пировке параметров и т. д. (см. гл. 13 в [1]). При решении помещен- ниже задач предполагается использование только метода главных онент (МГК).

Метод главных компонент. Во многих задачах обработки много- ых наблюдений и, в частности, в задачах классификации исследова- интересуют в первую очередь лишь те признаки, которые обнаружи- т наибольшую изменчивость (наибольший разброс) при переходе от го объекта к другому.

С другой стороны, не обязательно для описания состояния объек- спользовать какие-то из исходных, непосредственно замеренных на признаков. Так, например, для определения специфики фигуры че- ка при покупке одежды достаточно назвать значения двух признаков мер–рост), являющихся производными от измерений ряда параметров уры. При этом, конечно, теряется какая-то доля информации (портной еряет до одиннадцати параметров на клиенте), как бы огрубляются агрегировании) получающиеся при этом классы. Однако, как пока- исследования, к вполне удовлетворительной классификации людей с ки зрения специфики их фигуры приводит система, использующая три знака, каждый из которых является некоторой комбинацией от боль- о числа непосредственно замеряемых на объекте параметров.

Именно эти принципиальные установки заложены в сущность того ейного преобразования исходной системы признаков, которое приводит лавным компонентам. Формализуются же эти установки следующим азом.

Следуя общей оптимизационной постановке задачи снижения размер- ти (4.2) и полагая анализируемый признак X p-мерной случайной ве- иной с вектором средних значений $\boldsymbol{a} = (a^{(1)}, \ldots, a^{(p)})$ и ковариацион- матрицей $\boldsymbol{\Sigma} = (\sigma_{ij})$ $(i, j = 1, 2, \ldots, p)$, вообще говоря, неизвестными, еделим в качестве класса $\mathbf{F}(X)$ допустимых преобразований исследуе- признаков $x^{(1)}, x^{(2)}, \ldots, x^{(p)}$ их всевозможные *линейные ортогональ- нормированные комбинации*, т. е.

$$\mathbf{F} = \left\{ Z\colon\ z^{(j)} = \sum_{\nu=1}^{p} c_{j\nu}\bigl(x^{(\nu)} - a^{(\nu)}\bigr),\ j = 1, 2, \ldots, p \right\},$$

$$\sum_{\nu=1}^{p} c_{j\nu}^2 = 1 \quad \text{и} \quad \sum_{\nu=1}^{p} c_{j\nu} c_{k\nu} = 0 \tag{4.3}$$

$j = 1, 2, \ldots, p$ и $k = 1, 2, \ldots, p$, но $j \neq k$, а в качестве крите- (меры) информативности p'-мерной системы показателей $Z(X) =$

$(z^{(1)}(X), z^{(2)}(X), \ldots, z^{(p')}(X))$ выражение

$$I_{p'}\big(Z(X)\big) = \frac{\mathbf{D}\, z^{(1)} + \cdots + \mathbf{D}\, z^{(p')}}{\mathbf{D}\, x^{(1)} + \cdots + \mathbf{D}\, x^{(p)}}. \tag{4.4}$$

Тогда при любом фиксированном $p' = 1, 2, \ldots, p$ вектор искомых вспомогательных переменных $\widetilde{Z}(X) = (\widetilde{z}^{(1)}(X), \ldots, \widetilde{z}^{(p')}(X))^\top$ определяется как такая линейная комбинация

$$\widetilde{Z} = \mathbf{L}X \tag{4.5}$$

(где X — вектор-столбец *центрированных* исходных переменных, а

$$\mathbf{L} = \begin{pmatrix} l_{11} & \ldots & l_{1p} \\ \ldots & \ldots & \ldots \\ l_{p'1} & \ldots & l_{p'p} \end{pmatrix} -$$

матрица, строки которой удовлетворяют условию ортогональности), что

$$I_{p'}\big(\widetilde{z}^{(1)}(X), \ldots, \widetilde{z}^{(p')}(X)\big) = \max_{Z(X) \in \mathbf{F}} I_{p'}\big(Z(X)\big).$$

Полученные таким образом переменные $\widetilde{z}^{(1)}(X), \ldots, \widetilde{z}^{(p)}(X)$ и называют главными компонентами вектора X. Отсюда вытекает следующее определение главных компонент.

Первой главной компонентой *$\widetilde{z}^{(1)}(X)$ исследуемой системы показателей $X = (x^{(1)}, \ldots, x^{(p)})^\top$ называется такая нормированно-центрированная линейная комбинация этих показателей, которая среди всех прочих нормированно-центрированных линейных комбинаций переменных $x^{(1)}, \ldots, x^{(p)}$ обладает наибольшей дисперсией.*

k-й главной компонентой *$\widetilde{z}^{(k)}(X)$ $(k = 2, 3, \ldots, p)$ исследуемой системы показателей $X = (x^{(1)}, \ldots, x^{(p)})^\top$ называется такая нормированно-центрированная линейная комбинация этих показателей, которая не коррелирована с $k-1$ предыдущими главными компонентами и среди всех прочих нормированно-центрированных и некоррелированных с предыдущими $k-1$ главными компонентами линейных комбинаций переменных $x^{(1)}, \ldots, x^{(p)}$ обладает наибольшей дисперсией.*

З а м е ч а н и е 1 (переход к центрированным переменным). Поскольку, как увидим ниже, решение задачи (а именно вид матрицы линейного преобразования $\mathbf{L}$) зависит только от элементов ковариационной матрицы $\mathbf{\Sigma}$, которые в свою очередь не изменяются при замене исходных переменных $x^{(j)}$ переменными $x^{(j)} - c^{(j)}$ ($c^{(j)}$ — произвольные постоянные а), то в дальнейшем будем считать, что исходная система показателей *центрирована*, т.е. что $\mathbf{E}\,x^{(j)} = 0$, $j = 1, 2, \ldots, p$. В статистической тике этого добиваются, переходя к наблюдениям $\widetilde{x}_i^{(j)} = x_i^{(j)} - \overline{x}^{(j)}$, где $= \sum_{i=1}^{n} x_i^{(j)}/n$ (для упрощения обозначений волнистую черту над цен- ованной переменной и над главной компонентой в дальнейшем ста- не будем).

З а м е ч а н и е 2 (переход к выборочному варианту). Поскольку в ьных статистических задачах располагаем *лишь оценками* $\widehat{\boldsymbol{a}}$ и $\widehat{\boldsymbol{\Sigma}}$ со- тственно вектора средних $\boldsymbol{a}$ и ковариационной матрицы $\boldsymbol{\Sigma}$, то во всех нейших рассуждениях под $\widehat{a}^{(j)}$ понимается $\overline{x}^{(j)}$, а под $\widehat{\sigma}_{kj}$ — выбороч- ковариация $\widehat{\sigma}_{kj} = \sum_{i=1}^{n}(x_i^{(k)} - \overline{x}^{(k)})(x_i^{(j)} - \overline{x}^{(j)})/n$ $(j, k = 1, 2, \ldots, p)$.

З а м е ч а н и е 3. Использование главных компонент оказывается олее естественным и плодотворным в ситуациях, в которых все ком- нты $x^{(1)}, x^{(2)}, \ldots, x^{(p)}$ исследуемого вектора X имеют *общую физиче- природу* и соответственно *измерены в одних и тех же единицах*. К м примерам можно отнести исследование структуры бюджета време- ндивидуумов (все $x^{(j)}$ измеряются в единицах времени), исследование ктуры потребления семей (все $x^{(j)}$ измеряются в денежных единицах), едование общего развития и умственных способностей индивидуумов мощью специальных тестов (все $x^{(j)}$ измеряются в баллах), разного ро- нтропологические исследования (все $x^{(j)}$ измеряются в единицах меры ы) и т.д. Если же признаки $x^{(1)}, x^{(2)}, \ldots, x^{(p)}$ измеряются в различ- единицах, то результаты исследования с помощью главных компонент т существенно зависеть от выбора масштаба и природы единиц из- ния. Поэтому в подобных ситуациях исследователь предварительно ходит к вспомогательным безразмерным признакам $x^{*(j)}$, например, мощью нормирующего преобразования

$$x_i^{*(j)} = \frac{x_i^{(j)} - \overline{x}^{(j)}}{\sqrt{\widehat{\sigma}_{jj}}}, \qquad (4.1')$$

$$j = 1, 2, \ldots, p;$$
$$i = 1, 2, \ldots, n,$$

$\widehat{\sigma}_{jj}$ соответствует ранее введенным обозначениям, а затем строит глав- компоненты относительно этих вспомогательных признаков X^* и их риационной матрицы $\widehat{\boldsymbol{\Sigma}}_{X^*}$, которая, как легко видеть, является одно- менно выборочной *корреляционной* матрицей $\widehat{\mathbf{R}}$ исходных наблюдений

Вычисление главных компонент. Из определения главных компонент следует, что для вычисления первой главной компоненты необходимо решить оптимизационную задачу вида (4.2), т. е. в данном случае:

$$\begin{cases} \mathbf{D}\,(l_1X) \to \max\limits_{l_1}; \\ l_1 l_1^\top = 1, \end{cases} \tag{4.6}$$

где l_1 — первая строка матрицы $\mathbf{L}$ (см. (4.5)). Учитывая центрированность переменной X (т. е. $\mathbf{E}\,X = \mathbf{0}$) и то, что $\mathbf{E}\,(XX^\top) = \boldsymbol{\Sigma}$, имеем

$$\mathbf{D}\,(l_1X) = \mathbf{E}\,(l_1X)^2 = \mathbf{E}\,(l_1XX^\top l_1^\top) = l_1\boldsymbol{\Sigma} l_1^\top.$$

Следовательно, задача (4.6) может быть записана

$$\begin{cases} l_1\boldsymbol{\Sigma} l_1^\top \to \max\limits_{l_1}; \\ l_1 l_1^\top = 1, \end{cases} \tag{4.6'}$$

Вводя функцию Лагранжа $\varphi(l_1,\lambda) = l_1\boldsymbol{\Sigma} l_1^\top - \lambda(l_1l_1^\top - 1)$ и дифференцируя ее по компонентам вектор-столбца $l_1^\top$, имеем:

$$\frac{\partial\varphi}{\partial l_1^\top} = 2\boldsymbol{\Sigma} l_1^\top - 2\lambda l_1^\top,$$

что дает систему уравнений для определения l_1:

$$(\boldsymbol{\Sigma} - \lambda\mathbf{I})l_1^\top = \mathbf{0} \tag{4.7}$$

(здесь $\mathbf{0} = (0,0,\ldots,0)^\top$ — p-мерный вектор-столбец из нулей).

Для того чтобы существовало ненулевое решение системы (4.7) (а оно должно быть ненулевым, так как $l_1l_1^\top = 1$), матрица $\boldsymbol{\Sigma} - \lambda\mathbf{I}$ должна быть *вырожденной*, т. е.

$$|\boldsymbol{\Sigma} - \lambda\mathbf{I}| = 0. \tag{4.8}$$

Этого добиваются подбором соответствующего значения λ. Уравнение (4.8) (относительно λ) называется *характеристическим* для матрицы $\boldsymbol{\Sigma}$. Известно (см. [1], Приложение 2), что при симметричности и неотрицательной определенности матрицы $\boldsymbol{\Sigma}$ (каковой она и является как всякая ковариационная или корреляционная матрица) это уравнение имеет p вещественных неотрицательных корней $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$, называемых *характеристическими* (или *собственными*) *значениями* матрицы $\boldsymbol{\Sigma}$.

Учитывая, что $\mathbf{D}\,\widetilde{z}^{(1)} = \mathbf{D}\,(l_1X) = l_1\boldsymbol{\Sigma}\,l_1^\top$ и $l_1\boldsymbol{\Sigma}\,l_1^\top = \lambda$ (последнее соотношение следует из соотношения (4.7) после его умножения слева на l_1, с учетом $l_1l_1^\top = 1$), получаем $\mathbf{D}\,\widetilde{z}^{(1)}(X) = \lambda$.

ому для обеспечения максимальной величины дисперсии пере-
$^{(1)}$ нужно выбрать из p собственных значений матрицы $\mathbf{\Sigma}$ *наи-*
т. е.

$$\mathbf{D}\,\widetilde{z}^{(1)}(X) = \lambda_1.$$

тавляем λ_1 в систему уравнений (4.7) и, решая ее относительно
$_p$, определяем компоненты вектора l_1.

им образом, ***первая главная компонента получается как линей-***
бинация $\widetilde{z}^{(1)}(X) = l_1 X$, где l_1 — собственный вектор матрицы
ветствующий наибольшему собственному числу этой матрицы.
ее аналогично можно показать, что $\widetilde{z}^{(k)}(X) = l_k X$, где l_k — соб-
й вектор матрицы $\mathbf{\Sigma}$, соответствующий k-му по величине соб-
му значению λ_k этой матрицы.

им образом, соотношения для определения всех p главных ком-
вектора X могут быть представлены в виде (4.5), где $\widetilde{Z} =$
$, \widetilde{z}^{(p)})^\top$, $X = (x^{(1)}, \ldots, x^{(p)})^\top$, а матрица $\mathbf{L}$ состоит из строк
$, \ldots, l_{jp})$, $j = \overline{1,p}$, являющихся собственными векторами матрицы
ветствующими собственным числам λ_j. При этом сама матрица $\mathbf{L}$
етствии с условиями (4.3) является *ортогональной*, т. е.

$$\mathbf{L}\mathbf{L}^\top = \mathbf{L}^\top\mathbf{L} = \mathbf{I}. \tag{4.9}$$

дальнейшем в целях упрощения обозначений мы будем опускать
ну» над переменными главных компонент, т. е. обозначать главные
енты просто $Z = (z^{(1)}, z^{(2)}, \ldots, z^{(p')})$.

сновные числовые характеристики главных компонент.
елим основные числовые характеристики (средние значения, дис-
, ковариации) главных компонент в терминах основных числовых
теристик исходных переменных и собственных значений матрицы

) $\mathbf{E}\,Z = \mathbf{E}(\mathbf{L}\,X) = \mathbf{L}\cdot\mathbf{E}\,X = \mathbf{0}$;

) ковариационная матрица вектора главных компонент:

$$\begin{aligned}\mathbf{\Sigma}_Z = \mathbf{E}(ZZ^\top) &= \mathbf{E}\big((\mathbf{L}\,X)(\mathbf{L}\,X)^\top\big) = \mathbf{E}(\mathbf{L}XX^\top\mathbf{L}^\top)\\ &= \mathbf{L}\cdot\mathbf{E}(XX^\top)\cdot\mathbf{L}^\top = \mathbf{L}\cdot\mathbf{\Sigma}\cdot\mathbf{L}^\top.\end{aligned}$$

Умножая слева соотношения

$$(\mathbf{\Sigma} - \lambda_k\mathbf{I})l_k^\top = \mathbf{0}, \qquad k = \overline{1,p},$$

на l_j $(j = \overline{1,p})$, получаем, что

$$\mathbf{L\Sigma L}^\top = \mathbf{\Sigma}_Z = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix}. \tag{4.10}$$

Из (4.10), в частности, следует подтверждение взаимной некоррелированности главных компонент, а также тот факт, что $\mathbf{D}\, z^{(k)} = \lambda_k$ $(k = \overline{1,p})$;

в) сумма дисперсий исходных признаков равна сумме дисперсий всех главных компонент. Действительно,

$$\sum_{k=1}^{p} \mathbf{D}\, z^{(k)} = \operatorname{tr} \mathbf{\Sigma}_Z = \operatorname{tr}(\mathbf{L\Sigma L}^\top) = \operatorname{tr}\big((\mathbf{L\Sigma})\mathbf{L}^\top\big)$$

$$= \operatorname{tr}\big(\mathbf{L}^\top(\mathbf{L\Sigma})\big) = \operatorname{tr}\big((\mathbf{L}^\top\mathbf{L})\mathbf{\Sigma}\big) = \operatorname{tr}\mathbf{\Sigma} = \sum_{k=1}^{p} \mathbf{D}\, x^{(k)};$$

г) обобщенная дисперсия исходных признаков X $(\det \mathbf{\Sigma})$ равна обобщенной дисперсии главных компонент. Действительно, обобщенная дисперсия вектора Z равна

$$\det \mathbf{\Sigma}_Z = \det(\mathbf{L\Sigma L}^\top) = \det\big((\mathbf{L\Sigma})\mathbf{L}^\top\big) = \det\big(\mathbf{L}^\top(\mathbf{L\Sigma})\big)$$
$$= \det\big((\mathbf{LL}^\top)\mathbf{\Sigma}\big) = \det(\mathbf{\Sigma}). \tag{4.10$'$}$$

Следствие 1. *Из б) и в), в частности, следует, что критерий информативности метода главных компонент* (4.4) *может быть представлен в виде*

$$I_{p'}\big(Z(X)\big) = \frac{\lambda_1 + \dots + \lambda_{p'}}{\lambda_1 + \dots + \lambda_p}, \tag{4.4$'$}$$

где $\lambda_1, \lambda_2, \dots, \lambda_p$ — собственные числа ковариационной матрицы $\mathbf{\Sigma}$ вектора X, расположенные в порядке убывания.

Кстати, представление $I_{p'}(Z(X))$ в виде (4.4$'$) дает исследователю некоторую основу, опорную точку зрения, при вынесении решения о том, сколько последних главных компонент можно без особого ущерба изъять из рассмотрения, сократив тем самым размерность исследуемого пространства.

Действительно, анализируя с помощью (4.4$'$) изменение относительно доли дисперсии, вносимой первыми p' главными компонентами, в зависимости от числа этих компонент, можно разумно определить число

нент, которое целесообразно оставить в рассмотрении. Так, при из-
нии $I_{p'}$, изображенном на рис. 4.1, очевидно, целесообразно было бы
атить размерность пространства с $p = 10$ до $p' = 3$, так как добавле-
всех остальных семи главных компонент может повысить суммарную
ктеристику рассеяния не более чем на 10%.

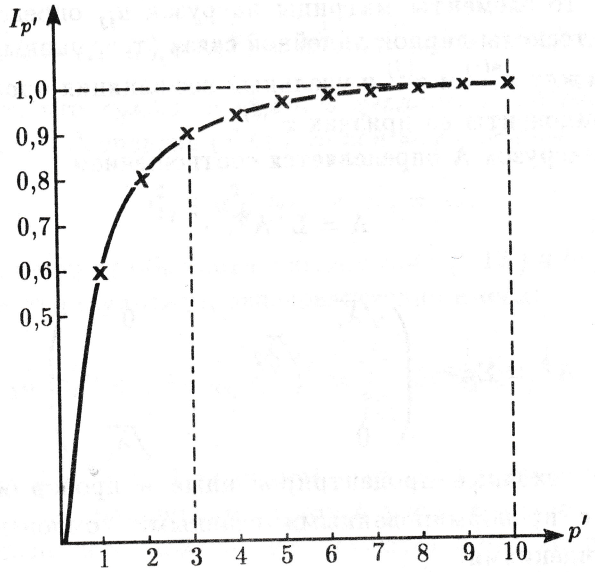

Рис. 4.1. Изменение относительной доли суммарной дисперсии исследуемых признаков, обусловленной первыми p' главными компонентами, в зависимости от p' (случай $p = 10$)

Следствие 2. *Если X^* — вектор нормированных признаков* $, \ldots, x^{*(p)}$, *т. е.* $\mathbf{E}\, x^{*(j)} = 0$ *и* $\mathbf{D}\, x^{*(j)} = 1$ *для* $j = \overline{1,p}$, *то соглас- замечанию 3 ковариационная и корреляционные матрицы совпадают* *е.* $\boldsymbol{\Sigma}_{X^*} = \mathbf{R}$) *и из б) и в) следует*

$$\operatorname{tr} \boldsymbol{\Sigma}_Z = \operatorname{tr} \boldsymbol{\Sigma}_{X^*} = \operatorname{tr} \mathbf{R} = p$$

$$\sum_{k=1}^{p} \lambda_k = p.$$

огда критерий информативности (4.4′) *может быть представлен в иде*

$$I_{p'}\big(Z(X)\big) = \frac{\lambda_1 + \cdots + \lambda_{p'}}{p}. \qquad (4.4'')$$

Матрица «нагрузок» $\mathbf{A} = (a_{ij})$, $i, j = 1, 2, \ldots, p$, главных компонент на исходные признаки также является важной характеристикой главных компонент. Если анализируемые переменные $X = (x^{(1)}, x^{(2)}, \ldots, x^{(p)})^\top$ предварительно процентрированы и пронормированы (см. выше замечания 1 и 3), т. е. если главные компоненты строятся для признаков $X^* = (x^{*(1)}, x^{*(2)}, \ldots, x^{*(p)})^\top$, $\mathbf{E}\, x^{*(i)} = 0$, $\mathbf{D}\, x^{*(i)} = 1$, $i = 1, 2, \ldots, p$, то элементы матрицы нагрузок a_{ij} определяют одновременно степень тесноты парной линейной связи (т. е. парный коэффициент корреляции) между $x^{*(i)}$ и $z^{(j)}$ и удельный вес влияния *пронормированной* j-й главной компоненты на признак $x^{*(i)}$.

Матрица нагрузок $\mathbf{A}$ определяется соотношением

$$\mathbf{A} = \mathbf{L}^\top \mathbf{\Lambda}^{\frac{1}{2}}, \tag{4.11}$$

где

$$\mathbf{\Lambda}^{\frac{1}{2}} = \mathbf{\Sigma}_Z^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & & & 0 \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ 0 & & & \sqrt{\lambda_p} \end{pmatrix}.$$

Она связывает исходные процентрированные и пронормированные переменные X^* с пронормированными главными компонентами $z_{\text{н}}^{(j)} = z^{(j)}/\sqrt{\lambda j}$ соотношениями:

$$X^* = \mathbf{L}^\top \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{\Lambda}^{-\frac{1}{2}} Z = \mathbf{A}\, Z_{\text{н}}, \tag{4.12}$$

т. е.

$$x^{*(i)} = a_{i1} z_{\text{н}}^{(1)} + a_{i2} z_{\text{н}}^{(2)} + \cdots + a_{ip} z_{\text{н}}^{(p)}. \tag{4.12$'$}$$

Таким образом, коэффициент a_{ij} действительно определяет удельный вес влияния j-й нормированной главной компоненты $z_{\text{н}}^{(j)}$ на i-й исходный признак. Это подтверждается, в частности, и следующим фактом:

$$\begin{aligned} r\big(x^{*(i)}, z^{(j)}\big) &= \frac{\mathbf{E}\,[(x^{*(i)} - \mathbf{E}\, x^{*(i)})(z^{(j)} - \mathbf{E}\, z^{(j)})]}{\sqrt{\mathbf{D}\, x^{*(i)}\, \mathbf{D}\, z^{(j)}}} \\ &= \frac{\mathbf{E}\,(x^{*(i)} z^{(j)})}{\sqrt{\lambda_j}} = \mathbf{E}\left(x^{*(i)} \frac{z^{(j)}}{\sqrt{\lambda_j}}\right) \\ &= \mathbf{E}\,\big(x^{*(i)} z_{\text{н}}^{(j)}\big) = \mathbf{E}\left[\big(a_{i1} z_{\text{н}}^{(1)} + \cdots + a_{ip} z_{\text{н}}^{(p)}\big) z_{\text{н}}^{(j)}\right] \\ &= a_{ij}\, \mathbf{E}\left(\big(z_{\text{н}}^{(j)}\big)^2\right) = a_{ij}, \end{aligned} \tag{4.13}$$