Линейная алгебра

Дима Трушин

Семинар 8

Проекции

Пусть V – некоторое векторное пространство и $U, W \subseteq V$ – некоторые подпространства. Будем говорить, что V раскладывается в прямую сумму этих подпространств, если $U \cap W = 0$ и $V = U + W$, т.е. любой вектор $v \in V$ представляется в виде $v = u + w$, где $u \in U$ и $w \in W$ (то есть $U + W = \{u + w \mid u \in U,\ w \in W\}$). Думать про это надо так, U и W – это непересекающиеся подпространства и V является наименьшим пространством их содержащим.Такое разложение всегда получается так: берем какой-нибудь базис $e_1, \ldots, e_n$ пространства V, делим его на две части $e_1, \ldots, e_k$ и $e_{k+1}, \ldots, e_n$ и полагаем $U = \langle e_1, \ldots, e_k \rangle$ и $W = \langle e_{k+1}, \ldots, e_n \rangle$. Если пространство V является прямой суммой подпространств U и W, то мы будем обозначать это дело следующим образом $V = U \oplus W$. В этом случае любой вектор v единственным образом раскладывается в виде $v = u + w$, где $u \in U$ и $w \in W$. Еще в этом случае $\dim U + \dim W = \dim V$.

Утверждение. *Пусть V – евклидово пространство и $U \subseteq V$ – произвольное подпространство. Тогда $V = U \oplus U^\perp$.*

Таким образом в евклидовом пространстве V при фиксированном подпространстве $U \subseteq V$, любой вектор $v \in V$ единственным образом раскладывается в сумму $v = \operatorname{pr}_U v + \operatorname{ort}_U v$, где $\operatorname{pr}_U v \in U$ и $\operatorname{ort}_U v \in U^\perp$.

Определение. Если V – евклидово пространство, $U \subseteq V$ – произвольное подпространство и $v \in V$, то

- Вектор $\operatorname{pr}_U v$ называется ортогональной проекцией v на U.
- Вектор $\operatorname{ort}_U v$ называется ортогональной составляющей v относительно U.

Обратите внимание, что ортогональная проекция v на U – это проекция v на U вдоль $U^\perp$, а ортогональная составляющая – проекция v на $U^\perp$ вдоль U.

Формула БАБА

Давайте я в начале разберу задачу нахождения проекции вектора на подпространство вдоль другого подпространства (здесь нам не нужно никакое скалярное произведение). Пусть V – некоторое векторное пространство и $V = U \oplus W$. Тогда на пространстве V задан оператор проекции $P\colon V \to V$ такой, что $\ker P = W$ и $P|_U = \mathrm{Id}$, то есть, если $v \in V$ раскладывается в сумму $v = u + w$, где $u \in U$ и $w \in W$, то $Pv = u$ – оператор вычисления проекции на U вдоль W.

Теперь мы хотим научиться эффективно считать P. Для этого предположим $V = \mathbb{R}^n$, $U = \langle u_1, \ldots, u_k \rangle$, $W = \{y \in \mathbb{R}^n \mid Ay = 0\}$, где $A \in \mathrm{M}_{s\,n}(\mathbb{R})$. В этом случае $P\colon \mathbb{R}^n \to \mathbb{R}^n$ задается некоторой матрицей. Наша задача – найти эту матрицу.

Предположим для простоты, что векторы $u_1, \ldots, u_k$ образуют базис U, а строки матрицы A линейно независимы. Определим матрицу $B = (u_1|\ldots|u_k) \in \mathrm{M}_{n\,k}(\mathbb{R})$. Тогда утверждаются следующие вещи:

1. Количество столбцов B совпадает с количеством строк A, то есть $k = s$.
2. Матрица AB обратима.
3. Оператор проекции задается формулой $P = B(AB)^{-1}A$. Мнемоническое правило «БАБА».

Доказательство. Матрица A задает линейное отображение $A\colon \mathbb{R}^n \to \mathbb{R}^s$ такое, что $\ker A = W$ и $\operatorname{Im} A = \mathbb{R}^s$ (так как строки матрицы A линейно независимы, то $\operatorname{rk} A = s$, но $\operatorname{rk} A = \dim \operatorname{Im} A$). Матрица B задает отображение $B\colon \mathbb{R}^k \to \mathbb{R}^n$ такое, что $\operatorname{Im} B = U$ и $\ker B = 0$ (так как столбцы B линейно независимы).

(1) Мы знаем, что

$$\begin{array}{c} \dim U + \dim W = n \\ \dim \ker A + \dim \operatorname{Im} A = n \end{array} \quad \text{то есть} \quad \begin{array}{c} k + \dim W = n \\ \dim W + s = n \end{array} \quad \text{откуда} \quad s = k$$

(2) Теперь рассмотрим отображение $AB\colon \mathbb{R}^k \to \mathbb{R}^k$. Заметим, что $\operatorname{Im} B \cap \ker A = U \cap W = 0$. Значит $\ker AB = 0$, то есть AB – обратимый оператор.

(3) Теперь выведем формулу для P. Пусть $v = u + w$, где $v \in \mathbb{R}^n$ – произвольный вектор, $u \in U$ и $w \in W$ – его единственное разложение по прямой сумме подпространств. Тогда $Av = Au + Aw = Au$. С другой стороны, так как $u \in U$, мы имеем $u = Bx$ для некоторого $x \in \mathbb{R}^k$. Тогда $Av = ABx$. Так как AB обратимая квадратная матрица, имеем $x = (AB)^{-1}Av$. Значит $u = Bx = B(AB)^{-1}Av$, что и требовалось. □

Обратите внимание, что проектор P на U вдоль W зависит от двух подпространств, а не только от U. Если вы измените одно из них, то проектор изменится.

Формула Атата

Теперь я хочу разобрать случай проектора на подпространство вдоль его ортогонального дополнения. Такой проектор называется ортопроектором. Пусть $V = \mathbb{R}^n$ со стандартным скалярным произведением $(x, y) = x^t y$ и пусть подпространство $U \subseteq V$ задано своим базисом $U = \langle u_1, \ldots, u_k \rangle$. Составим матрицу $A = (u_1|\ldots|u_k) \in \mathrm{M}_{n\,k}(\mathbb{R})$. Тогда $U^\perp = \{y \in \mathbb{R}^n \mid A^t y = 0\}$. Пусть теперь $v \in V$ – произвольный вектор и $v = \operatorname{pr}_U v + \operatorname{ort}_U v$. Тогда формула «БАБА» превращается в $\operatorname{pr}_U v = A(A^tA)^{-1}A^t v$. Мнемоническое правило для запоминания: в евклидовом пространстве БАБА – это Атата.

Обратите внимание, что проектор P всегда зависит от двух подпространств: то, на которое проектируем U, и то, вдоль которого проектируем W. Но в случае ортогонального проектирования $W = U^\perp$, потому ортопроектор P реально зависит только от одного подпространства.

Метод наименьших квадратов

Пусть мы хотим решить систему $Ax = b$, где $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$ – столбец неизвестных. И предположим, что система не имеет решений, но от этого наше желание ее решить не становится слабее. Давайте обсудим, как удовлетворить наши желания в подобной ситуации и когда такие ситуации обычно встречаются.

Введем на пространстве $\mathbb{R}^m$ стандартное скалярное произведение $(x, y) = x^t y$. Тогда, на процесс решения системы можно смотреть так: мы подбираем $x \in \mathbb{R}^n$ так, чтоб $|Ax - b| = 0$. Если решить систему невозможно, то этот подход подсказывает, как надо поступить. Надо пытаться минимизировать расстояние между Ax и b. То есть решить задачу

$$\begin{aligned} &|Ax - b| \to \min \\ &x \in \mathbb{R}^n \end{aligned}$$

Теперь давайте поймем, как надо решать такую задачу. Пусть матрица A имеет вид $A = (A_1|\ldots|A_n)$, где $A_i \in \mathbb{R}^m$ – ее столбцы. Тогда система $Ax = b$ означает, $x_1A_1 + \ldots + x_nA_n = b$. То есть система разрешима тогда и только тогда, когда $b \in \langle A_1, \ldots, A_n \rangle$. Значит наша задача минимизировать расстояние между b и $\langle A_1, \ldots, A_n \rangle$. Мы можем разложить вектор b на проекцию и ортогональную составляющую относительно $\langle A_1, \ldots, A_n \rangle$. Обычная теорема пифагора говорит, что минимум расстояния достигается на $b_0 = \operatorname{pr}_{\langle A \rangle} b$. В этом случае вместо исходной системы $Ax = b$ мы должны решить систему $Ax = b_0$. И если x_0 – ее решение, то $|Ax_0 - b|$ как раз и будет минимальным.

Давайте теперь предположим, что столбцы матрицы A линейно независимы. Тогда по формуле «Атата» мы знаем, что $b_0 = A(A^tA)^{-1}A^t b$. Кроме этого должно выполняться $b_0 = Ax_0$. Так как столбцы A линейно независимы, такое x_0 должно быть единственным. Но мы видим, что в качестве x_0 подходит $x_0 = (A^tA)^{-1}A^t b$.

Двойственность для подпространств

Утверждение. *Пусть $V = \mathbb{R}^n$ – евклидово пространство. Тогда:*

1. *Для любого подпространства $W \subseteq V$ выполнено*

$$\dim W^\perp + \dim W = \dim V$$

2. *Для любого подпространства $W \subseteq V$, $V = W \oplus W^{\perp}$.*

3. *Для любого подпространства $W \subseteq V$ выполнено $(W^{\perp})^{\perp} = W$.*

4. *Для любых подпространств $W \subseteq E \subseteq V$ верно, что $W^{\perp} \supseteq E^{\perp}$. Причем $W = E$ тогда и только тогда, когда $W^{\perp} = E^{\perp}$.*

5. *Для любых подпространств $W, E \subseteq V$ выполнено равенство*
$$(W + E)^{\perp} = W^{\perp} \cap E^{\perp}$$

6. *Для любых подпространств $W, E \subseteq V$ выполнено равенство*
$$(W \cap E)^{\perp} = W^{\perp} + E^{\perp}$$

Самосопряженные операторы

Пусть V – евклидово пространство и пусть $\phi\colon V \to V$ – линейный оператор. Тогда *сопряженный* к нему линейный оператор ϕ^* – это такой оператор, что $(\phi(v), u) = (v, \phi^*(u))$ для всех $v, u \in V$. Оператор называется *самосопряженным*, если $\phi^* = \phi$.

Теперь разберемся, что происходит в ортонормированном базисе. В этом случае $V = \mathbb{R}^n$, $(x, y) = x^t y$, а $\phi(x) = Ax$, а $\phi^*(x) = Bx$. Тогда условие $(Ax, y) = (x, By)$ означает $x^t A^t y = x^t By$. То есть $B = A^t$. То есть матрица для ϕ^* это A^t. Значит самосопряженный оператор в ортонормированном базисе задается симметричной матрицей.

В случае произвольного базиса скалярное произведение задается $(x, y) = x^t By$, где B – симметричная невырожденная положительно определенная матрица. Тогда если $\phi x = Ax$ и $\phi^* x = A'x$, то условие $(Ax, y) = (x, A'y)$ расписывается так: $(Ax)^t By = x^t BA'y$. То есть $x^t A^t By = x^t BA'y$ для всех $x, y \in \mathbb{R}^n$. Последнее значит, что $A^t B = BA'$. Значит $A' = B^{-1} A^t B$ – это формула связывает матрицу ϕ и ϕ^* в произвольных базисах.

Утверждение. *Пусть $\phi\colon V \to V$ – самосопряженный оператор в евклидовом пространстве. Тогда*

1. *Все его собственные значения вещественны.*

2. *Собственные вектора с разными собственными значениями ортогональны друг другу.*

3. *Существует ортонормированный базис пространства V состоящий из собственных векторов ϕ.*

4. *В некотором ортонормированном базисе матрица ϕ имеет диагональный вид, с вещественными числами на диагонали.*

Переформулируем это утверждение на языке матриц.

Утверждение. *Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – симметрическая матрица. Тогда*

1. *Все собственные значения A вещественные.*

2. *Все собственные вектора с разными собственными значениями ортогональны.*

3. *Существует ортогональная матрица $C \in \mathrm{M}_n(\mathbb{R})$ такая, что $C^{-1}AC$ является диагональной вещественной матрицей.*[1]

Самосопряженный оператор называется *положительным*, если все его собственные значения **неотрицательные**. Да, да, именно так. Нулевая матрица тоже считается положительным оператором. Вот такая вот дурацкая терминология.

Алгоритм разложения симметрических матриц

Дано Матрица $A \in \mathrm{M}_n(\mathbb{R})$ такая, что $A^t = A$.

[1]Обратите внимание, чтот тут нет разницы между $C^{-1}AC$ и $C^t AC$, так как C ортогональная.

Задача Найти разложение $A = C\Lambda C^t$, где $C \in \mathrm{M}_n(\mathbb{R})$ – ортогональная матрица, $\Lambda \in \mathrm{M}_n(\mathbb{R})$ – диагональная матрица.

Алгоритм

1. Найти собственные значения матрицы A.
   (a) Составить характеристический многочлен $\chi(\lambda) = \det(A - \lambda E)$.
   (b) Найти корни $\chi(\lambda)$ с учетом кратностей: $\{(\lambda_1, n_1), \ldots, (\lambda_k, n_k)\}$, где λ_i – корни, n_i – кратности.
2. Для каждого λ_i найти ортонормированный базис в пространстве собственных векторов отвечающему λ_i.
   (a) Найти ФСР системы $(A - \lambda_i E)x = 0$. Пусть это будет $v_1^i, \ldots, v_{n_i}^i$. Обратите внимание, что их количество будет в точности равно кратности n_i.
   (b) Ортогонализовать $v_1^i, \ldots, v_{n_i}^i$ методом Грама-Шмидта. Обратите внимание, после ортогонализации останется ровно n_i векторов.
   (c) Сделать каждый вектор длинны один: $v_j^i \mapsto \frac{v_j^i}{|v_j^i|}$.
3. Матрица Λ будет диагональной с числами $\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_k$ на диагонали, где каждое λ_i повторяется n_i раз. Обратите внимание, всего получится n чисел.
4. Матрица C будет составлена из столбцов $v_1^1, \ldots, v_{n_1}^1, v_1^2, \ldots, v_{n_2}^2, \ldots, v_1^k, \ldots, v_{n_k}^k$. Обратите внимание, порядок собственных векторов соответствует порядку собственных значений в матрице Λ.

Сингулярное разложение (SVD)

Это утверждение я сформулирую только на матричном языке.

Утверждение. *Пусть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$. Тогда существуют*

1. *ортогональная матрица $C \in \mathrm{M}_m(\mathbb{R})$.*
2. *ортогональная матрица $D \in \mathrm{M}_n(\mathbb{R})$.*
3. *положительные вещественные числа $\lambda_1 \geqslant \ldots \geqslant \lambda_s > 0$*

такие, что $A = C\Lambda D^t$, где

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_s & \\ & & & 0 \end{pmatrix} \in \mathrm{M}_{m\,n}(\mathbb{R})$$

Пусть столбцы матрицы C – это вектора v_i, а столбцы матрицы D – это вектора u_i. Тогда утверждение означает, что

$$A = \lambda_1 v_i u_i^t + \ldots + \lambda_s v_s u_s^t$$

То есть мы представили матрицу A в виде «ортогональной» суммы матриц ранга один, в том смысле, что все v_i ортогональны друг другу и все u_i ортогональны друг другу.

Усеченное сингулярное разложение Пусть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $C \in \mathrm{M}_{m\,s}(\mathbb{R})$, $D \in \mathrm{M}_{n\,s}(\mathbb{R})$ и $\Lambda \in \mathrm{M}_s(\mathbb{R})$ – диагональная матрица с числами $\lambda_1 \geqslant \ldots \geqslant \lambda_s > 0$ на диагонали. Предположим, что столбцы матриц C и D ортонормированны (то есть все между собой ортогональны и длины один). Тогда равенство вида $A = C\Lambda D^t$ называется усеченным сингулярным разложением.

Если нам известно сингулярное разложение $A = C\Lambda D^t$, то усеченное из него делается так: 1) составим матрицу C', состоящую из первых s столбцов матрицы C, 2) составим матрицу D', состоящую из первых s столбцов матрицы D, 3) определим матрицу Λ' как квадратную s на s матрицу с диагональю из матрицы Λ. Тогда $A = C'\Lambda'(D')^t$ будет усеченным разложением.

Замечание Философский смысл этого разложения следующий. Пусть наша матрица – это квадратная черно-белая картинка, где числа – интенсивности черного цвета. На вектора v_i и u_i надо смотреть как на «ортогональные» компоненты «базовых» цветовых интенсивностей. А λ_i – это мощности этих самых сигналов. Потому, если λ_i достаточно малы, то наш глаз не способен различить соответствующие сигналы. Потому, если мы выкинем их из нашей матрицы, то на глаз, матрица A не будет отличаться от полученной.

Обычно в реальной жизни выходит, что достаточно только первых штук пять слагаемых. Тогда $A' = \lambda_1 v_1 u_1^t + \ldots + \lambda_5 v_5 u_5^t$ будет на глаз не отличима от A. В чем же польза от такого? На хранение матрицы A нам потребуется mn чисел. Для хранения матрицы A' нам надо 5 чисел λ_i и еще 5 пар векторов v_i и u_i, на хранение каждого из которых надо m и n чисел соответственно. И того затраты $5 + 5m + 5n = 5(m + n + 1)$. Это дает огромный выигрыш в количестве хранимой информации и является основой для многих алгоритмов архивации с потерей данных вроде JPG.

Алгоритм нахождения сингулярного разложения

Дано Матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$.

Задача Найти разложение $A = C\Lambda D^t$, где $C \in \mathrm{M}_m(\mathbb{R})$ ортогональная, $D \in \mathrm{M}_n(\mathbb{R})$ ортогональная, $\Lambda \in \mathrm{M}_{m\,n}(\mathbb{R})$ содержит на диагонали элементы $\lambda_1 \geqslant \ldots \geqslant \lambda_s > 0$, а все остальные нули.

Алгоритм

1. Составим матрицу $S = AA^t \in \mathrm{M}_m(\mathbb{R})$. Тогда $S = C\Lambda\Lambda^t C^t$.

2. Так как $S^t = S$. То с помощью алгоритма для симметрических матриц найдем ее разложение $S = \bar{C}\bar{\Lambda}\bar{C}^t$. Причем, обязательно получится, что диагональная матрица $\bar{\Lambda}$ состоит из неотрицательных элементов.

3. Тогда $C = \bar{C}$, а $\Lambda\Lambda^t = \bar{\Lambda}$. То есть $\lambda_i^2 = \bar{\lambda}_i$. Так как $\lambda_i > 0$, то они находятся как $\lambda_i = \sqrt{\bar{\lambda}_i}$.

4. Теперь надо найти D из условия $A = C\Lambda D^t$. Обратите внимание Λ не обязательно квадратная и не обязательно обратимая. Заметим
$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_s & \\ & & & 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_s & \\ & & & E \end{pmatrix} \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{pmatrix}$$
Значит
$$\begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_s & \\ & & & E \end{pmatrix}^{-1} C^t A = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{pmatrix} D^t$$
Представим D в виде $D = (D_1|D_2)$, где D_1 – состоит из первых s столбцов. Тогда
$$\begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_s & \\ & & & E \end{pmatrix}^{-1} C^t A = \begin{pmatrix} D_1^t \\ 0 \end{pmatrix}$$
Отсюда нашлась D_1.

5. Матрица D_2 однозначно не восстанавливается. Чтобы найти какую-то D, надо дополнить столбцы D_1 до базиса в $\mathbb{R}^n$ и ортонормировать полученную систему. Тогда это даст все столбцы матрицы D. Обратите внимание, первые s столбцов должны остаться столбцами из D_1.

Замечания

1. Надо заметить, что нельзя попытаться составить матрицу A^tA и из нее найти матрицу D. Так как матрицы D и C определены не однозначно и зависят друг от друга. Если вы нашли какую-то матрицу C, то к ней подойдет не любая найденная матрица D, а только та, что является решением $A = C\Lambda D^t$.

2. Приведенным выше алгоритмом имеет смысл пользоваться, если у матрицы A количество строк меньше, чем количество столбцов. Если же столбцов меньше, чем строк, то надо найти сингулярное разложение для $A^t = C\Lambda D^t$. Тогда $A = D\Lambda^t C^t$ будет искомым сингулярным разложением для A.

Алгоритм нахождения усеченного сингулярного разложения

Дано Матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$.

Задача Найти разложение $A = C\Lambda D^t$, где $C \in \mathrm{M}_{m\,s}(\mathbb{R})$, $D \in \mathrm{M}_{n\,s}(\mathbb{R})$ – матрицы с ортонормированными столбцами, $\Lambda \in \mathrm{M}_s(\mathbb{R})$ – диагональная матрица с элементами $\lambda_1 \geqslant \ldots \geqslant \lambda_s > 0$ на диагонали.

Алгоритм

1. Составим матрицу $S = AA^t \in \mathrm{M}_m(\mathbb{R})$. Тогда $S = C\Lambda^2 C^t$.

2. Надо найти ненулевые собственные значения матрицы (с учетом кратности) S и упорядочить их по невозрастанию:

   (a) Составить характеристический многочлен $\chi(\lambda) = \det(S - \lambda E)$.

   (b) Найти корни $\chi(\sigma)$ с учетом кратностей: $\{(\sigma_1, n_1), \ldots, (\sigma_k, n_k)\}$, где σ_i – ненулевые корни, n_i – кратности. Мы считаем, что $\sigma_1 \geqslant \ldots \geqslant \sigma_k > 0$.

3. Для каждого σ_i найти ортонормированный базис в пространстве собственных векторов отвечающему σ_i.

   (a) Найти ФСР системы $(S - \sigma_i E)x = 0$. Пусть это будет $v_1^i, \ldots, v_{n_i}^i$. Обратите внимание, что их количество будет в точности равно кратности n_i.

   (b) Ортогонализовать $v_1^i, \ldots, v_{n_i}^i$ методом Грама-Шмидта. Обратите внимание, после ортогонализации останется ровно n_i векторов.

   (c) Сделать каждый вектор длинны один: $v_j^i \mapsto \frac{v_j^i}{|v_j^i|}$.

4. Составим все векторы v_j^i в столбцы матрицы C так, что сначала идут векторы для σ_1, потом для σ_2 и тд.

5. Положим $\lambda_1 = \sqrt{\sigma_1}, \ldots, \lambda_s = \sqrt{\sigma_s}$. Определим $\Lambda \in \mathrm{M}_s(\mathbb{R})$ как диагональную матрицу с числами $\lambda_1, \ldots, \lambda_s$ на диагонали.

6. Найдем матрицу D решая уравнение $A = C\Lambda D^t$. Так как столбцы C ортогональны, то получим $D^t = \Lambda^{-1} C^t A$.