МЕТОД ГЛАВНЫХ КОМПОНЕНТ

§ 1. Геометрия главных компонент

После того, как объекты разбиты на однородные группы (классы), возникает задача изучения *взаимосвязей признаков* внутри отдельного класса. На практике чаще всего встречаются следующие два вида зависимостей: а) объекты образуют «облако» эллиптического типа (рис. 1, *а*), б) объекты располагаются в окрестности некоторой кривой (поверхности) (рис. 1, *б*).

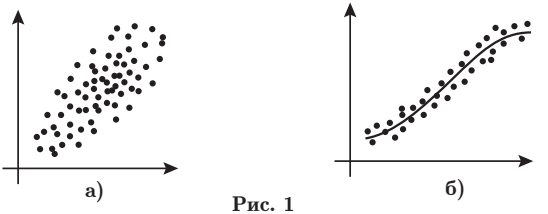

Рис. 1

В случае а) оба признака являются «полноценными» случайными величинами, и изучению подлежит уровень зависимости (корреляции) между ними. Случай б) соответствует «функциональной» зависимости между признаками,испорченной шумом. Зависимости первого вида изучаются в этой главе методами *корреляционного анализа.* Методы, позволяющие во втором случае построить интересующую исследователя кривую (поверхность), относятся к так называемому *регрессионному анализу.*

Допустим, что каждый из n объектов описывается m признаками (координатами), и представим данные (для отдельного класса объектов) в форме таблицы $\boldsymbol{X} = \|x_{il}\|_{n\times m}$. Вычислим для каждого признака (столбца матрицы $\boldsymbol{X}$) *среднее значение* $\overline{x}_l = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_{il}$ и центрируем данные: $x'_{il} = x_{il} - \overline{x}_l$. Далее будем считать x_{il} уже *центрированными*: $\overline{x}_l = 0$ для $l = 1, \ldots, m$. (Д1)

Обозначим через $\widehat{\boldsymbol{\Sigma}} = \|\widehat{\sigma}_{kl}\|_{m\times m}$ *выборочную ковариационную матрицу* (центрированных) признаков: $\widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n}\,\boldsymbol{X}^T\boldsymbol{X}$ (т. е. $\widehat{\sigma}_{kl}$ — выборочная ковариация k-го и l-го столбцов матрицы $\boldsymbol{X}$).

Поскольку $\widehat{\boldsymbol{\Sigma}}$ — матрица ковариаций, она неотрицательно определена. Следовательно, существует ортогональная матрица $\boldsymbol{C}$, приводящая $\widehat{\boldsymbol{\Sigma}}$ к главным осям: $\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}\boldsymbol{C} = \boldsymbol{\Lambda}$.

Здесь $\boldsymbol{\Lambda}$ — диагональная матрица с неотрицательными элементами $\lambda_1 \geqslant \ldots \geqslant \lambda_m$ на главной диагонали, которые являются корнями уравнения $\det(\widehat{\boldsymbol{\Sigma}} - \lambda\boldsymbol{E}) = 0$. Они называются *собственными значениями* матрицы $\widehat{\boldsymbol{\Sigma}}$. Предположим, что все λ_l *положительны*

и различны (для экспериментальных данных x_{il} это условие выполняется практически всегда). При этом столбцы $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m$ матрицы $\boldsymbol{C}$ (*главные оси или компоненты*) определяются однозначно с точностью до выбора направления оси (одновременного изменения знака всех координат вектора $\boldsymbol{c}_l$). Они образуют новый ортонормированный базис в $\mathbb{R}^m$, обладающий **рядом важных свойств**:

1) Проекции объектов на первую главную компоненту $\boldsymbol{c}_1$ имеют *наибольшую* выборочную дисперсию среди проекций на всевозможные направления $\boldsymbol{d}$ в пространстве $\mathbb{R}^m$.

   Доказательство. Вектор проекций $\boldsymbol{y}$ на направление $\boldsymbol{d}$ ($\boldsymbol{d}^T\boldsymbol{d}=1$) задается равенством $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{d}$. Ввиду допущения Д1

   $$\overline{\boldsymbol{y}}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{m}x_{il}d_l=\sum_{l=1}^{m}\overline{x}_l d_l=0.$$

   Тогда *выборочная дисперсия проекций на направление* $\boldsymbol{d}$ равна

   $$S^2(\boldsymbol{d})=\frac{1}{n}\sum_{i=1}^{n}y_i^2=\frac{1}{n}\,\boldsymbol{y}^T\boldsymbol{y}=\frac{1}{n}\,(\boldsymbol{X}\boldsymbol{d})^T\boldsymbol{X}\boldsymbol{d}=\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}. \qquad (1)$$

   Тем самым задача сводится к максимизации по $\boldsymbol{d}$ квадратичной формы $\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}$ при условии $\boldsymbol{d}^T\boldsymbol{d}=1$. Для ее решения применим метод неопределенных множителей Лагранжа.
   Приравнивая нулю частные производные по переменным d_l *функции Лагранжа*

   $$F(\boldsymbol{d},\lambda)=\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}-\lambda(\boldsymbol{d}^T\boldsymbol{d}-1)=\sum_{k=1}^{m}\sum_{l=1}^{m}\widehat{\sigma}_{kl}d_k d_l-\lambda\sum_{l=1}^{m}d_l^2+\lambda,$$

   приходим к системе (линейных относительно $d_1,\ldots,d_m$) уравнений

   $$\frac{\partial}{\partial d_l}F(\boldsymbol{d},\lambda)=2\sum_{k=1}^{m}\widehat{\sigma}_{kl}d_k-2\lambda d_l=0,\quad l=1,\ldots,m.$$

   Ее можно записать в матричной форме:

   $$\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}-\lambda\boldsymbol{d}=\boldsymbol{0}\Longleftrightarrow(\widehat{\boldsymbol{\Sigma}}-\lambda\boldsymbol{E})\boldsymbol{d}=\boldsymbol{0}. \qquad (2)$$

   Поскольку $\boldsymbol{d}^T\boldsymbol{d}=1$, нас интересуют только ненулевые решения. Для них должно выполняться условие $\det(\widehat{\boldsymbol{\Sigma}}-\lambda\boldsymbol{E})=0$, т. е. искомое направление $\boldsymbol{d}$ обязано быть собственным вектором $\widehat{\boldsymbol{\Sigma}}$, отвечающим собственному значению λ. Умножая соотношение (2) на $\boldsymbol{d}^T$ слева, получим

   $$\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}=\lambda\,\boldsymbol{d}^T\boldsymbol{d}=\lambda.$$

   Левая часть есть $S^2(\boldsymbol{d})$ (см. формулу (1)). Следовательно, λ должно равняться наибольшему собственному значению λ_1 матрицы $\widehat{\boldsymbol{\Sigma}}$, а $\boldsymbol{d}=\boldsymbol{c}_1$. ■

2) Второй собственный вектор $\boldsymbol{c}_2$ характеризуется тем, что выборочная дисперсия проекций объектов на ось $\boldsymbol{c}_2$ максимальна среди всех направлений $\boldsymbol{d}$, *ортогональных вектору* $\boldsymbol{c}_1$, т. е. таких, что $\boldsymbol{d}^T\boldsymbol{c}_1 = 0$.

   Доказательство. Функция Лагранжа в этом случае выглядит так:

   $$F(\boldsymbol{d},\lambda,\mu) = \boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d} - \lambda(\boldsymbol{d}^T\boldsymbol{d} - 1) - \mu(\boldsymbol{d}^T\boldsymbol{c}_1 - 0).$$

   Вычисляя ее частные производные по d_l, приходим к системе

   $$\widehat{\boldsymbol{\Sigma}}\boldsymbol{d} - \lambda\boldsymbol{d} - \mu\boldsymbol{c}_1 = \boldsymbol{0}. \tag{3}$$

   Транспонируем равенство (3), умножим на $\boldsymbol{c}_1$ справа и воспользуемся тем, что $\boldsymbol{c}_1$ — собственный вектор с собственным значением λ_1 (т. е. $\widehat{\boldsymbol{\Sigma}}\boldsymbol{c}_1 = \lambda_1\boldsymbol{c}_1$), а также условием $\boldsymbol{d}^T\boldsymbol{c}_1 = 0$:

   $$\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{c}_1 - \lambda\,\boldsymbol{d}^T\boldsymbol{c}_1 - \mu\,\boldsymbol{c}_1^T\boldsymbol{c}_1 = 0 \iff \lambda_1\cdot 0 - \lambda\cdot 0 - \mu\cdot 1 = 0.$$

   Отсюда $\mu = 0$. С учетом этого из (3) вытекает, что $\boldsymbol{d}$ — собственный вектор матрицы $\widehat{\boldsymbol{\Sigma}}$ с собственным значением λ. Умножая равенство (3) слева на $\boldsymbol{d}^T$, выводим, что $\lambda = \lambda_2$ и $\boldsymbol{d} = \boldsymbol{c}_2$. ■

3) При $l \geqslant 3$ аналогично устанавливается, что $\boldsymbol{c}_l$ — направление с наибольшей выборочной дисперсией проекций объектов среди направлений, ортогональных векторам $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{l-1}$.

4) Сумма выборочных дисперсий исходных признаков (столбцов матрицы $\boldsymbol{X}$) $\widehat{\sigma}_{11} + \ldots + \widehat{\sigma}_{mm} = \operatorname{tr}\widehat{\boldsymbol{\Sigma}}$ в силу подобия матриц $\widehat{\boldsymbol{\Sigma}}$ и $\boldsymbol{\Lambda}$ (см. П10) равна $\operatorname{tr}\boldsymbol{\Lambda} = \lambda_1 + \ldots + \lambda_m = S^2(\boldsymbol{c}_1) + \ldots + S^2(\boldsymbol{c}_m)$, т. е. сумме выборочных дисперсий проекций объектов на (новые) главные оси. Эта величина может рассматриваться как *мера общего разброса* объектов относительно их центра масс. Представляет интерес *относительная доля разброса, приходящаяся на k первых главных осей*,

   $$\gamma_k = (\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_m), \quad k \leqslant m.$$

   Если эта величина при некотором k достаточно близка к 1, то возможно *уменьшение размерности* пространства признаков за счет перехода от m исходных признаков к k новым признакам — первым главным компонентам. На практике нередко удается ограничиться двумя или тремя компонентами без существенной потери информации. Объекты описываются координатами в новых осях, которым специалисты-прикладники, как правило, могут придать содержательную интерпретацию.

	1	2	3	4	5	6
1	129	64	95	17,5	11,2	13,8
2	154	74	76	20,0	14,2	16,5
3	170	87	71	17,9	12,3	15,9
4	188	94	73	19,5	13,3	14,8
5	161	81	55	17,1	12,1	13,0
6	164	90	58	17,5	12,7	14,7
7	203	109	65	20,7	14,0	16,8
8	178	97	57	17,3	12,8	14,3
9	212	114	65	20,5	14,3	15,5
10	221	123	62	21,2	15,2	17,0
11	183	97	52	19,3	12,9	13,5
12	212	112	65	19,7	14,2	16,0
13	220	117	70	19,8	14,3	15,6
14	216	113	72	20,5	14,4	17,7
15	216	112	75	19,6	14,0	16,4

	1	2	3	4	5	6
16	205	110	68	20,8	14,1	16,4
17	228	122	78	22,5	14,2	17,8
18	218	112	65	20,3	13,9	17,0
19	190	93	78	19,7	13,2	14,0
20	212	111	73	20,5	13,7	16,6
21	201	105	70	19,8	14,3	15,9
22	196	106	67	18,5	12,6	14,2
23	158	71	71	16,7	12,5	13,3
24	255	126	86	21,4	15,0	18,0
25	234	113	83	21,3	14,8	17,0
26	205	105	70	19,0	12,4	14,9
27	186	97	62	19,0	13,2	14,2
28	241	119	87	21,0	14,7	18,3
29	220	111	88	22,5	15,4	18,0
30	242	120	85	19,9	15,3	17,6

	1	2	3	4	5	6
31	199	105	73	23,4	15,0	19,1
32	227	117	77	25,0	15,3	18,6
33	228	122	82	24,7	15,0	18,5
34	232	123	83	25,3	16,8	15,5
35	231	121	78	23,5	16,5	19,6
36	215	118	74	25,7	15,7	19,0
37	184	100	69	23,3	15,8	19,7
38	175	94	73	22,2	14,8	17,0
39	239	124	77	25,0	16,8	27,0
40	203	109	70	23,3	15,0	18,7
41	226	118	72	26,0	16,0	19,4
42	226	119	77	26,5	16,8	19,3
43	210	103	72	20,5	14,0	16,7

Рис. 2

Пример 1.
В таблице на рис. 2 указаны размеры челюстей и зубов тридцати собак (номера 1–30), двенадцати волков (номера 31–42) и ископаемого черепа неизвестного животного (номер 43), найденного в четвертичном слое.

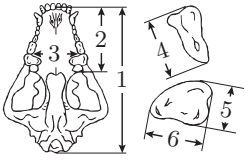

Рис. 3

Рис. 3 демонстрирует измеряемые характеристики: 1 — длина черепа, 2 — длина верхней челюсти, 3 — ширина верхней челюсти; следующие измерения относятся к зубам: 4 — длина верхнего карнивора, 5 — длина первого верхнего моляра, 6 — ширина первого верхнего моляра.

Найдем и интерпретируем главные компоненты для этих данных. Исходные признаки следующие: 1 — длина черепа, 2 — длина верхней челюсти, 3 — ширина верхней челюсти, 4 — длина верхнего карнивора, 5 — длина первого верхнего моляра, 6 — ширина первого верхнего моляра.

Используем нормировку одномерных наблюдений $Z_1, \ldots, Z_n$ вида:

$Z'_i = (Z_i - \overline{Z})/S$, где $\overline{Z} = \frac{1}{n}\sum Z_i$ — среднее арифметическое,

$S^2 = \frac{1}{n}\sum (Z_i - \overline{Z})^2$ — выборочная дисперсия.

	1	2	3	4	5	6
1	1	0,96	0,35	0,61	0,72	0,59
2		1	0,20	0,66	0,74	0,59
3			1	0,37	0,35	0,35
4				1	0,89	0,76
5					1	0,79
6						1

Рис. 4

$\boldsymbol{c}_1$	$\boldsymbol{c}_2$	$\boldsymbol{c}_3$	$\boldsymbol{c}_4$	$\boldsymbol{c}_5$	$\boldsymbol{c}_6$
0,43	0,23	0,53	0,11	0,05	0,68
0,43	0,38	0,39	0,01	−0,20	−0,69
0,23	−0,89	0,38	−0,02	0,00	−0,13
0,44	−0,07	−0,40	−0,52	−0,58	0,18
0,46	0,02	−0,27	−0,31	0,78	−0,09
0,42	−0,10	−0,44	0,79	−0,09	−0,01

Рис. 5

Очевидно, что при нормировке данных с помощью средних арифметических и стандартных отклонений признаков выборочная ковариационная матрица $\widehat{\boldsymbol{\Sigma}}$ совпадает с выборочной корреляционной матрицей исходных признаков. Ее нетрудно подсчитать. Результаты вычислений представлены таблицей на рис. 4.

Обратим внимание, что длина черепа (признак 1) и длина верхней челюсти (признак 2) сильно коррелируют ($\widehat{\rho}_{12} = 0{,}96$), поэтому целесообразно оставить в модели только один из них. Признаки 4, 5 и 6, относящиеся к зубам, также очень тесно связаны между собой, поскольку $\widehat{\rho}_{45} = 0{,}89$, $\widehat{\rho}_{46} = 0{,}76$ и $\widehat{\rho}_{56} = 0{,}79$.

На рис. 5 приведены собственные векторы $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_6$, вычисленные *степенным методом* (см. § 3). Собственные значения λ_k, соответствующие им доли следа $\lambda_k / \sum \lambda_l$ (в %) и накопленные доли γ_k (в %) указаны в следующей таблице:

Номера компонент	1	2	3	4	5	6
Собственные значения	4,100	0,883	0,639	0,259	0,097	0,022
Проценты от следа	68,3	14,7	10,7	4,3	1,6	0,4
Накопленные проценты	68,3	83,0	93,7	98,0	99,6	100,0

На первые 3 компоненты приходится 93,7% полной дисперсии «облака». При этом первая компонента имеет смысл *общего размера*. Это следует из того, что все координаты у $\boldsymbol{c}_1$ одного знака и примерно одинаковы по величине, т. е. при проецировании на эту ось координаты нормированных признаков просто складываются.

Вторая компонента, по существу, отвечает за *ширину верхней челюсти* (признак 3), поскольку третья координата у $\boldsymbol{c}_2$ по абсолютной величине равна $0{,}89 \approx 1$. Эта ось отражает различие в пропорциях челюстей и отличает удлиненные формы от укороченных (гончих и колли от бульдогов и боксеров). На вторую ось волки проецируются в основном рядом с немецкими овчарками.

Третья ось противопоставляет размеры челюстей размерам зубов: первые три координаты у $\boldsymbol{c}_3$ примерно равны по абсолютной величине последним трем координатам, но противоположны по знаку. Другими словами, третья ось отражает *относительную* (по сравнению с размерами черепа) *величину зубов*. Она позволяет отличить животных с развитыми зубами (волки, немецкие овчарки, доберманы) от собак других пород (сенбернары, мастифы, сеттеры).

5) Пусть M_k — это подпространство, натянутое на главные оси $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$. Оказывается, при проецировании объектов на произвольное подпространство L_k размерности k в $\mathbb{R}^m$ геометрическая структура *искажается в наименьшей степени*, если этим подпространством является M_k :
а) сумма квадратов расстояний от объектов до их проекций на L_k минимальна, когда $L_k = M_k$ (в этом случае она равна $n(\lambda_{k+1} + \ldots + \lambda_m)$;
б) при проецировании на M_k наименее искажается сумма квадратов расстояний между всевозможными парами объектов (для M_k ее изменение составляет $n^2(\lambda_{k+1} + \ldots + \lambda_m)$);
в) когда $L_k = M_k$, в наименьшей степени искажаются расстояния от объектов до их центра масс (совпадающего с началом координат $\mathbf{0}$ ввиду допущения Д1), а также углы между всевозможными парами прямых, соединяющих объекты с $\mathbf{0}$.

Замечание 1. Следует иметь в виду, что главные компоненты вычисляются по выборочной ковариационной матрице $\widehat{\boldsymbol{\Sigma}}$ и поэтому *зависят от масштаба признаков*. Скажем, если один из признаков принимает значения от 0 до 100, а другие — от 0 до 10, то независимо от структуры данных первый признак будет отождествляться с первой главной компонентой. Чтобы избежать этого, обычно перед вычислением главных компонент данные нормируют.

§2. Вычисление главных компонент

Следующий простой алгоритм (*степенной метод*) позволяет вычислить для симметричной (не обязательно неотрицательно определенной) матрицы $\boldsymbol{A}$ несколько *наибольших по модулю* собственных значений λ_l и соответствующих собственных векторов $\boldsymbol{c}_l$.
Сначала разберем частный случай этого алгоритма — вычисление λ_1 и $\boldsymbol{c}_1$.

Возьмем произвольный ненулевой вектор $\boldsymbol{x}_0$ из $\mathbb{R}^m$. Пусть выполнены два условия: $|\lambda_1| > |\lambda_2| \geqslant \ldots \geqslant |\lambda_m|$ и $\boldsymbol{c}_1^T \boldsymbol{x}_0 \neq 0$. Положим $\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{x}_i$ (т. е. $\boldsymbol{x}_i = \boldsymbol{A}^i \boldsymbol{x}_0$, что объясняет название метода). Тогда последовательность значений $t_{i+1} = \boldsymbol{x}_i^T \boldsymbol{x}_{i+1}/|\boldsymbol{x}_i|^2$ сходится к λ_1 со скоростью геометрической прогрессии со знаменателем $\gamma = |\lambda_2/\lambda_1| < 1$.

Доказательство. Обозначим через $b_1, \ldots, b_m$ координаты вектора $\boldsymbol{x}_0$ в базисе $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$, т. е. $\boldsymbol{x}_0 = b_1\boldsymbol{c}_1 + \ldots + b_m\boldsymbol{c}_m$. Так как $\boldsymbol{A}\boldsymbol{c}_l = \lambda_l \boldsymbol{c}_l$, то $\boldsymbol{A}^i \boldsymbol{c}_l = \lambda_l^i \boldsymbol{c}_l$. Отсюда $\boldsymbol{x}_i = \lambda_1^i b_1 \boldsymbol{c}_1 + \ldots + \lambda_m^i b_m \boldsymbol{c}_m = \lambda_1^i (b_1 \boldsymbol{c}_1 + \boldsymbol{\delta}_i)$, где $\boldsymbol{\delta}_i = b_2(\lambda_2/\lambda_1)^i \boldsymbol{c}_2 + \ldots + b_m(\lambda_m/\lambda_1)^i \boldsymbol{c}_m$, причем в силу первого условия $|\boldsymbol{\delta}_i| = O(\gamma^i) \to 0$ при $i \to \infty$.

В частности, когда все $\lambda_l > 1$, это означает, что при каждой итерации длина вектора $\boldsymbol{x}_i$ увеличивается, но быстрее остальных растет его координата по оси $\boldsymbol{c}_1$. При этом направление $\boldsymbol{x}_i$ приближается к направлению первой главной оси (рис.6). ■

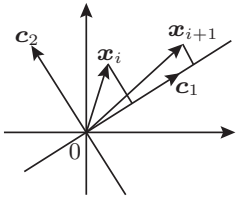

Рис. 6

Замечание 2. Если $|\lambda_1| > 1$, то $|\boldsymbol{x}_i| \to \infty$, а если $|\lambda_1| < 1$, то $|\boldsymbol{x}_i| \to 0$ при $i \to \infty$. При вычислении на компьютере в первом случае возможно переполнение, во втором — исчезновение порядка. Предотвратить эти ситуации позволяет нормировка векторов $\boldsymbol{x}_i$ на каждом шаге.

Учитывая замечание 2, приходим к следующему **алгоритму для вычисления λ_1 и $\boldsymbol{c}_1$.**

1. Положим $i = 0$, $t_0 = 0$ и зададим произвольный ненулевой вектор $\boldsymbol{x}_0$ из $\mathbb{R}^m$. Нормируем $\boldsymbol{x}_0$: пусть $\boldsymbol{e}_0 = \boldsymbol{x}_0/|\boldsymbol{x}_0|$ (обычно сразу берут $\boldsymbol{e}_0 = (1,0,\ldots,0)$).
2. Вычислим $\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{e}_i$ и $t_{i+1} = \boldsymbol{e}_i^T \boldsymbol{x}_{i+1}$.
3. Нормируем $\boldsymbol{x}_{i+1}$: $\boldsymbol{e}_{i+1} = \boldsymbol{x}_{i+1}/|\boldsymbol{x}_{i+1}|$.
4. Если $|t_{i+1} - t_i| \leqslant \varepsilon$, где ε — достаточно малое число, то положим $\lambda_1 = t_{i+1}$ и $\boldsymbol{c}_1 = \boldsymbol{e}_{i+1}$ и закончим итерационный процесс. Иначе увеличим i на 1 и перейдем к шагу 2.

Замечание 3. Если при выборе $\boldsymbol{x}_0$ случится, что $\boldsymbol{c}_1^T\boldsymbol{x}_0 = 0$, то через несколько итераций за счет погрешностей округления, как правило, появится ненулевая координата у $\boldsymbol{x}_i$ в направлении $\boldsymbol{c}_1$, и процесс, хотя и с некоторым запаздыванием, выйдет на первое собственное значение. Можно также стартовать с разных $\boldsymbol{x}_0$ и сравнивать результаты.

Пусть $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_k| > |\lambda_{k+1}| \geqslant \ldots \geqslant |\lambda_m|$. Для вычисления $k > 1$ первых главных компонент кроме умножения на $\boldsymbol{A}$ и нормировки понадобится еще ортогонализация. Именно, предположим, что λ_l и $\boldsymbol{c}_l$ уже известны при всех $l \leqslant k-1$. Чтобы найти λ_k и $\boldsymbol{c}_k$, выполним **следующие шаги**.

1. Положим $i = 0$, $t_0 = 0$ и зададим произвольный вектор $\boldsymbol{x}_0$ из $\mathbb{R}^m$ (надо только, чтобы $\boldsymbol{c}_k^T\boldsymbol{x}_0 \neq 0$).
2. Проведем ортогонализацию $\boldsymbol{x}_i$ к векторам $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_{k-1}$: $\boldsymbol{y}_i = \boldsymbol{x}_i - (\boldsymbol{c}_1^T\boldsymbol{x}_i)\,\boldsymbol{c}_1 - \ldots - (\boldsymbol{c}_{k-1}^T\boldsymbol{x}_i)\,\boldsymbol{c}_{k-1}$.
3. Нормируем $\boldsymbol{y}_i$: $\boldsymbol{e}_i = \boldsymbol{y}_i/|\boldsymbol{y}_i|$.
4. Вычислим $\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{e}_i$ и $t_{i+1} = \boldsymbol{e}_i^T\boldsymbol{x}_{i+1}$.
5. Нормируем $\boldsymbol{x}_{i+1}$: $\boldsymbol{z}_{i+1} = \boldsymbol{x}_{i+1}/|\boldsymbol{x}_{i+1}|$.
6. Если $|t_{i+1} - t_i| \leqslant \varepsilon$, то положим $\lambda_k = t_{i+1}$, $\boldsymbol{c}_k = \boldsymbol{z}_{i+1}$ и закончим процесс. В противном случае приравняем $\boldsymbol{x}_{i+1} = \boldsymbol{z}_{i+1}$, увеличим i на 1 и перейдем к шагу 2.

В заключение обсудим, насколько следует доверять вычисленным значениям λ_l и $\boldsymbol{c}_l$. Следующий пример показывает, что для числа объектов n порядка нескольких десятков разброс этих значений может оказаться *достаточно большим.*

Пример 2. ***Статистика главных компонент для нормальной модели.*** Предположим, что объекты представляют собой выборку размера n из закона $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ с невырожденной матрицей $\boldsymbol{\Sigma}$, у которой все собственные значения λ_l $(l = 1,\ldots,m)$ различны. Тогда:

1) собственные значения $\widehat{\lambda}_1,\ldots,\widehat{\lambda}_m$ и отвечающие им собственные векторы $\widehat{\boldsymbol{c}}_1,\ldots,\widehat{\boldsymbol{c}}_m$ выборочной ковариационной матрицы $\widehat{\boldsymbol{\Sigma}}$ являются оценками максимального правдоподобия для соответствующих $\lambda_1,\ldots,\lambda_m$, $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m$, и, следовательно, обладают асимптотической эффективностью;

2) $\sqrt{n}\,(\widehat{\lambda}_l - \lambda_l) \xrightarrow{d} \xi \sim \mathcal{N}(0, 2\lambda_l^2)$ при $n \to \infty$;

3) $\sqrt{n}\,(\widehat{\boldsymbol{c}}_l - \boldsymbol{c}_l) \xrightarrow{d} \boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$ при $n \to \infty$, где элементами матрицы $\boldsymbol{V}$ служат $v_{kl} = \lambda_l \sum\limits_{k\neq l} |\boldsymbol{c}_l|^2\lambda_k/(\lambda_k - \lambda_l)^2$.

Построим на основе свойства 2 асимптотический доверительный интервал для λ_1 в примере 1. Так как свойство 2 равносильно сходимости $\sqrt{n/2}\,(\widehat{\lambda}_l/\lambda_l - 1) \xrightarrow{d} \xi/(\lambda_l\sqrt{2}) \sim \mathcal{N}(0,1)$, то при достаточно больших n приближенно с вероятностью $1-\alpha$

выполняется неравенство

$$\widehat{\lambda}_l \Big/ \left(1 + x_{1-\alpha/2}\sqrt{2/n}\right) < \lambda_l < \widehat{\lambda}_l \Big/ \left(1 - x_{1-\alpha/2}\sqrt{2/n}\right),$$

где $x_{1-\alpha/2}$ — квантиль уровня $(1-\alpha/2)$ закона $\mathcal{N}(0,1)$.
Для $\alpha = 5\%$ по таблице Т2 находим, что $x_{1-\alpha/2} = 1{,}96$. При $n = 43$ получаем 95%-ный доверительный интервал для λ_1 от 2,88 до 7,1.