МдАД: Линейная алгебра **Осень 2018**

Линейная алгебра 4: 20 октября

Преподаватель: Антон Савостьянов *Ассистент: Даяна Мухаметшина*

Контакты: *Антон Савостьянов, почта:* a.s.savostyanov@gmail.com, *telegram:* @mryodo
Даяна Мухаметшина, почта: dayanamuha@gmail.com, *telegram:* @anniesss1

Правила игры: *Домашние задания следует присылать в читаемом виде на почту преподавателя не позднее указанного при выдаче задания крайнего срока (дедлайна).*
При выполнении домашнего задания приветствуется использование среды LaTeX; допустим набор в редакторах Word (Libreoffice, Google Docs) и отсканированные письменные материалы.
Выполненное домашнее задание должно содержать решение задачи, по которому возможно восстановить авторский ход решения, а не только ответ.

4.1 Скалярное произведение и ортогональные системы векторов

Определение 4.1. Линейно независимые вектора $v_1, v_2, \dots v_k$ называются ортогональной системой, если для любой пары $i \neq j$ скалярное произведение $(v_i, v_j) = 0$.

Определение 4.2. Длиной вектора v называется величина $||v||= \sqrt{(v,v)}$. Ортогональная система, где каждый из векторов имеет длину 1, называется ортонормированной. Заметьте, что любую ортогональную систему можно привести к ортонормированной, поделив каждый вектор на его длину.

Рассмотрим систему линейно независимых векторов $v_1, v_2, \dots v_k$; они задают некоторое линейное пространство L. Найдем какой-нибудь ортогональный базис в данном пространстве.

Пусть такой базис называется $u_1, \dots u_k$. Тогда положим $u_1 = v_1$. Теперь, чтобы вектор u_2 был ортогонален всем векторам из уже построенного базиса, из него нужно убрать составляющие, неортогональные предыдущим векторам:

$$v_2 = au_1 + u_2 \;\Rightarrow\; (u_1, v_2) = a(u_1, u_1) + 0 \;\Rightarrow\; u_2 = v_2 - \frac{(u_1, v_2)}{(u_1, u_1)} u_1$$

Оператор $p_{u_1}(v_2) = \frac{(u_1,v_2)}{(u_1,u_1)} u_1$ называется оператором проекции вектора v_2 на вектор u_1. Тогда для построения u_3 требуется:

$$u_3 = v_3 - p_{u_1}(v_3) - p_{u_2}(v_3)$$

$$u_p = v_p - \sum_{t=1}^{p-1} p_{u_t}(v_p)$$

Описанную процедуру называют процессом ортогонализации Грамма-Шмидта.

Упражнение 1. Методом Грамма-Шмидта ортонормируйте следующие системы векторов:

(a) $e_1 = (1, 1), e_2 = (1, 2)$ в $\mathbb{R}^2$;
(b) $e_1 = (1, 1, 1), e_2 = (1, 2, 3)$ в $\mathbb{R}^3$;

Определение 4.3. Ортогональным дополнением подпространства L пространства V называется такое множество векторов $L^{\perp} = \{x \mid \forall v \in L\ (x, v) = 0\}$.

Упражнение 2. Докажите, что ортогональное дополнение пространства тоже является пространством.

Упражнение 3. Пользуясь известными фактами для линейных однородных систем, установите связь между размерностями $\dim L$ и $\dim L^{\perp}$.

Упражнение 4. Для следующего подпространства $U \subset \mathbb{R}^n$ со стандартным скалярным произведением опишите ортогональное дополнение к нему и постройте в нем ортонормированный базис:

(a) U — линейная оболочка вектора $(1, 1, 1)$;
(b) U — множество решений системы $x + 2y + 3z = 0$;
(c) U — линейная оболочка векторов $(1, 1, 1)$ и $(1, 2, 3)$.

Что же значат слова «со стандартным скалярным произведением»? Действительно, оказывается, что скалярным произведением называется любая функция $\beta(x, y)$, для которой

(a) $\beta(x, x) \geq 0$, причем $\beta(x, x) = 0$ только если $x = 0$;
(b) $\beta(x, y) = \beta(y, x)$
(c) $\beta(ax + by, z) = a\beta(x, z) + b\beta(y, z)$

Тогда несложно понять, что скалярных произведений можно задать достаточно много. Например, рассмотрим пространство V с базисом $e_1, e_2, \ldots e_n$. Пусть исследуется произведение векторов $x = (x_1, x_2, \ldots x_n)$ и $y = (y_1, y_2, \ldots y_n)$. Тогда:

$$(x, y) = (x_1 e_1 + x_2 e_2 + \ldots + x_n e_n, y_1 e_1 + y_2 e_2 + \ldots + y_n e_n) = \sum_{i,j=1}^{n} x_i y_j (e_i, e_j)$$

Таким образом, скалярное произведение полностью определяется скалярными произведениями базисных векторов. Несложно заметить, что тогда скалярное произведение может быть задано матрично:

$$(x, y) = x^T \begin{pmatrix} (e_1, e_1) & (e_1, e_2) & \ldots & (e_1, e_n) \\ (e_2, e_1) & (e_2, e_2) & \ldots & (e_2, e_n) \\ \vdots & \vdots & \ddots & \vdots \\ (e_n, e_1) & (e_n, e_2) & \ldots & (e_n, e_n) \end{pmatrix} y = x^T G y$$

Матрица G называется матрицей Грамма и задает скалярное произведение. Исходя из его свойств, матрица Грамма должна быть симметрична и положительно определена. Как выглядит матрица Грамма для ортонормированного базиса.

Упражнение 5. Рассмотрим пространство многочленов не старше 3ей степени. Выберите в нем некоторый базис и посчитайте матрицу Грамма в нем, считая, что скалярное произведение считает по правилу $(f, g) = \int_0^1 f(x)g(x)dx$. Получился ли ортонормированный базис? Как достичь его при помощи операций с квадратичными формами?

4.2 Квадратичные формы (добавление)

Определение 4.4. Вид квадратичной формы $q(x)$, при котором ее матрица $Q(x)$ принимает диагональный вид, называется каноническим видом квадратичной формы.

Определение 4.5. Вид квадратичной формы $q(x)$, при котором ее матрица $Q(x)$ принимает диагональный вид, причем на диагонали стоят только числа 0, 1 и -1 называется нормальным видом квадратичной формы.

Для симметрических матриц (а все матрицы квадратичных форм именно такие) справедлива такая замечательная теорема:

Теорема 4.6.

1. *Все собственные числа симметрической матрицы A вещественны;*
2. *Собственные вектора, принадлежащие различным собственным числам симметричной матрицы A попарно ортогональны (то есть их скалярное произведение равно 0);*
3. *Существует ортогональная замена переменных $X = TY$, приводящая квадратичную форму $f(X)$ к каноническому виду, где на диагонали матрицы квадратичной формы стоят собственные числа.*

Упражнение 6. Для каждой квадратичной формы найдите матрицу в стандартном базисе; ортонормированный базис, в котором матрица квадратичной формы диагональна; ортонормированный базис, в котором матрица квадратичной формы имеет нормальный вид:

(a) $2x^2 + 2xy + 2y^2$
(b) $5x^2 - 2xy - 2xz + 5xy - 2yz + 5z^2$

4.3 Метод наименьших квадратов

Часто на практике при исследовании какого-нибудь природного или социального явления делается допущение, что это явление описывается линейной формулой. Точнее, предполагается, что некоторая величина b линейно зависит от величин $a_1, a_2, \ldots, a_n$ и мы хотим найти эту зависимость, $b = a_1x_1 + \ldots + a_nx_n$ (модель линейной регрессии), то есть хотим найти неизвестные коэффициенты $x_1, \ldots, x_n$.

Для нахождения искомой зависимости $a_1, a_2, \ldots a_n$ проводится большое число измерений ($m >> n$, то есть число неизвестных существенно меньше числа уравнений; такие системы называют переопределенными, поскольку они с большой вероятностью не имеют точного решения), которые приводят к следующей системе:

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n = b_2 \\ \ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{cases} \iff Ax = b$$

Такая система как правило несовместна. Вместо ее решения, приходится искать наилучшее решение — то есть такой вектор x, который минимизирует суммарную ошибку в каждом уравнении.

Таких измерений ошибки в математике и машинном обучении рассматривается довольно много: сейчас нас интересует самый базовый случай, когда ошибка моделируется квадратичной функцией; такой метод получил название метода наименьших квадратов.

Определение 4.7. Псевдорешением («наилучшим решением») системы уравнений $Ax = b$, где система переопределена, называется такой вектор x^*, который минимизирует сумму квадратов разностей левых и правыъ частей уравнений, то есть

$$(a_{11}x_1+\ldots+a_{1n}x_n-b_1)^2+(a_{21}x_1+\ldots+a_{2n}x_n-b_2)^2+\ldots+(a_{m1}x_1+\ldots+a_{mn}x_n-b_m)^2 \to \min$$

среди всех возможных векторов $x \in \mathbb{R}^n$. Значением функции, заданной выше, называется квадратичным отклонением.

Несложно понять, что в матричном виде мы вычислили вектор $(Ax - b)$ и минизируем его длину (вернее, ее квадрат):

$$||Ax - b||^2 = (Ax - b, Ax - b) \to \min$$

Геометрически, решение такой системы методом МНК заключается в нахождении такой точки P, которая лежит ближе всего к точке b, при этом находясь в пространстве столбцов матрицы A. Несложно понять, что если нам требуется найти наиболее близкую точку к заданной в выбранном подпространстве, то достаточно просто обнаружить ее проекцию на заданное подпространство. Тогда искомое x^* должно быть проекцией b на подпространство, порожденное векторами $A^1, A^2, \ldots A^n$, а вектор $Ax * -b$ будет ортогональной составляющей («высотой») вектора b к пространству, порожденному $A^1, A^2, \ldots A^n$.

Заметим также, что вектор $Ax - b$ ортогонален пространству $A^1, A^2, \ldots A^n$, а значит ортогонален любой их линейной комбинации Av, то есть $(Ax - b, Av) = 0$ для любого вектора v. Отсюда получим следующую теорему:

Теорема 4.8. *Пусть все столбцы матрицы A, $A^1, A^2, \dots A^n$, линейно независимы. Тогда искомый вектор проекции x^*, известный как псведорешение, вычисляется по формуле:*

$$x^* = (A^T A)^{-1} A^T b$$

Определение 4.9. Если все столбцы матрицы A, $A^1, A^2, \dots A^n$, линейно независимы, то матрица $(A^T A)$ обратима. В данном случае матрицу $(A^T A)^{-1} A^T$ называют псевдообратной для столбцовой матрицы.

Упражнение 7. Найдите псевдорешения системы:

$$(a) \begin{cases} x + y = 0 \\ x - y = 0 \\ y = 1 \end{cases}$$

$$(b) \begin{cases} x + y - z = 1 \\ 2x + 3y + z = 6 \\ 3x - y + 4z = 6 \\ x + 2y + 3z = 7 \end{cases}$$

4.4 Сингулярное разложение

Определение 4.10. Ортогональным оператором называется оператор, сохраняющий расстояние, то есть

$$(x, y) = (Ax, Ay)$$

Для матрицы такого оператора верно свойство:

$$AA^T = A^T A = E$$

Несложно заметить, что ортогональный оператор всегда обратим (почему?) и $A^{-1} = A^T$

Для ортогонального оператора всегда верны следующие свойства:

(a) Строки и столбцы его матрицы образуют систему ортонормированных векторов. Таким образом, скалярное произведение любой строки на себя дает 1, а в противном случае 0.

(b) Определитель матрицы такого оператора всегда равен 1.

(c) Такой оператор сохраняет ортонормированность базиса (то есть отображает ортонормированный базис в ортонормированный).

(d) Действие оператора — движения пространства — могут быть сведены к конечному числу поворотов и отражений относительно некоторых осей, вне зависимости от размерности изучаемого пространства.

Упражнение 8. Подумайте, где вы уже видели такой ортогональный оператор?

Теорема 4.11. *Для любой вещественной матрицы A произвольного размера $m \times n$ существуют ортогональные матрицы U и V (обязательно квадратные размеров $m \times m$ и $n \times n$ соответственно) такие, что*

$$U^T AV = \Lambda,$$

где Λ — прямоугольная в общем случае матрица, у которой не нули стоят только на главной диагонали, причем количество ненулевых элементов на диагонали равно рангу матрицы A, а сами диагональные элементы упорядочены по невозрастанию.

Матрицы U и V называются левыми и правыми сингулярными векторами, а диагональные значения матрицы Λ называются сингулярными числами (значениями).

Заметим также, что засчет обнуления части столбцов и строк левых и правых сингулярных векторов, связанных с выходом за ранг матрицы A, такое разложение существенно экономит память, требуемую для хранения матрицы: в случае матрицы A требовалось mn ячеей памяти, в то время как сейчас вычеркивая ненужные нулевые строки и столбцы в матрице U останется mr элементов, в матрице Λ — r^2 элементов (или просто r, если хранить флаг о диагональности) и в матрице V — rm элементов; всего эта сумма не превосходит nr по памяти, при условии того, что одна размерность матрицы существенно больше другой (или ранг матрицы мал).

Аналогично наблюдается еще одно, наиболее полезное свойство сингулярного разложения: информация в матрице распределена соответственно невозрастанию сингулярных чисел. Эта фраза означает, что если выбрать самые маленькие сингулярные числа в матрице Λ и заменить их на 0, то мы во-первых, получим приближение исходной матрицы матрицей меньшего ранга, а во-вторых, это будет наилучшее приближение (с точки зрения Фробениусовой нормы; то есть фактически дисперсность векторов в матрице объясняется в большей степени сингулярными векторами и числами, соответствующими большим сингулярным значениям). На данной теореме основан метод уменьшения размерности данных, известный как метод главных компонент, principal component analysis.

Алгоритм поиска сингулярного разложения

(a) Вычислить матрицу $A^T A$. Убедитесь, что матрица симметрична и имеет действительный спектр.

(b) Решите характеристическое уравнение

$$\det(AA^T - \lambda E) = 0,$$

то есть найдите собственные числа данной матрицы.

(c) Найдите собственные вектора матрицы AA^T, соответсвующие найденным собственным числам, решая систему $(AA^T - \lambda E)x = 0$. Ее решениями будет левые сингулярные вектора.

(d) Запишите найденные вектора по столбцам в матрицу U, предварительно их отнормировав.

(e) Найдите собственные вектора матрицы $A^T A$, соответсвующие найденным собственным числам, решая систему $(A^T A - \lambda E)x = 0$. Ее решениями будет правые сингулярные вектора.

(f) Запишите найденные вектора по столбцам в матрицу V, предварительно их отнормировав.

(g) Альтернативно, зная матрицу U, можно вычислить матрицу V из разложения $U^T AV = \Lambda$.

Геометрическая интерпретация может быть приведена следующая: рассмотрим образ линейного преобразования A, примененного к единичной сфере — это гиперэллипсоид. Величины его главных полуосей — это сингулярные числа, а векторы, задающие направления главных полуосей — это столбцы матрицы U (векторы, задающие прообразы главных осей эллипсоида задаются столбцами матрицы V).

Упражнение 9. Найдите сингулярное разложение матрицы A. Установите ее наилучшее одноранговое приближение. Сильно ли различаются матрицы?

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$