

1. Сингулярное разложение (SVD)

Определение 1.1. Сингулярное разложение (SVD) — декомпозиция вещественной матрицы с целью ее приведения к каноническому виду.

Теорема 1.1. Пусть $A \in R^{m \times n}$. Существуют такие ОНБ $e_{k=1}^{k^n} \subset C^m$ и положительные числа $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r, 0 \leq r \leq \min(m, n)$, что $Ae^k = \begin{cases} \sigma_k q^k, k \leq r, \\ 0, k > r, \end{cases}$

$\sigma_1, \sigma_2, \dots, \sigma_r$ называются сингулярными числами матрицы A . Базисы $\{e^k\}_{k=1}^n, \{q^k\}_{k=1}^m$, обеспечивающие выполнение соотношения теоремы, называют сингулярными базисами матрицы A . Понятно, что r есть размерность $Im(A)$, т. е. r — ранг матрицы A .

Доказательство. Матрица A^*A самосопряжена и неотрицательна. Действительно,

$$(A^*A)^* = A^*A,$$

$$(A^*Ax, x) = (Ax, Ax) \geq 0, \forall x \in C^n.$$

Поэтому $\exists e_{k=1}^{k^n}$ - ОНБ собственных векторов матрицы $A^*A \Rightarrow$

$$A^*Ae^k = \sigma_k^2 e^k, \sigma_k^2 \geq 0, k = 1, 2, \dots, n \Rightarrow$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r, 0 \leq r \leq n, \sigma_{r+1} = \dots = \sigma_n = 0$. Положим $z^k = Ae^k$ для $k = 1, 2, \dots, r$ и

$$(z^p, z^q) = (Ae^p, Ae^q) = (A^*Ae^p, e^q) = \sigma_p^2 (e^p, e^q).$$

Значит, $(z^p, z^q) = \begin{cases} \sigma_p^2, p = q, \\ 0, p \neq q, \end{cases}$ значит, векторы $q^k = \sigma_k^{-1} Ae^k, k = 1, 2, \dots, r$, образуют

ортонормированную систему в пространстве C^m . Если $r < m$, дополним ее произвольно векторами $q^k, k = r+1, r+2, \dots, m$, до ОНБ пространства C^m . Из этого следует утверждение теоремы. \square

Определение 1.2. Сингулярным выражением матрицы A - выражение вида

$$A = V \Sigma W^*,$$

где $\Sigma \in R^{m \times n}$, у которой элементы, лежащие на главной диагонали - это сингулярные числа (а все элементы, не лежащие на главной диагонали - нулевые), а матрицы

$$V = \{q^k\}_{k=1}^m \in R^{m \times m}, W = \{e^k\}_{k=1}^n \in R^{n \times n}.$$

Геометрический смысл:

Пусть A - линейный оператор, отображающий R^n в себя через линейные операторы вращения, растяжения. Компоненты сингулярного разложения показывают эти геометрические преобразования.

2. Метод главных компонент

Определение 2.1. Метод Главных Компонент (*Principal Components Analysis, PCA*) — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации.

Теорема 2.1. Для данной матрицы $A \exists A_k$ - ее аппроксимация: $\text{rank}(A_k) = k \leq \text{rank}(A)$ и $\forall B, B \in R^{m \times n}, \text{rank}(B) = k : \|A - A_k\|_F \leq \|A - B\|_F$, где $\|\cdot\|_F$ - норма Фробениуса.

2.1. Формальная постановка задачи

Пусть имеется n числовых признаков $f_j(x), j = 1, \dots, n$. Объекты обучающей выборки будем отождествлять с их признаковыми описаниями: $x_i(f_1(x_i), \dots, f_n(x_i)), i = 1, \dots, l$. Рассмотрим матрицу F , строки которой соответствуют признаковым описаниям обучающих объектов:

$$F_{l \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_n) & \dots & f_n(x_n) \end{pmatrix} = \begin{pmatrix} x_1 \\ \dots \\ x_l \end{pmatrix}$$

Обозначим через $z_i = (g_1(x_i), \dots, g_m(x_i))$ признаковые описания тех же объектов в новом пространстве $Z = R^m$ меньшей размерности, $m < n$:

$$G_{l \times n} = \begin{pmatrix} g_1(x_1) & \dots & g_n(x_1) \\ \dots & \dots & \dots \\ g_1(x_n) & \dots & g_n(x_n) \end{pmatrix} = \begin{pmatrix} z_1 \\ \dots \\ z_l \end{pmatrix}$$

Потребуем, чтобы исходные признаковые описания можно было восстановить по новым описаниям с помощью некоторого линейного преобразования, определяемого матрицей $U = (u_{js})_{n \times m}$:

$$\hat{f}_j(x) = \sum_{s=1}^m g_s(x) u_{js}, \quad j = 1, \dots, n, \quad x \in X,$$

или в векторной записи: $\hat{x} = zU^T$. Восстановленное описание \hat{x} не обязано в точности совпадать с исходным описанием x , но их отличие на объектах обучающей выборки должно быть как можно меньше при выбранной размерности m . Будем искать одновременно и матрицу новых признаков G , и матрицу линейного преобразования U , при которых суммарная невязка $\Delta^2(G, U)$ восстановленных описаний минимальна:

$$\Delta^2(G, U) = \sum_{i=1}^l \|\hat{x}_i U^T - x_i\|^2 = \|GU^T - F\|^2 \rightarrow \min_{G, U}$$

Будем предполагать, что матрицы G и U невырождены: $\text{rank}(G) = \text{rank}(U) = m$. Иначе существовало бы представление $\hat{G}\hat{U}^T = GU^T$ с числом столбцов в матрице \hat{G} , меньшим m . Поэтому интересны лишь случаи, когда $m \leq \text{rank}(F)$.

Определение 2.2. u_1, \dots, u_m будем называть главными компонентами.

Теорема 2.2. Если $m \leq \text{rank}(f)$, то минимум $\Delta^2(G, U)$ достигается, когда столбцы матрицы U есть собственные векторы $F^T F$, соответствующие m максимальным собственным значениям. При этом $G = FU$, матрицы U и G ортогональны.

Геометрический смысл:

Метод аппроксимирует n -размерное облако наблюдений до n -мерного эллипсоида, полуоси которого будут являться главными компонентами.