**Сингулярное разложение (SVD) и Метод главных компонент (PCA)**

**Сингулярное разложение** — разложение прямоугольной матрицы, имеющее широкое применение, в силу своей наглядной геометрической интерпретации, при решении многих прикладных задач. Переформулировка сингулярного разложения, так называемое разложение Шмидта , имеет приложения в квантовой теории информации, например, в квантовой запутанности.

Сингулярное разложение матрицы M позволяет вычислять сингулярные числа данной матрицы, а также левые и правые сингулярные векторы матрицы M:

- Левые сингулярные векторы матрицы M – это собственные векторы матрицы $MM^*$.
- Правые сингулярные векторы матрицы M – это собственные векторы матрицы $M^*M$.

Где $M^*$ – эрмитово-сопряжённая матрица к матрице $M$, для вещественной матрицы $M^* = M^T$.

Сингулярные числа матрицы не следует путать с собственными числами той же матрицы.

Сингулярное разложение является удобным при вычислении ранга матрицы , ядра матрицы и псевдообратной матрицы .

Сингулярное разложение также используется для приближения матриц матрицами заданного ранга.

**Определение**

Пусть матрица $M$ порядка $m \times n$ состоит из элементов из поля $K$, где $K$ – либо поле вещественных чисел, либо поле комплексных чисел.

**Сингулярные числа и сингулярные векторы**

Неотрицательное вещественное число $\sigma$ называется сингулярным числом матрицы $M$, когда существуют два вектора единичной длины $u \in K^m$ и $v \in K^n$ такие, что:

$Mv = \sigma u$, и $M^*u = \sigma v$ Такие векторы $u$ и $v$ называются, соответственно, левым сингулярным вектором и правым сингулярным вектором, соответствующим сингулярному числу $\sigma$.

**Разложение матрицы**

Сингулярным разложением матрицы $M$ порядка $m \times n$ является разложение вида

$M = U\Sigma V^*$ где $\Sigma$ — диагональная матрица размера $m \times n$ с неотрицательными элементами, у которой элементы, лежащие на главной диагонали — это сингулярные числа, а матрицы $U$ (порядка $m$) и $V$ (порядка $n$) – это две унитарные матрицы, состоящие из левых и правых сингулярных векторов соответственно ($V^*$ — это сопряжённо-транспонированная матрица к $V$).

**Геометрический смысл**

Пусть матрице A поставлен в соответствие линейный оператор. Сингулярное разложение можно переформулировать в геометрических терминах. Линейный оператор, отображающий элементы пространства $\mathbb{R}^n$ в себя, представим в виде последовательно выполняемых линейных операторов вращения и растяжения. Поэтому компоненты сингулярного разложения наглядно показывают геометрические изменения при отображении линейным оператором $A$ множества векторов из векторного пространства в себя или в векторное пространство другой размерности.

Для более визуального представления рассмотрим сферу $S$ единичного радиуса в пространстве $\mathbb{R}^n$. Линейное отображение $T$ отображает эту сферу в эллипсоид пространства $\mathbb{R}^m$. Тогда ненулевые сингулярные значения диагонали матрицы $\Sigma$ являются длинами полуосей этого эллипсоида. В случае когда $n = m$ и все сингулярные величины различны и отличны от нуля, сингулярное разложение линейного отображения $T$ может быть легко проанализировано как последствие трех действий: рассмотрим эллипсоид $T(S)$ и его оси; затем рассмотрим направления в $\mathbb{R}^n$, которые отображение $T$ переводит в эти оси. Эти направления ортогональны. Вначале применим изометрию $\mathbf{v}^*$, отобразив эти направления на координатные оси $\mathbb{R}^n$. Вторым шагом применим эндоморфизм $\mathbf{d}$, диагонализированный вдоль координатных осей и расширяющий/сжимающий эти направления, используя длины полуосей $T(S)$ как коэффициенты растяжения. Тогда произведение $\mathbf{d} \otimes \mathbf{v}^*$ отображает единичную сферу на изометричный эллипсоид $T(S)$. Для определения последнего шага u просто применим изометрию к этому эллипсоиду так, чтобы перевести его в $T(S)$. Произведение $\mathbf{u} \otimes \mathbf{d} \otimes \mathbf{v}^*$ совпадает с $T$.

**Метод главных компонент** — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Вычисление главных компонент может быть сведено к вычислению сингулярного разложения матрицы данных или к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных.

**Формальная постановка задачи**

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

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

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

$$G_{l \times m} = \begin{pmatrix} g_1(x_1) & \cdots & g_m(x_1) \\ \ldots & \ldots & \ldots \\ g_1(x_l) & \cdots & g_m(x_l) \end{pmatrix} = \begin{pmatrix} z_1 \\ \ldots \\ z_l \end{pmatrix}.$$

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

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

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 признаковых описаний $G$, и матрицу линейного преобразования $U$, при которых суммарная невязка $\Delta^2(G, U)$ восстановленных описаний минимальна:

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

где все нормы евклидовы. Будем предполагать, что матрицы $G$ и $U$ невырождены: $\operatorname{rank} G = \operatorname{rank} U = m$. Иначе существовало бы представление $\bar{G}\bar{U}^T = GU^T$ с числом столбцов в матрице $\bar{G}$, меньшим $m$. Поэтому интересны лишь случаи, когда $m \leq \operatorname{rank} F$.

**Теорема 1.** *Если $m \leq \operatorname{rank} F$, то минимум $\Delta^2(G, U)$ достигается, когда столбцы матрицы $U$ есть собственные векторы $F^T F$, соответствующие $m$ максимальнып собственным значениям. При этом $G = FU$, матрицы $U$ и $G$ ортогональны.*

*Доказательство.* Запишем необходимые условия минимума:

$$\begin{cases} \frac{\partial \Delta^2}{\partial G} = \left(GU^T - F\right) U = 0 \\ \frac{\partial \Delta^2}{\partial U} = G^T \left(GU^T - F\right) = 0 \end{cases}$$

Поскольку искомые матрицы $G$ и $U$ невырождены, отсюда следует:

$$\begin{cases} G = FU \left(U^T U\right)^{-1} \\ U = F^T G \left(G^T G\right)^{-1} \end{cases}$$

Функционал $\Delta^2(G, U)$ зависит только от произведения матриц $GU^T$, поэтому решение задачи $\Delta^2(G, U) \to \min_{G,U}$ определено с точностью до произвольного невырожденного преобразования $R : GU^T = (GR)\left(R^{-1}U^T\right)$. Распорядимся свободой выбора $R$ так, чтобы матрицы $U^T U$ и $G^T G$ оказались диагональными. Покажем, что это всегда возможно. Пусть $\tilde{G}\tilde{U}^T -$ произвольное решение задачи. Матрица $\tilde{U}^T\tilde{U}$ симметричная, невырожденная, положительно определенная, поэтому существует невырожденная матрица $S_{m\times m}$ такая, что $S^{-1}\tilde{U}^T\tilde{U}\left(S^{-1}\right)^T = I_m$

Матрица $S^T\tilde{G}^T\tilde{G}S$ симметричная и невырожденная, поэтому существует ортогональная матрица $T_{m\times m}$ такая, что $T^T\left(S^T\tilde{G}^T\tilde{G}S\right)T = \operatorname{diag}(\lambda_1,\ldots,\lambda_m) \equiv \Lambda -$ диагональная матрица. По определению ортогональности $T^TT = I_m$ Преобразование $R = ST$ невырождено. Положим $G = \tilde{G}R, U^T = R^{-1}\tilde{U}^T$. Тогда

$$G^TG = T^T\left(S^T\tilde{G}^T\tilde{G}S\right)T = \Lambda$$
$$U^TU = T^{-1}\left(S^{-1}\tilde{U}^T\tilde{U}\left(S^{-1}\right)^T\right)\left(T^{-1}\right)^T = \left(T^TT\right)^{-1} = I_m$$

В силу $GU^T = \tilde{G}\tilde{U}^T$ матрицы $G$ и $U$ являются решением задачи $\Delta^2(G,U) \to \min_{G,U}$ и удовлетворяют необходимому условию минимума. Подставим матрицы $G$ и $U$ в

$$G = FU\left(U^TU\right)^{-1}$$
$$U = F^TG\left(G^TG\right)^{-1}$$

Благодаря диагональности $G^TG$ и $U^TU$ соотношения существенно упростятся:

$$\begin{cases} G = FU \\ U\Lambda = F^TG \end{cases}$$

Подставим первое соотношение во второе, получим $U\Lambda = F^TFU$. Это означает, что столбцы матрицы $U$ обязаны быть собственными векторами матрицы $F^TF$, а диагональные элементы $\lambda_1,\ldots,\lambda_m$ соответствующими им собственными значениями. Аналогично, подставив второе соотношение в первое, получим $G\Lambda = FF^TG$, то есть столбцы матрицы $G$ являются собственными векторами $FF^T$, соответствующими тем же самым собственным значениям. Подставляя $G$ и $U$ в функционал $\Delta^2(G,U)$, находим:

$$\Delta^2(G,U) = \left\|F - GU^T\right\|^2 = \operatorname{tr}\left(F^T - UG^t\right)\left(F - GU^T\right) = \operatorname{tr}F^T\left(F - GU^T\right) =$$
$$\operatorname{tr}F^TF - \operatorname{tr}F^TGU^T = \|F\|^2 - \operatorname{tr}U\Lambda U^T = \|F\|^2 - \operatorname{tr}\Lambda = \sum_{j=1}^{n}\lambda_j - \sum_{j=1}^{m}\lambda_j - \sum_{j=m+1}^{n}\lambda_j$$

где $\lambda_1,\ldots,\lambda_n$ - все собственные значения матрицы $F^TF$. Минимум $\Delta^2$ достигается, когда $\lambda_1,\ldots,\lambda_m -$ наибольшие $m$ из $n$ собственных значений.

Собственные векторы $u_1,\ldots,u_m$, отвечающие максимальным собственным значениям, называют главными компонентами. $\square$

### Связь с сингулярным разложением

Если $m = n$, то $\Delta^2(G,U) = 0$. В этом случае представление $F = GU^T$ является точным и совпадает с сингулярным разложением: $F = GU^T = VDU^T$, если положить $G = VD$ и $\Lambda = D^2$. При этом матрица $V$ ортогональна: $V^TV = I_m$.

Если $m < n$, то представление $F \approx GU^T$ является приближённым. Сингулярное разложение матрицы $GU^T$ получается из сингулярного разложения матрицы $F$ путём отбрасывания (обнуления) $n - m$ минимальных собственных значений.