3 Метод главных компонент.

Метод главных компонент осуществляет переход к новой совокупности некоррелированных признаков $\eta_1, \ldots, \eta_m$, каждый из которых является линейной комбинацией исходных признаков $\xi_1,\ \xi_2,\ \ldots,\ \xi_p$. При этом линейные комбинации выбираются таким образом, чтобы среди всех возможных линейных нормированных комбинаций исходных признаков первая главная компонента η_1 обладала наибольшей дисперсией. Геометрически это выглядит как ориентация новой координатной оси η_1 вдоль направления наибольшей вытянутости эллипсоида рассеивания исследуемой выборки в пространстве признаков $\xi_1,\ \xi_2,\ \ldots,\ \xi_p$. Вторая главная компонента имеет наибольшую дисперсию среди всех оставшихся линейных преобразований, некоррелированных с первой главной компонентой. Она интерпретируется как направление наибольшей вытянутости эллипсоида рассеивания, перпендикулярное первой главной компоненте. Следующие главные компоненты определяются по аналогичной схеме.

В дальнейшем из полученных величин можно оставить только $m \quad (m < p)$ наиболее значимых факторов, вносящих максимальный вклад в суммарную дисперсию и использовать эти величины как некие интегральные факторы, характеризующие всю совокупность признаков.

Пусть $\xi = (\xi_1,\ \xi_2,\ \ldots,\ \xi_p)$ – центрированная многомерная случайная величина с матрицей ковариаций Σ, то есть $E\xi = (0, \ldots, 0)$ и $E(\xi_i \cdot \xi_j) = \sigma_{i\,j}$. Положим

$$\eta_i = \beta^{(i)T}\xi = \beta_1^{(i)}\xi_1 + \beta_2^{(i)}\xi_2 + \ldots + \beta_p^{(i)}\xi_p, \tag{16}$$

где $\beta^{(i)}$ – векторы неизвестных коэффициентов преобразования, $i = \overline{1, m}$, $(m \leq p)$, причем параметр m заранее не определен, а определяется в процессе получения главных компонент. Будем называть величины η_i главными компонентами (факторами в терминах факторного анализа). В матричной форме приведение вектора исходных признаков ξ к главным компонентам можно записать как:

$$\eta = B^T\xi,$$

где

$$B = \begin{pmatrix} \beta_1^{(1)} & \beta_1^{(2)} & & \beta_1^{(p)} \\ \beta_2^{(1)} & \beta_2^{(2)} & & \beta_2^{(p)} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_p^{(1)} & \beta_p^{(2)} & & \beta_p^{(p)}. \end{pmatrix}$$

Так как величины ξ_i центрированы, то $M(\eta_i) = 0,\ i = \overline{1, m}$, а дисперсия главных компонент определяется как:

$$D\left(\eta_i\right) = E\left(\eta_i\eta_i^T\right) = \beta^{(i)T}E\left(\xi\xi^T\right)\beta^{(i)} = \beta^{(i)T}\Sigma\beta^{(i)} \tag{17}$$

На коэффициенты преобразования $\beta^{(i)}$ накладываем условиям нормировки:

$$||\beta^{(i)}||^2 = \beta^{(i)T}\beta^{(i)} = 1. \tag{18}$$

Рассмотрим далее процесс нахождения главных компонент.

Шаг 1. Будем искать первую компоненту в виде

$$\eta_1 = \beta^{(1)T}\xi,$$

где вектор параметров преобразования $\beta^{(1)}$ будем выбирать таким образом, чтобы дисперсия первой главной компоненты была наибольшей, то есть будем решать задачу оптимизации:

$$D\eta_1 = \beta^{(1)T}\Sigma\beta^{(1)} \to \max_{\beta^{(1)}}$$

при ограничении:

$$\beta^{(1)T}\beta^{(1)} = 1.$$

Таким образом, получаем задачу исследования функции на условный экстремум. Составим функцию Лагранжа:

$$L(\lambda) = \beta^{(1)T}\Sigma\beta^{(1)} - \lambda(\beta^{(1)T}\beta^{(1)} - 1) = \sum_i \sum_j \beta_i^{(1)}\beta_j^{(1)}\sigma_{ij} - \lambda\left(\sum_i \left(\beta_i^{(1)}\right)^2 - 1\right)$$

Исследуем функцию Лагранжа на экстремум:

$$\frac{\partial L}{\partial \beta^{(1)}} = 2\Sigma\beta^{(1)} - 2\lambda\beta^{(1)} = 0 \tag{19}$$

$$\frac{\partial L}{\partial \lambda} = \beta^{(1)T}\beta^{(1)} - 1 = 0 \tag{20}$$

Из первого уравнения системы получаем уравнение $(\Sigma - \lambda \cdot E)\beta^{(1)} = 0$. Так как нас интересуют лишь нетривиальные решения этого матричного уравнения, то для их нахождения необходимо потребовать, чтобы матрица $\Sigma - \lambda \cdot E$ была вырожденной, то есть

$$|\Sigma - \lambda \cdot E| = 0.$$

Получаем характеристическое уравнение для нахождения собственных значений (множителей Лагранжа) ковариационной матрицы Σ, которое имеет в общем случае p корней с учетом их кратности. Какой же из корней выбрать для получения вектора коэффициентов $\beta^{(1)}$.

Далее, рассмотрим уравнение $(\Sigma - \lambda \cdot E)\beta^{(1)} = 0$ для нахождения вектора коэффициентов $\beta^{(1)}$. Если это уравнение умножить слева на $\beta^{(1)T}$, то

$$\beta^{(1)T}\Sigma\beta^{(1)} = D\eta_1 = \beta^{(1)T}\lambda\beta^{(1)} = \lambda \quad \text{так как} \quad \beta^{(1)T}\beta^{(1)} = 1.$$

Таким образом, дисперсия первой главной компоненты совпадает с собственным значением характеристического уравнения, а значит, необходимо выбрать наибольшее собственное значение, так как мы хотим найти вектор коэффициентов $\beta^{(1)}$ с наибольшей дисперсией главных компонент. Обозначим наибольшее среди всех собственных значений λ_1. Тогда вектор $\beta^{(1)}$ представляет собой собственный вектор, соответствующий собственному значению λ_1. Найдя максимальное среди всех собственных значений λ_1 и соответствующий ему собственный вектор $\beta^{(1)}$, получаем первую главную компоненту $\eta_1 = \beta^{(1)T}\xi$.

Шаг 2. Для нахождения второй главной компоненты необходимо найти вектор коэффициентов $\beta^{(2)}$. Для этого будем искать компоненту с наибольшей дисперсий среди оставшихся компонент и, чтобы эта компнента была некоррелирована с первой,а также удовлетворяла условию нормировки. Условие некоррелированности примет вид:

$$0 = E(\eta_2 \cdot \eta_1^T) = E(\beta^{(2)T}\xi \cdot (\beta_1\xi)^T) = \beta^{(2)T}E(\xi\xi^T)\beta^{(1)} = \beta^{(2)T}\Sigma\beta^{(1)}$$

Поскольку на первом шаге показано, что $\Sigma\beta^{(1)} = \lambda_1\beta^{(1)}$, то условие некоррелированности первой и второй главных компонент можно переписать в виде:

$$\beta^{(2)T}\beta^{(1)} = 0,$$

то есть некоррелированность геометрически означает ортоганальность векторов $\beta^{(1)}$ и $\beta^{(2)}$.

Задача оптимизации для нахождения второй главной компоненты имеет вид:

$$\begin{cases} D\eta_2 = \beta^{(2)T}\,\Sigma\,\beta^{(2)} \to \max\limits_{\beta^{(2)}} \quad \text{при ограничениях} \\ \beta^{(2)T}\beta^{(2)} = 1 \quad \text{условие нормировки,} \\ \beta^{(2)T}\beta^{(1)} = 0 \quad \text{условие некоррелированности.} \end{cases} \tag{21}$$

Составим функцию Лагранжа:

$$L(\lambda, \nu) = \beta^{(2)T}\,\Sigma\,\beta^{(2)} - \lambda(\beta^{(2)T}\beta^{(2)} - 1) - 2\nu\beta^{(2)T}\lambda_1\beta^{(1)}$$

Исследуем ее на экстремум:

$$\begin{cases} \dfrac{\partial L}{\partial \beta^{(2)}} = 2\Sigma\beta^{(2)} - 2\lambda\beta^{(2)} - 2\nu\lambda_1\beta^{(1)} = 0 \\ \dfrac{\partial L}{\partial \lambda} = \beta^{(2)T}\beta^{(2)} - 1 = 0 \\ \dfrac{\partial L}{\partial \nu} = \beta^{(2)T}\beta^{(1)} = 0 \end{cases} \tag{22}$$

Умножая первое уравнение в системе (22) на $\beta^{(1)T}$, получим:

$$2\beta^{(1)T}\Sigma\beta^{(2)} - 2\lambda\beta^{(1)T}\beta^{(2)} - 2\nu\beta^{(1)T}\lambda_1\beta^{(1)} = 2\beta^{(1)T}\Sigma\beta^{(2)} - 2\lambda\left(\beta^{(2)T}\beta^{(1)}\right)^T - 2\nu\lambda_1 =$$
$$2\left(\beta^{(2)T}\Sigma\beta^{(1)}\right)^T - 2\nu\lambda_1 = 2\beta^{(2)T}\lambda_1\beta^{(1)} - 2\nu\lambda_1 = -2\nu\lambda_1 = 0.$$

Так как $\lambda_1 \neq 0 \Rightarrow \nu = 0$. Тогда возвращаясь к первому уравнению в системе (22) получаем:

$$2\Sigma\beta^{(2)} - 2\lambda\beta^{(2)} = 0 \Rightarrow (\Sigma - \lambda)\,\beta^{(2)} = 0.$$

Так как нас интересует ненулевой вектор коэффициентов $\beta^{(2)}$, то это матричное уравнение будет иметь нетривиальные решения, если их будет бесконечно много, то есть когда $|\Sigma - \lambda \cdot E| = 0$. Так как дисперсия второй компоненты снова будет равна λ, то для нахождения второй главной компоненты необходимо взять второе по величине собственное значение для ковариационной матрицы Σ. А тогда собственный вектор, соответствующий этому собственному значению и будет вектором искомых коэффициентов $\beta^{(2)}$. Таким образом, получаем вторую главную компоненту : $\eta_2 = \beta^{(2)}\,\xi$.

Продолжая этот процесс далее, рассмотрим $(k+1)$ **шаг**.
Необходимо найти компоненту $\eta_{k+1} = \beta^{(k+1)}\,\xi$, которая бы имела максимальную дисперсию по сравннению со всеми оставшимися ненайденными компонентами и была бы некоррелирована с компонентами, найденными на предыдущих шагах $\eta_1, \ldots, \eta_k$, то есть условие некоррелированности будет представлять собой систему из k уравнений:

$$\beta^{(k+1)T}\,\beta^{(i)} = 0, \quad i = \overline{1,k}.$$

Функция Лагранжа по аналогии со вторым шагом:

$$L(\lambda, \nu_1, \ldots, \nu_k) = \beta^{(k+1)T}\,\Sigma\,\beta^{(k+1)} - \lambda(\beta^{(k+1)T}\beta^{(k+1)} - 1) - 2\sum_{i=1}^{k} \nu_i \beta^{(k+1)T} \lambda_i \beta^{(i)}$$

Дифференцируем функцию Лагранжа по β^{k+1}:

$$\frac{\partial L}{\partial \beta^{(k+1)}} = 2\Sigma\beta^{(k+1)} - 2\lambda\beta^{(k+1)} - 2\sum_{i=1}^{k} \nu_i \lambda_i \beta^{(i)} = 0$$

Умножая полученное уравнение слева на $\beta^{(1)T}, \beta^{(2)T}, \ldots, \beta^{(k)T}$, получим k уравнений, аналогичных шагу 2, из которых получим, что все $\nu_i = 0, \ i = 1, ..., k$. А значит, вектор $\beta^{(k+1)}$ находится из уравнения $(\Sigma - \lambda)\,\beta^{(k+1)} = 0$, то есть тоже является собственным вектором ковариационной матрицы Σ. Соответствующее ему собственное значение по аналогичным причинам находится из уравнения $|\Sigma - \lambda \cdot E| = 0$ и является максимальных, среди всех корней характеристического уравнения, если не брать в расчет найденные первые k собственных значений.

Покажем, что собственные значения выбираются не просто по величине наибольшего значения среди всех оставшихся корней, а являются "упорядоченными"относительно главной диагонали элементами ковариационной матрицы вектора главных компонент η.