Анализ главных компонент. Примеры

Пример 1.

Рассмотрим случай двух переменных x_1 и x_2 . То есть в таблице данных два столбца.

Найдем главные компоненты. Задача простая, она позволяет провести все вычисления вручную и тем самым проверить наше понимание темы.

Шаг 1

Обозначим коэффициент корреляции r. Тогда корреляционная матрица R имеет вид

$$R=\begin{pmatrix}1 & r\\ r & 1\end{pmatrix}$$

Рассмотрим случай, когда $r>0$.

Шаг 2

Найдем собственные числа матрицы R, они равны дисперсиям главных компонент.

Собственные числа λ будут корнями уравнения

$$|R-\lambda\cdot E|=0$$

где E – единичная матрица,

$$E=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

Приравниваем определитель нулю, решаем квадратное уравнение

$$(1-\lambda)^2-r^2=0$$

Получаем два решения $\lambda_1=1+r$ и $\lambda_2=1-r$

Вопрос 1. Могут ли в данной задаче собственные числа быть отрицательными?
Вопрос 2. Могут ли собственные числа быть отрицательными в других практических задачах?
Вопрос 3. Могут ли некоторые собственные числа равняться нулю?
Вопрос 4. Могут ли все собственные числа равняться нулю?
Вопрос 5. Причем тут дисперсия?
Вопрос 6. В нашей задаче сумма собственных чисел оказалась равна двум. Можно ли это было предвидеть до нахождения значений?
Вопрос 7. Причем здесь trace(R) - след матрицы R?

Шаг 3

Найдем собственные векторы матрицы ***R***. Они нужны для вычисления значений главных компонент.

Первая главная компонента соответствует наибольшему собственному числу. Так как ***r***>0, то $\lambda_1 > \lambda_2$

Обозначим первый собственный вектор $\boldsymbol{a_1} = \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix}$. Он соответствует первому собственному числу λ_1 .

Из определения собственного вектора получаем уравнение для $\boldsymbol{a_1}$ в матричном виде $R\,a_1 = \lambda_1 a_1$. Оно эквивалентно системе уравнений

$$a_{11} + r a_{12} = (1+r) a_{11}$$
$$a_{11} + r a_{12} = (1+r) a_{12}$$

После упрощений получаем два одинаковых уравнения $a_{11} = a_{12}$

С учетом нормировки $\boldsymbol{a_1}^T \cdot \boldsymbol{a_1} = 1$ находим $a_{11} = a_{12} = \frac{1}{\sqrt{2}}$ или $a_{11} = a_{12} = -\frac{1}{\sqrt{2}}$

Аналогично, для второго собственного вектора $\boldsymbol{a_2} = \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix}$ получаем $a_{21} = \frac{1}{\sqrt{2}}$ и $a_{22} = -\frac{1}{\sqrt{2}}$. Или $a_{21} = -\frac{1}{\sqrt{2}}$ и $a_{22} = \frac{1}{\sqrt{2}}$

<u>Шаг 4</u>

Теперь можно записать формулы для главных компонент

$$y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 = \frac{1}{\sqrt{2}}(x_1 + x_2)$$
$$y_2 = a_{21} \cdot x_1 + a_{22} \cdot x_2 = \frac{1}{\sqrt{2}}(x_1 - x_2)$$

Вопрос 8. Что известно про третью главную компоненту?

Задача 1. Повторить выкладки для случая $r < 0$

Замечание. Случай $r = 0$ похитрее. Без доказательства заметим, что можно выбрать бесконечно много пар собственных векторов. Подойдут любые два ортонормированных вектора.

Замечание. Любой собственный вектор можно умножить на минус единицу. Традиция предписывает выбор варианта, при котором первая координата вектора неотрицательная.

Замечание. В нашей задаче главные компоненты не зависят от коэффициента корреляции ***r***. Это неожиданно, но так бывает. Но посчитаем долю общей дисперсии (total variance), приходящейся на первую главную компоненту. Она равна $\frac{1+r}{2}$. Чем больше корреляция, тем сильнее дублируется информация в переменных x_1 и x_2, тем более информативной будет первая компонента.

Деталь 4. Известно, что если из собственных векторов составить матрицу, она будет ортогональной. Обозначим эту матрицу ***A***. Справедливо

$$A^{-1}\cdot R\cdot A = A^{T}\cdot R\cdot A = \Lambda$$

где

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Проверим справедливость равенства в нашей задаче

$$A = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}}\cdot\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$A^{T}\cdot R\cdot A = \frac{1}{\sqrt{2}}\cdot\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\cdot\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\cdot\frac{1}{\sqrt{2}}\cdot\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = .$$

$$. = \frac{1}{2}\cdot\begin{pmatrix} 2(1+r) & 0 \\ 0 & 2(1-r) \end{pmatrix} = \begin{pmatrix} 1+r & 0 \\ 0 & 1-r \end{pmatrix} = \Lambda$$