\dim(\ker(A^*A)) = 0$, а тогда $\mathrm{rang}(A^*A) = n$

В общем случае мы доказали, что $\ker(A^*A) \subset \ker(A)$, отсюда вытекает неравенство $\dim(\ker(A^*A)) <= \dim(\ker(A))$, то равносильно $n - \mathrm{rang}(A^*A) <= n - \mathrm{rang}(A)$, что приводит нас к $\mathrm{rang}(A^*A) >= \mathrm{rang}(A)$

3.9 Теорема о существовании псевдообратной матрицы для матриц полного столб. и строч. ранга

Если A матрица (m, n), такая что $n < m$ и что $\mathrm{rang}(A) = n$ (матрица полного столбцового ранга), то A^+ существует и равна

$$A^+ = (A^*A)^{-1}A^* \tag{10}$$

Для матрицы полного строчного ранга

$$A^+ = A^*(AA^*)^{-1} \tag{11}$$

Проверим аксиомы для матрицы полного столбцового ранга

- $A(A^*A)^{-1}A^*A = A$
- $(A^*A)^{-1}A^*A(A^*A)^{-1}A^* = (A^*A)^{-1}A^*$
- $(A(A^*A)^{-1}A^*)^* = A(A^*A)^{-1}A^*$
- $((A^*A)^{-1}A^*A)^* = (A^*A)^{-1}A^*A$

Это сводится к (3 точно выполняется)

- $AE = A$
- $(A^*A)^{-1}EA^* = (A^*A)^{-1}A^*$
- $(E)^* = E$

Следовательно, данная матрица удовлетворяет всем четырем аксиомам и единственна

Аналогично, для матриц полного строчного ранга

3.10 Теорема о существовании псевдообратной матрицы для любой матрицы

3.10.1 Лемма об обратной матрице произвдения матриц

Если $A = BC$, где B — матрица полного столбцового ранга, а C — матрица полного строкового ранга, то тогда $A^+ = B^+C^+$,

Доказательство

Рассмотрим, аксиомы для A, пусть псевдообратнная матрица $A^+ = C^+B^+$, если для нее будут выполнены аксиомы, то тогда по условию единственности, это и есть псевдообртаная матрица

- $AC^+B^+A = A$
- $C^+B^+AC^+B^+ = C^+B^+$
- $(AC^+B^+)^* = AC^+B^+$
- $(C^+B^+A)^* = C^+B^+A$

Заменим A на BC, тогда

- $BCC^+B^+BC = BC$
- $C^+B^+BCC^+B^+ = C^+B^+$
- $(BCC^+B^+)^* = BCC^+B^+$
- $(C^+B^+BC)^* = C^+B^+BC$

Так как мы знаем, что за матрицы B и C мы можем выразить их псевдообратные матрицы явно

- $BC(C^*(CC^*)^{-1})((B^*B)^{-1}B^*)BC = BC$
- $(C^*(CC^*)^{-1})((B^*B)^{-1}B^*)BC(C^*(CC^*)^{-1})((B^*B)^{-1}B^*) = (C^*(CC^*)^{-1})((B^*B)^{-1}B^*)$
- $(BC(C^*(CC^*)^{-1})((B^*B)^{-1}B^*))^* = BC(C^*(CC^*)^{-1})((B^*B)^{-1}B^*)$
- $((C^*(CC^*)^{-1})((B^*B)^{-1}B^*)BC)^* = (C^*(CC^*)^{-1})((B^*B)^{-1}B^*)BC$

Начнем раскрывать скобки

- $BC = BC$
- $(C^*(CC^*)^{-1})((B^*B)^{-1}B^*) = (C^*(CC^*)^{-1})((B^*B)^{-1}B^*)$
- $(B((B^*B)^{-1}B^*))^* = B(B^*B)^{-1}B^*$
- $((C^*(CC^*)^{-1})C)^* = (C^*(CC^*)^{-1})C$

Легко заметить, что все 4 аксиомы выполняются

3.10.2 Основная теорема

Для любой матрицы существует псевдообратная матрица

3.10.3 Начало доказательства

Мы знаем, что такая матрица есть для невырожденных матриц, матриц полного столбцового и строкового рангов и нулевых матриц, рассмотрим ситуацию для других матриц, которые не попадают в эту категорию. Мы знаем, что для любой матрицы можно применить скелетное разложение, так и сделаем для произвольной матрицы A, тогда $A = FG$, по свойству леммы о произведении матриц мы знаем, что $A^+ = G^+F^+$. При этом, мы знаем, что для любой матрицы полного столбцового ранга и полного строчного ранга, существует псвдообратная матрица, следовательно $A^+ = G^*(GG^*)^{-1}(F^*F)^{-1}F^*$. Следовательно, для любой матрицы существует псевдообратная.

3.11 Суть псевдообратной матрицы

В случае несовместных систем псевдообратная матрица позволяет получить наиболее точное приближенное решение по МНК. Сформулируем задачу более точно. Вектор $\bar{u}$ называется приближенным решением системы $A\bar{x} = \bar{b}$ по МНК или псевдорешением, если для любого $\bar{x}$ выполняется

$$\|A\bar{x} - \bar{b}\| \geq \|A\bar{u} - \bar{b}\| \tag{12}$$

Используется Евклидова норма, легко заметить, что если система имеет решение, то оно совпадает с псевдорешением

3.12 Теорема о псевдорешении

Вектор $\bar{u} = A^{+}\bar{b}$ является псевдорешением и имеет наименьшую длину из всех псевдорешений

Доказательство

Заметим, что $\mathrm{Im}(A)$ порождено столбцами матрицы A, а все векторы $\ker(A)$ ортогональны строкам матрицы A. Следовательно для любых $\bar{x} \in \mathrm{Im}(A)$ и $\bar{y} \in \ker(A^*)$ выполняется $\bar{x} \cdot \bar{y} = 0$. Аналогично и в обратную сторону для $\bar{x} \in \mathrm{Im}(A^*)$ и $\bar{y} \in \ker(A)$

3.12.1 Лемма

Для любого $\bar{x} \in \mathrm{Im}(M)$ и $\bar{y} \in \mathrm{Im}(A)$, выполняется $\bar{x} \cdot \bar{y} = 0$, если $M = AA^{+} - E$

Доказательство

Если мы покажем, что $\mathrm{Im}(M) \subset \ker(A^*)$, тогда точно лемма будет доказана по свойству выше. Рассмотрим вектор $y = Mx \in \mathrm{Im}(M)$, покажем, что $A^*y = 0$. $A^*Mx = 0$, следовательно $(A^*AA^{+} - A^*)x = 0$. По аксиоме три $(A^*(AA^{+})^* - A^*)x = 0$. Это равняется $((AA^{+}A)^* - A^*)x = 0$. Что приводится к $(A^* - A^*)x = 0$. Таким образом данное равенство выполняется всегда. **Ч.Т.Д**

3.12.2 Возврат к доказательству

Если $\bar{x} \cdot \bar{y} = 0$, то тогда для $\bar{z} = \bar{x} + \bar{y}$ выполняется следующее равенство $\|\bar{z}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2$. Покажем это $\|\bar{z}\| = \sqrt{\bar{z} \cdot \bar{z}}$, раскрывая это получим: $\sqrt{(\bar{x} + \bar{y}) \cdot (\bar{x} + \bar{y})}$, что приводит нас к выражению $\sqrt{\bar{x} \cdot \bar{x} + \bar{y} \cdot \bar{y}}$. Следовательно мы получаем, что $\|\bar{z}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2$. Ч.Т.Д

Отсюда мы получаем неравенство: $\|\bar{z}\| \geq \|\bar{x}\|$ (или $\bar{y}$)

Покажем, что $u = A^+b$ является псевдорешением. Рассмотрим $f(x) = Ax - b$ нам нужно показать, что такой вектор имеет длину больше чем $AA^+b - b$. Мы можем заметить, что второе выражние равно $(AA^+ - E)b$. А $Ax \in \mathrm{Im}(A)$, следовательно эти два вектора перпендикулярны.

Мы можем записать $f(x)$ как $Ax - Au + Au - b$, где $u = A^+b$

$$Ax - Au + Au - b = A(x-u) + (AA^+ - E)b \tag{13}$$

Легко увидеть, что данные слагаемые перпендикулярны, таким образом исходя из неравенства выше

$$\|A(x-u) + (AA^+ - E)b\| \geq \|(AA^+ - E)b\| \tag{14}$$

Раскрывая скобки получим, что для любого x

$$\|f(x)\| \geq \|AA^+b - b\| \tag{15}$$

Легко заметить, что равенство достигается при

$$A(x - A^+b) = 0 \tag{16}$$

Покажем, что такое решение имеет минимальную длину по сравнению с другими псевдорешениями

Пусть x другое псевдорешение, т.е $A(x-u) = 0$

Покажем, что $(x-u) \perp u$, при условии $A(x-u) = 0$. Рассмотрим их произвдение

$$(x-u)^* \cdot u = b^* {A^+}^* (x-u) \tag{17}$$

При этом ${A^+}^* = (A^+AA^+)^* = {A^+}^* A^+ A^*$. Итого мы получаем, что скалярное произвдение имеет вид

$$b^* {A^+}^* A^+ A(x-u) = 0 \tag{18}$$

Ч.Т.Д

Используя прошлое неравенство мы получим, что

$$\|x - u + u\| \geq \|u\| \tag{19}$$

Мы доказали, что псевдорешение A^+b имеет минимальную длину среди всех возможных

3.12.3 Как описать все псевдорешения?

Все псевдорешения могут быть описаны формулой

$$A(x - A^+b) = 0 \tag{20}$$

Можно заметить, что, если

$$x = A^+b - (A^+A - E)y \tag{21}$$

где y — это произвольный вектор, то тогда подставив такой x в формулу мы получим, что

$$A(A^+b - (A^+A - E)y - A^+b) = -A(A^+A - E)y = -(A - A)y = 0 \tag{22}$$

Следовательно любой такой вектор подходит в качестве псевдорешения, т.е минимизирует норму между собой и целевым вектором

4 Матричные разложения

4.1 LU разложение

LU разложение позволяет представить матрицу как произведение двух матриц L — нижнетреугольная матрица с единицами на диагонали, а U - верхнетреугольная

4.1.1 Теорема об LU разложении

Пусть A — матрица (n, n) и все угловые подматрицы невырождены. Тогда существует единсвтенное разложение $A = LU$, где L — нижнетреугольная с единицами на диагонали, а U — верхнетреугольная матрица

Доказательство

Докажем по индукции, что такое разложение существует, база индукции выглядит так: для матрицы $(1, 1)$ мы получим $L = \left(\begin{smallmatrix} 1 \end{smallmatrix} \right)$ и $U = \left(\begin{smallmatrix} a_{11} \end{smallmatrix} \right)$

Пусть разложение существует для любой матрицы размерами $(n-1, n-1)$ для которой выполнены требования, покажем что в таком случае оно существует для любой матрицы A с размерами (n, n). Представим A в блочном виде

$$A = \begin{bmatrix} A_{11} & a_{12} \\ a_{21} & a_{nn} \end{bmatrix} \tag{23}$$

Тут матрица A_{11} имеет размеры $(n-1, n-1)$ по условию все угловые миноры ненулевые, следовательно для данной подматрицы сущесвтвует LU разложение $A_{11} = L_{11}U_{11}$

Мы хотим представить A как LU, рассмотрим как должны выглядеть такие матрицы в предположении, что L и U будут занимать соответсвующие им блоки

$$L = \begin{bmatrix} L_{11} & 0 \\ l_{21} & 1 \end{bmatrix} \tag{24}$$

и

$$U = \begin{bmatrix} U_{11} & u_{12} \\ 0 & u_{nn} \end{bmatrix} \tag{25}$$

Тогда произведение матриц задается как

$$LU = \begin{bmatrix} L_{11}U_{11} & L_{11}u_{12} \\ l_{21}U_{11} & l_{21}u_{12} + u_{nn} \end{bmatrix} \tag{26}$$

Нужно заметить, что l_{12} — вектор строка

Запишем наши равенства

$$\begin{cases} L_{11}u_{12} = a_{12} \\ l_{21}U_{11} = a_{21} \\ l_{21}u_{12} + u_{nn} = a_{nn} \end{cases}$$

Тут l_{12} — вектор строка и u_{12} — вектор столбец

L_{11} однозначно обратима, а U_{11} так как все угловые миноры неотрицатльны ($\det(A_{11}) = \det(L_{11}) \cdot \det(U_{11})$ и при этом $\det(A_{11}) \neq 0$)

$$\begin{cases} u_{12} = L_{11}^{-1}a_{12} \\ l_{21} = a_{21}U_{11}^{-1} \\ u_{nn} = a_{nn} - a_{21}U_{11}^{-1}L_{11}^{-1}a_{12} = a_{nn} - a_{21}A_{11}^{-1}a_{12} \end{cases}$$

Мы можем заметить, что такое разложение единственно, и для него строго необходимо, чтобы матрица в левом верхнем блоке обращалась

4.1.2 Практический подход для LU разложения

Мы можем привести матрицу A к нижнетреугольному виду с помощью метода Гаусса, а после найти U как $U = L^{-1}A$

4.1.3 Где используется

Можно выделить следующие use-case:

- Быстрое решение систем вида $Ax = y$. Такая система превращается в две треугольные системы при ее представлении как: $LUx = y$

- Численно стабильное решение задачи нахождения det, так как мы можем использовать натуральный логарифм от детерминанта и при этом сохранять численную стабильность решения, так как $\det(A) = \det(L) \cdot \det(U) = 1 \cdot \prod_i^n U_{ii}$

- Более удобное обращение раскладываемой матрицы

4.2 LUP-разложение

Данное разложение является обобщением LU разложения на случай любой квадратной невырожденной матрицы. Любая матрица может быть представлена в виде:

$$PA = LU \tag{27}$$

Т.е мы выполняем разложение для матрицы PA, где P отвчеает за перестановку строк в матрице A. Мы всегда можем привести невырожденную матрицу A к такому виду, чтобы к ней можно было применить LU разложение, это следует из метода окаймляющих миноров для определения ранга.

Хорошим алгоритмом для LUP разложения является алгоритм Дулиттла (рассмотрим позже), который позоляет удобно векторизовать вычисления. Однако алгоритмическая сложность все также составляет $O(n^3)$

4.3 LDL^T разложение

В случае если матрица A является эрмитивой ($A^* = A$) и положительно определенной, то тогда LU разложение будет связанно с LDL^T разложением. Покажем это:

Так как матрица A положительно определена все угловые миноры положительны (по критерию Сильвестра), следовательно существует LU разложение. Также по условию эрмитовости мы получаем:

$$A = LU = A^* = U^*L^* \tag{28}$$

Т.е мы получаем равенство которое нам нужно решить

$$LU = U^*L^* \tag{29}$$

Заметим, что всегда существует такая матрица D, что

$$U = DL^* \tag{30}$$

так как L^* всегда обратима и мы можем выразить D как UL^{*-1}. U должна быть верхнетреугольной, следовательно матрица D будет диагонаальной, так как она не должна менять структуру матрицы L^*, которая сама является верхнетреугольной.

Подставим $U = DL^*$ в исходное уравнение. Тогда:

$$LDL^* = LD^*L^* \tag{31}$$

Так как $D^* = D$ равенство выполняется (D — диагональная)

Таким образом, мы показали, что любую эрмитову положительно опредленную матрицу можно представить в виде следующего произвдения матриц, причем такое разложение единственно

$$A = LDL^* \tag{32}$$

где L — нижнетреугольная матрица, а D — диагональная

Заметим интересный факт: у D на диагонали стоят только положительные числа, так как матрица A положительно определена. Проверим этот факт: мы знаем, что $x^*Ax > 0$ для любого x. Т.е $x^*LDL^*x > 0$. Выразим $y = L^*x$, при этом L^* — невырожденная. Тогда выражение равносильно $y^*Dy > 0$ для любого y. Так как матрица D диагональная:

$$y^*Dy = \sum_{i=1}^{n} d_{ii} \cdot y_i^2 \tag{33}$$

где d_{ii} - элемент матрицы D

Следовательно все $d_{ii} > 0$, иначе найдется вектор для которого условие будет невыполнено

4.4 Разложение Холецкого

Пусть нам дана эрмитова квадратная положительно определенная матрица, тогда она представима в виде:

$$A = U^*U = LL^* \tag{34}$$

При этом на главной диагонали L или U будут находиться вещесвтенные положительные числа.

Заметим, что мы можем представить нашу матрицу так (прошлое разложение):

$$A = LDL^* \tag{35}$$

Так как матрица D — диагональная и имеет только положительные элементы на диагонали, мы можем определить диагональную матрицу $D^{0.5}$, такую что $D^{0.5}D^{0.5} = D$.

Следовательно:

$$A = LD^{0.5}D^{0.5*}L^* \tag{36}$$

Заметим, что тогда

$$A = L_1L_1^* \tag{37}$$

Где $L_1 = LD^{0.5}$, при этом $D^{0.5}$ не меняет структуру матрицы L, так как является диагональной матрицей. Также стоит отметить, что на диагонали у $D^{0.5}$ находятся положительные числа, а именно квадраты значений с диагонали D. Таким образом на диагонали матрицы L_1 будут стоять только положительные числа, что и требовалось. Следовательно мы показали, что можно единсвтенным образом представить матрицу A как:

$$A = LL^* \tag{38}$$

На практике лучше считать LDL^T разложение, так как оно не включает в себя взятие квадрантных корней

4.5 QR - разложение

QR разложение представляет собой создание оронормированного базиса в матричном виде. Вспомним про алгоритм ортоганализации: мы берем первый вектор как есть. Теперь нам нужно сделать так, чтобы второй вектор стал перпендикулярен первому. Мы можем прдставить вектор a_2 как сумму двух компонент: перпендикуляра к a_1 и некого остатка. Т.е

$$a_2 = b_2 + rem_2 \tag{39}$$

При этом $b_2 \perp a_1$

Мы можем найти остаток как $\frac{\langle a_2, a_1 \rangle}{\langle a_1, a_1 \rangle} a_1$

Можно считать, что на каждом этапе мы строим плоскость из векторов которые нашли до этого. При этом новый вектор не прнадлежит этой плоскости так как тогда не выполнялось бы условие о ЛНЗ. Мы всегда можем провести от плоскости перпендикуляр к некой точке, тогда наш вектор вне этой плоскости представим как сумма перпендикуляра от плоскости и некого вектора который находится на плоскости и ведет к точке где начинается перпендикуляр. Благодаря данному механизму любой новый вектор перпендикулярен всем предыдущим и все будущие вектора будут перпендикулярны текущему, так как он будет использоваться для создания плоскости. После того как все вектора будут найдены их будет необходимо отнормировать.

Что такое QR разложение?

Это представление матрицы A в виде

$$A = QR \tag{40}$$

где в столбцах матрицы Q стоят вектора из ортонормированного базиса (i.e $Q^*Q = E$). А R — верхнетреугольная матрица.

Самым простым способом нахождения разложения является нахождение Q через ортоганализацю Грамма Шмидта, и последующее нахождение R как $R = Q^{-1}A = Q^*A$

Однако эффективнее использовать алгоритм Хаусхолдера, для матрицы (m, n), где $m > n$. Его алгоритмическая сложность составляет $O(2mn^2 - \frac{2n^3}{3})$.

4.6 Спектральное разложение

Некоторые квадратные матрицы можно представить в виде:

$$A = UDU^* \tag{41}$$

где D — диагональная матрица с собственными числами на диагонали, а U — унитарная матрица (i.e $UU^* = U^*U = E$) (вообще то она ортонормированна по столбцам, но в случае квадратных матриц это приводит и к ортонормированности по строкам)

Для этого нужно затребовать, чтобы матрица была

- Эрмитовой (все ее собственные числа будут вещесвтенными)
- Унитарной (все ее собственные числа будут по модулю равны 1)
- Нормальной ($A^*A = AA^*$)

Для его нахождения мы сначала должны найти все собственные числа матрицы A, после найти весь набор собственных векторов A. Мы можем привести наши собственные вектора к ортонормированному базису и записать эти вектора в столбцы матрицы U. А в D мы запишем собтсвенные числа матрицы A.

Покажем, что нормальность необходима для данного разложения, пусть матрица не является нормальной, тогда $A^*A \neq AA^*$. Подставим сюда матрицу в разложенном виде

$$UDU^*UDU^* \neq UDU^*UDU^* \tag{42}$$

Сокращая часть множителей

$$UDDU^* \neq UDDU^* \tag{43}$$

Как мы видим возникает противоречие.

Таже из нормальности следует, что матрица имеет все собственные вектора (лень доказывать)

4.7 Сингулярное разложение (SVD)

Данное разложение является обобщением спектрального разложения на случай произвольной матрицы. Пусть A матрица (m, n), тогда собственные числа матрицы $A{*}A$, которая является Эрмитовой, будут вещественными числами. Таким образом мы можем взять корень из этих чисел.

$$\sigma(A)_i = \sqrt{\lambda(A^*A)_i} \tag{44}$$

где $\sigma(A)_i$ — называется сингулярным числом матрицы A

Есть договоренность в нумерации сингулярных чисел, что

$$\sigma_1 \geq \sigma_2 ... \geq \sigma_n \tag{45}$$

Легко заметить, что $\mathrm{rang}(A^*A) = \mathrm{rang}(A)$, следовательно кол-во сингулярных значений равно рангу матрицы A

Теорема

Для любой матрицы A размерами (m, n) сущесвтвует сингулярное разложение, такое, что

$$A = U\Sigma V^* \tag{46}$$

где:

- U — унитарная матрица (m, m)
- Σ —диагональная матрица с размерами совпадающими с матрицей A на на диагонали у которой стоят сингулярные значения
- V — унитарная матрица (n, n)

Доказательство

Рассмотрим две матрицы A^*A и AA^*. Во-первых покажем, что их ненулевые собственные числа, которые соотв. ненулевым собственнным векторам одинаковы. Пусть $A^*Av = \lambda v$. Если мы умножим обе части на A то получим, что $AA^*(Av) = \lambda(Av)$. Таким образом собственный вектор AA^* соотв. собственному значению λ, может быть получен через собственный вектор A^*A, который соотв. этому же значению через умножение на A. Также мы показали, что ненулевые собтсвенные значения данных матриц совпадают.

Каждая из этих матриц является эрмитовой, следовательно они могут быть разложены через спектральное разложение.

$$A^*A = V\Lambda_{(n,n)}V^* \tag{47}$$

и

$$AA^* = U\Lambda_{(m,m)}U^* \tag{48}$$

(Предположим, что в обоих случаях собственные значения расположены по убыванию).

Заметим, что так как мы рассматриваем эрмитовы матрицы их собственные вектора перпендикулярны. Покажем, что это так, пусть A — эрмитова матрица и $A^* = A$. Рассмотрим собственные вектора k_1 и k_2. Тогда $Ak_1 = \lambda_1 k_1$ и $Ak_2 = \lambda_2 k_2$. Взглянем на скалярное произвдения векторов

$$\begin{aligned} k_1^* k_2 = \frac{1}{\lambda_1} k_1^* A^* k_2 = \frac{1}{\lambda_1} k_1^* A k_2 \\ = \frac{\lambda_2}{\lambda_1} k_1^* k_2 \end{aligned}$$

Следовательно

$$\lambda_1 k_1^* k_2 = \lambda_2 k_1^* k_2 \tag{49}$$

Таким образом

$$(\lambda_1 - \lambda_2) \cdot k_1^* k_2 = 0 \tag{50}$$

Так как собственные числа не равны, то это возможно только если $k_1^* k_2 = 0$. Ч.Т.Д

Исходя из этого, нам остается только отнормировать данные вектора, что вставить их в матрицы U и V

Рассмотрим матрицу $U\Sigma V^*$ составленную по тем же правилам, что и для сингулярного разложения. Заметим, что тогда AA^* и A^*A сводятся к своим спектральным разложениям. Теперь нам нужно показать, что

$$U\Sigma V^* = A \tag{51}$$

Мы можем выразить наш вектор через комбинацию ортонормированных векторов из V

$$x = \sum_{i=1}^{r} v_i w_i \tag{52}$$

Покажем, что $U\Sigma V^* v_i = Av_i$ для любого v_i. По свойству ортонормированности $U\Sigma V^* v_i = U\Sigma e_i$ (так как мы взяли вектор-столбец из V). $U\Sigma e_i = \sigma_i U e_i$ (тут e_i из n - мерного стал m - мерным)

4.7.1 Отступление

Также стоит отметить, что есть вторая версия разложения — усеченное разложение

Матрица A будет представлена как

$$A = U_{(m,r)} \Sigma_{(r,r)} V^*_{(r,n)} \tag{53}$$

Мы просто вырезаем все строки и столбцы, которые умножаются на ноль

4.7.2 Intuition данного разложения

Рассмотрим матрицу A размерами (m, n). Это отображение $f : R^n \to R^m$. Таким образом, данное разложение говорит нам, что в исходном и выходном пространстве найдутся такие ортонормированные базисы, что действие оператора будет описываться диагональной матрицей.

Пусть у нас есть базис $v_1, v_2 ... v_n$ в исходном пространстве и $u_1, u_2 ... u_m$ в выходном тогда:

$$f(\sum_{i=1}^{n} x_i \cdot v_i) = \sum_{i=1}^{m} x_i \cdot \sigma_i \cdot u_i \tag{54}$$

4.7.3 Связь с нормой Фробениуса

Во-первых покажем, что норма Фробениуса инвариантна относительно унитарных преобразований слева (Под унитарным преобразованием мы можем подразуемвать, что столбцы матрицы задают ортонормированный базис (но не обязательно строки) т.е $U^*U = E$). Т.е нам нужно показать, что

$$\|UA\|_{fro} = \|A\|_{fro} \tag{55}$$

где U — матрица с ортонормированными столбцами ($U^*U = E$)

Легко инуитивно заметить, что оронормированный базис не должен менять норму вектора так как мы просто поворачиваем весь базис относительно исходного $e_1, e_2 ... e_n$.

Также легко заметить, что

$$\|B\|_{fro} = \sqrt{\sum_{i=1}^{k} \|B_{(*,i)}\|^2_{L2}} \tag{56}$$

Т.е норма Фробениуса выражается через квадраты L2 норм столбцов

Нам достаточно показать, что переход к новому ортонормированному базису не меняет норму вектора столбца и тогда утверждение будет доказано. Рассмотрим выражние Ux. Нам необходимо показать, что $\|Ux\|_{L2} = \|x\|_{L2}$. Заметим, что

$$\|x\|_{L2} = \sqrt{\langle x, x\rangle} = \sqrt{x^* \cdot x} \tag{57}$$

при этом

$$\|Ux\|_{L2} = \sqrt{\langle Ux, Ux\rangle} = \sqrt{x^* U^* \cdot Ux} = \sqrt{x^* \cdot x} \tag{58}$$

Таким образом:

$$\|A\|_{fro} = \sqrt{\sum_{i=1}^{k} \|A_{(*,i)}\|_{L2}^2} \tag{59}$$

Для UA

$$\|UA\|_{fro} = \sqrt{\sum_{i=1}^{k} \|U \cdot A_{(*,i)}\|_{L2}^2} \tag{60}$$

Исходя из предыдущего выражения мы знаем, что

$$\|U \cdot A_{(*,i)}\|_{L2} = \|A_{(*,i)}\|_{L2} \tag{61}$$

В итоге

$$\|UA\|_{fro} = \sqrt{\sum_{i=1}^{k} \|U \cdot A_{(*,i)}\|_{L2}^2} = \sqrt{\sum_{i=1}^{k} \|A_{(*,i)}\|_{L2}^2} = \|A\|_{fro} \tag{62}$$

Рассмотрим норму матрицы A

$$A = U\Sigma V^* \tag{63}$$

Тогда

$$\begin{gathered}\|A\|_{fro} = \|U\Sigma V^*\|_{fro} = \|\Sigma V^*\|_{fro} \\ = \|V\Sigma^*\|_{fro} = \|\Sigma^*\|_{fro} = \|\Sigma\|_{fro} \\ = \sum \sigma(A)^2\end{gathered}$$