Метод главных компонент (МГК)
Principal component analysis (PCA)

12 мая 2020

Метод снижения размерности

Что имеем на «входе» задачи?

Ряд характеристик – признаков объектов, достаточно сильно скоррелированных между собой. Если информация дублируется, то почему бы не уменьшить количество признаков?

Метод снижения размерности

Что имеем на «входе» задачи?

Ряд характеристик – признаков объектов, достаточно сильно скоррелированных между собой. Если информация дублируется, то почему бы не уменьшить количество признаков?

Постановка задачи снижения размерности

От исходного признакового пространства размерности k, содержащего много дублирующей информации, перейти к новому признаковому пространству размерности p, где $p < k$. Корреляция между новыми признаками равна 0, то есть, дублирующая информация отсутствует.

Примеры: возможность построения индекса

Пример 1

1. количество раз обработки рук антисептиком за день
2. страх выходить на улицу
3. страх потери работы

Примеры: возможность построения индекса

Пример 1

1. количество раз обработки рук антисептиком за день
2. страх выходить на улицу
3. страх потери работы

Пример 2

1. доверие правительству
2. доверие парламенту
3. доверие политическим партиям

Примеры: возможность построения индекса

Пример 1

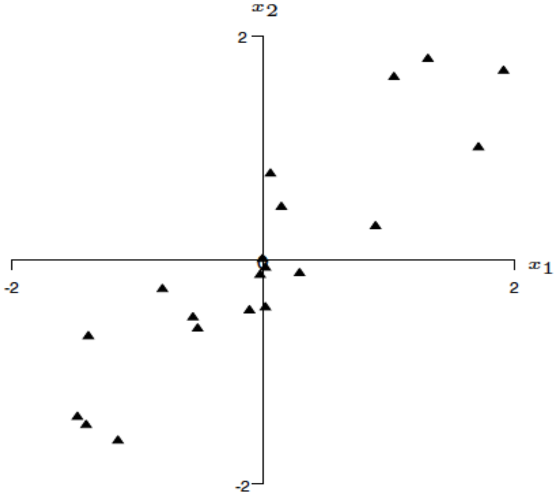
1. количество раз обработки рук антисептиком за день
2. страх выходить на улицу
3. страх потери работы

Пример 2

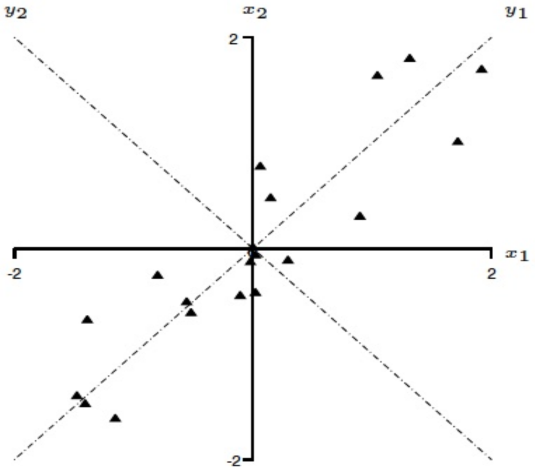
1. доверие правительству
2. доверие парламенту
3. доверие политическим партиям

Ваш пример?

Визуализация: на «входе» задачи



Визуализация: в результате преобразования



Как выполнить преобразование?

Алгоритм действий:

1. Переход к новой системе координат
   - Поворачиваем оси так, чтобы угол между осями составлял $90°$.
   - Минимизация суммы квадратов расстояний от точки до прямой.
2. Оставляем оси с большей долей дисперсии (информации), остальные оси отбрасываем.

Переход к новой системе координат

$\overrightarrow{y} = A\overrightarrow{x}$

Для перехода от $\overrightarrow{x}$ к $\overrightarrow{y}$ нам понадобится ортогональная матрица A.

- как столбцы, так и строки имеют длину $= 1$
- как столбцы, так и строки ортогональны
- $|\det(A)| = 1$
- $A^{-1} = A^T$

Про длину и ортогональность

Пример

$$\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

- Длина (вектор по первой строке) : $\sqrt{\left(\frac{1}{\sqrt{5}}\right)^2 + \left(\frac{-2}{\sqrt{5}}\right)^2} = 1$
- Ортогональность первого и второго столбцов: $\frac{1}{\sqrt{5}}\left(\frac{-2}{\sqrt{5}}\right) + \frac{2}{\sqrt{5}\sqrt{5}} = 0$ (скалярное произведение $= 0$)

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$
Вариация новых показателей:
$\overrightarrow{y}^T\overrightarrow{y} = (A\overrightarrow{x})^T(A\overrightarrow{x}) = \overrightarrow{x}^T A^T A\overrightarrow{x} =$

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$

Вариация новых показателей:

$\overrightarrow{y}^T\overrightarrow{y} = (A\overrightarrow{x})^T(A\overrightarrow{x}) = \overrightarrow{x}^T A^T A\overrightarrow{x} = \overrightarrow{x}^T I\overrightarrow{x} = \overrightarrow{x}^T\overrightarrow{x}$

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$
Вариация новых показателей:
$\overrightarrow{y}^T\overrightarrow{y} = (A\overrightarrow{x})^T(A\overrightarrow{x}) = \overrightarrow{x}^T A^T A\overrightarrow{x} = \overrightarrow{x}^T I\overrightarrow{x} = \overrightarrow{x}^T\overrightarrow{x}$

Принцип сохранения информации

Переходим от k-мерного к k-мерному ортогонализированному признаковому пространству (до удаления малоинформативных индексов). В результате МГК информация перераспределяется, при этом сохраняется полностью.

Для реализации МГК вспомним материал:

Собственные векторы и собственные значения

$M\overrightarrow{z} = \lambda\overrightarrow{z}$, где М – квадратная матрица, $\overrightarrow{z}$ – собственный вектор (ненулевой), λ – собственное значение.
В результате воздействия матрицы М получаем вектор, параллельный исходному вектору.

Для реализации МГК вспомним материал:

Собственные векторы и собственные значения

$M\overrightarrow{z} = \lambda\overrightarrow{z}$, где М – квадратная матрица, $\overrightarrow{z}$ – собственный вектор (ненулевой), λ – собственное значение.
В результате воздействия матрицы М получаем вектор, параллельный исходному вектору.

Характеристическое уравнение

$M\overrightarrow{z} = \lambda\overrightarrow{z} = \lambda I\overrightarrow{z}$, где I – единичная матрица.
$(M - \lambda I)\overrightarrow{z} = 0$
Может быть либо одно решение (нулевой $\overrightarrow{z}$), либо много решений. Для второго исхода матрица $(M - \lambda I)$ должна преобразовывать в 0 (вырожденная матрица):

$$\boxed{det(M - \lambda I) = 0}$$

Найдем собственные числа и векторы:

Пример:

Дана матрица $M = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$

Найдем собственные числа и векторы:

Пример:

Дана матрица $M = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$

Решение:

Найдем собственные числа:

$\det \begin{pmatrix} 5-\lambda & -4 \\ -4 & 5-\lambda \end{pmatrix} = 0$

Найдем собственные числа и векторы:

Пример:

Дана матрица M $= \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$

Решение:

Найдем собственные числа:

$$\det\begin{pmatrix} 5-\lambda & -4 \\ -4 & 5-\lambda \end{pmatrix} = 0 \rightarrow (5-\lambda)^2 - 16 = 0$$

Найдем собственные числа и векторы:

Пример:

Дана матрица M $= \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$

Решение:

Найдем собственные числа:

$$\det\begin{pmatrix} 5-\lambda & -4 \\ -4 & 5-\lambda \end{pmatrix} = 0 \rightarrow (5-\lambda)^2 - 16 = 0 \rightarrow \lambda^2 - 10\lambda + 9 = 0$$

Найдем собственные числа и векторы:

Пример:

Дана матрица $M = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$

Решение:

Найдем собственные числа:

$\det\begin{pmatrix} 5-\lambda & -4 \\ -4 & 5-\lambda \end{pmatrix} = 0 \rightarrow (5-\lambda)^2 - 16 = 0 \rightarrow \lambda^2 - 10\lambda + 9 = 0$

$\rightarrow \lambda_1 = 9; \lambda_2 = 1$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:
$(M - 9I)\overrightarrow{z} = 0$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:

$(M - 9I)\overrightarrow{z} = 0$

$\begin{pmatrix} -4 & -4 \\ -4 & -4 \end{pmatrix} \overrightarrow{z} = 0$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M-\lambda I)\overrightarrow{z}=0}$$

Подставим последовательно собственные числа:

$(M-9I)\overrightarrow{z}=0$

$\begin{pmatrix} -4 & -4 \\ -4 & -4 \end{pmatrix}\overrightarrow{z}=0 \rightarrow -4z_1-4z_2=0$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:

$(M - 9I)\overrightarrow{z} = 0$

$\begin{pmatrix} -4 & -4 \\ -4 & -4 \end{pmatrix} \overrightarrow{z} = 0 \rightarrow -4z_1 - 4z_2 = 0 \rightarrow z_1 = -z_2$

Общий вид собственного вектора (1): $\begin{pmatrix} -z_2 \\ z_2 \end{pmatrix}$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:
$(M - I)\overrightarrow{z} = 0$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:

$(M - I)\overrightarrow{z} = 0$

$\begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} \overrightarrow{z} = 0$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:
$(M - I)\overrightarrow{z} = 0$

$$\begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} \overrightarrow{z} = 0 \rightarrow 4z_1 - 4z_2 = 0$$

Найдем собственные числа и векторы:

Решение (продолжение):

$$\boxed{(M - \lambda I)\overrightarrow{z} = 0}$$

Подставим последовательно собственные числа:

$(M - I)\overrightarrow{z} = 0$

$\begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix}\overrightarrow{z} = 0 \rightarrow 4z_1 - 4z_2 = 0 \rightarrow z_1 = z_2$

Общий вид собственного вектора (2): $\begin{pmatrix} z_2 \\ z_2 \end{pmatrix}$

Работа с матрицами в рамках МГК

Имеем: ковариационная/корреляционная матрица C

Матрица C содержит информацию об исходном признаковом пространстве. Вспомним общий вид ковариационной матрицы:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

Работа с матрицами в рамках МГК

Имеем: ковариационная/корреляционная матрица C

Матрица C содержит информацию об исходном признаковом пространстве. Вспомним общий вид ковариационной матрицы:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

Находим в процессе: матрицу преобразования A

Матрица A – матрица, составленная из собственных векторов матрицы C.

Работа с матрицами в рамках МГК

Имеем: ковариационная/корреляционная матрица С

Матрица С содержит информацию об исходном признаковом пространстве. Вспомним общий вид ковариационной матрицы:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

Находим в процессе: матрицу преобразования А

Матрица А – матрица, составленная из собственных векторов матрицы С.

В результате: главные компоненты

Главная компонента (ГК): $y_{ji} = z_{1j}x_{1i} + ...z_{kj}x_{ki}$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda\overrightarrow{z}}$

Решим характеристическое уравнение:

$$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0$$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda\overrightarrow{z}}$

Решим характеристическое уравнение:

$$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0 \rightarrow (34-\lambda)(10-\lambda) - 25 = 0$$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda\overrightarrow{z}}$

Решим характеристическое уравнение:

$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0 \rightarrow (34-\lambda)(10-\lambda) - 25 = 0$

$\rightarrow \lambda^2 - 44\lambda + 315 = 0$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda\overrightarrow{z}}$

Решим характеристическое уравнение:

$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0 \to (34-\lambda)(10-\lambda) - 25 = 0$

$\to \lambda^2 - 44\lambda + 315 = 0 \to \lambda_1 = 35; \lambda_2 = 9$

$\lambda_1 = 35$ – вариация первой ГК

$\lambda_2 = 9$ – вариация второй (менее информативной) ГК

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix}\overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix}\overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$.

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = 5z_2 \\ 26z_2^2 = 1 \end{cases} \Rightarrow$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = 5z_2 \\ 26z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{5}{\sqrt{26}} \\ z_2 = \dfrac{1}{\sqrt{26}} \end{cases}$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \rightarrow -z_1 + 5z_2 = 0 \rightarrow z_1 = 5z_2$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = 5z_2 \\ 26z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{5}{\sqrt{26}} \\ z_2 = \dfrac{1}{\sqrt{26}} \end{cases}$$

Запишем ГК (1): $\boxed{\mathrm{y}_{1i} = \dfrac{5}{\sqrt{26}} x_{1i} + \dfrac{1}{\sqrt{26}} x_{2i}}$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0 \begin{pmatrix} 25 & 5 \\ 5 & 1 \end{pmatrix} \overrightarrow{z} = 0$

Запишем как систему уравнений:

$$\begin{cases} -5z_1 = z_2 \\ z_1^2 + 25z_1^2 = 1 \end{cases} \quad \Rightarrow$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0 \begin{pmatrix} 25 & 5 \\ 5 & 1 \end{pmatrix} \overrightarrow{z} = 0$

Запишем как систему уравнений:

$$\begin{cases} -5z_1 = z_2 \\ z_1^2 + 25z_1^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{1}{\sqrt{26}} \\ z_2 = \dfrac{-5}{\sqrt{26}} \end{cases}$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0 \begin{pmatrix} 25 & 5 \\ 5 & 1 \end{pmatrix} \overrightarrow{z} = 0$

Запишем как систему уравнений:

$$\begin{cases} -5z_1 = z_2 \\ z_1^2 + 25z_1^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{1}{\sqrt{26}} \\ z_2 = \dfrac{-5}{\sqrt{26}} \end{cases}$$

Запишем ГК (2): $\boxed{\mathrm{y}_{2i} = \dfrac{1}{\sqrt{26}} x_{1i} + \dfrac{-5}{\sqrt{26}} x_{2i}}$