Сингулярное разложение

Сингулярное разложение (Singular Values Decomposition, SVD) является удобным методом при работе с матрицами. Сингулярное разложение показывает геометрическую структуру матрицы и позволяет наглядно представить имеющиеся данные. Сингулярное разложение используется при решении самых разных задач — от приближения методом наименьших квадратов и решения систем уравнений до сжатия и распознавания изображений. Используются разные свойства сингулярного разложения, например, способность показывать ранг матрицы и приближать матрицы данного ранга. Так как вычисления ранга матрицы — задача, которая встречается очень часто, то сингулярное разложение является довольно популярным методом.

Определение SVD

Теорема 1 (Дж. Форсайт)**.** *Для любой вещественной $(n \times n)$-матрицы A существуют две вещественные ортогональные $(n \times n)$-матрицы U и V такие, что*

$$U^T AV = \Lambda.$$

Более того, можно выбрать U и V так, чтобы диагональные элементы Λ имели вид

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > \lambda_{r+1} = ... = \lambda_n = 0,$$

где r — ранг матрицы A. В частности, если A невырождена, то

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n > 0.$$

Индекс r элемента λ_r есть фактическая размерность собственного пространства матрицы A.

Столбцы матриц U и V называются соответственно левыми и правыми сингулярными векторами, а значения диагонали матрицы Λ называются сингулярными значениями.

Пусть A — $(m \times n)$-матрица и ей в соответствие поставлен линейный оператор, также обозначаемый A. Формулу сингулярного разложения $A = U\Lambda V^T$ можно переформулировать в геометрических терминах. Линейный оператор, отображающий элементы пространства $\mathbb{R}^m$ в элементы пространства $\mathbb{R}^n$ представим в виде последовательно выполняемых линейных операций вращения, растяжения и вращения. Число ненулевых элементов на диагонали матрицы Λ есть фактическая размерность матрицы A.

Поэтому компоненты сингулярного разложения наглядно показывают геометрические изменения при отображении линейным оператором A множества векторов из одного векторного пространства в другое.

Например, матрица

$$A = \begin{pmatrix} 0.96 & 1.72 \\ 2.28 & 0.96 \end{pmatrix}$$

имеет сингулярное разложение

$$A = U\Lambda V^T = \begin{pmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.8 & 0.6 \\ 0.6 & -0.6 \end{pmatrix}^T$$

Легко увидеть, что матрицы U и V ортогональны,

$$U^T U = UU^T = I, \text{ также } V^T V = VV^T = I,$$

и сумма квадратов значений их столбцов равна единице; для каждой из матриц $0.6^2 + 0.8^2 = 1$.

Наивный алгоритм SVD

Рассмотрим приближенное линейное описание матрицы $A = \{a_{ij}\}$ вида

$$a_{ij} = \sum_{k=1}^{r} u_{ik}\lambda_k v_{kj} + c_{ij}, \tag{1}$$

где $i = 1, ..., m$ и $j = 1, ..., n$.

Значения $u_{ik}, \lambda_k, v_{kj}$ для данного значения k найдены из условия минимума выражения

$$\varepsilon^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}^2, \tag{2}$$

при ограничениях нормировки

$$\sum_{j=1}^{n} u_{ik}^2 = \sum_{i=1}^{m} v_{kj}^2 = 1 \tag{3}$$

и упорядоченности $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r \geq ... \geq 0$.

Выражения (1), (2), (3) запишем в матричных обозначениях:

$$\begin{aligned} &A = U\Lambda V^T + C, \\ &\varepsilon^2 = \mathrm{tr}(CC^T) = \|C\|^2, \\ &U^T U = VV^T = I, \end{aligned}$$

где матрицы $U = \{u_{kj}\}, \Lambda = \mathrm{diag}\{\lambda_k\}, V = \{v_{ik}\}$. Если значение r достаточно велико, то C=(0). Так будет заведомо при $r \geq \min\{m, n\}$. Минимальное значение r, при котором выполнимо равенство $A = U\Lambda V^T$, равно рангу матрицы A.

Один из возможных алгоритмов нахождения сингулярного разложения заключается в следующем. Найдем последовательно векторы $\mathbf{u}_k, \mathbf{v}_k$ и сингулярные числа λ_k

для $k = 1, ..., r$. В качестве этих векторов берутся нормированные значения векторов $\mathbf{a}_k$ и $\mathbf{b}_k$ соответственно: $\mathbf{u}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|}$, $\mathbf{v}_k = \frac{\mathbf{b}_k}{\|\mathbf{b}k\|}$. Векторы $\mathbf{a}_k$ и $\mathbf{b}_k$ находятся как пределы последовательностей векторов $\{\mathbf{a}_{k_i}\}$ и $\{\mathbf{b}_{k_i}\}$, соответственно $\mathbf{a}_k = \lim(\mathbf{a}_{k_i})$ и $\mathbf{b}_k = \lim(\mathbf{b}_{k_i})$. Сингулярное число λ_k находится как произведение норм векторов: $\lambda_k = \|\mathbf{a}_k\| \cdot \|\mathbf{b}_k\|$.

Процедура нахождения векторов $\mathbf{u}_k, \mathbf{v}_k$ начинается с выбора наибольшей по норме строки $\mathbf{b}_{1_1}$ матрицы A. Для $k = 1$ формулы нахождения векторов $\mathbf{a}_{1_i}, \mathbf{b}_{1_i}$ имеют вид:

$$\mathbf{a}_{1_i} = \frac{A\mathbf{b}_{1_i}^T}{\mathbf{b}_{1_i}\mathbf{b}_{1_i}^T}, \qquad \mathbf{b}_{1_{i+1}} = \frac{\mathbf{a}_{1_i}^T A}{\mathbf{a}_{1_i}^T \mathbf{a}_{1_i}}, \qquad i = 1, 2, ...$$

Для вычисления векторов $\mathbf{u}_k, \mathbf{v}_k$ при $k = 2, ..., r$ используется вышеприведенная формула, с той разницей, что матрица A заменяется на скорректированную на k-м шаге матрицу $A_{k+1} = A_k - \mathbf{u}_k \lambda_k \mathbf{v}_k$.

TODO: Написать в чем недостаток по сравнению со стандартным разложением, включающем разложение Холесского. Картинка.

Сингулярное разложение и собственные значения матрицы

Сингулярное разложение обладает свойством, которое связывает задачу отыскания сингулярного разложения и задачу отыскания собственных векторов. (Собственный вектор $\mathbf{x}$ матрицы A — такой вектор, при котором выполняется условие $A\mathbf{x} = \lambda\mathbf{x}$, число λ называется собственным числом.) Так как матрицы U и V ортогональные, то есть

$$U^T U = V V^T = I, \tag{4}$$

где I — единичная матрица размерности $r \times r$, то из (4) следует, что

$$\begin{aligned} AA^T = U\Lambda V^T V \Lambda U^T = U\Lambda^2 U^T, \\ A^T A = V\Lambda U^T U \Lambda V^T = V\Lambda^2 V^T. \end{aligned} \tag{5}$$

Умножая оба выражения справа соответственно на U и V получаем

$$\begin{aligned} AA^T U = U\Lambda^2, \\ A^T A V = V\Lambda^2. \end{aligned} \tag{6}$$

Из выражения (6) следует, что столбцы матрицы U являются собственными векторами матрицы AA^T, а квадраты сингулярных чисел $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_r)$ ее собственными значениями. Также столбцы матрицы V являются собственными векторами матрицы $A^T A$, а квадраты сингулярных чисел являются ее собственными значениями.

Пространства матрицы и SVD

Сингулярное разложение позволяет найти ортогональные базисы различных векторных пространств разлагаемой матрицы. В теореме о сингулярном разложении (Форсайт) рассматривается матрица размером $(m \times m)$,

$$A_{(m\times m)} = U_{(m\times m)} \Lambda_{(m\times m)} V^T_{(m\times m)}.$$

Существует так называемое экономное представление сингулярного разложения матрицы.

$$A_{(m\times n)} = U_{(m\times m)}\Lambda_{(m\times n)}V^T_{(n\times n)}$$

Согласно этому представлению при $m > n$, диагональная матрица Λ имеет пустые строки, а при $m < n$ — пустые столбцы. Поэтому существует еще одно экономное представление

$$A_{(m\times n)} = U_{(m\times r)}\Lambda_{(r\times r)}V^T_{(r\times n)},$$

в котором $r = \min(m, n)$.

Нуль-пространство матрицы A — набор векторов $\mathbf{x}$, для которого справедливо высказывание $A\mathbf{x} = \mathbf{0}$. Собственное пространство матрицы A — набор векторов $\mathbf{b}$, при котором уравнение $A\mathbf{x} = \mathbf{b}$ имеет ненулевое решение для $\mathbf{x}$. Обозначим $\mathbf{u}_k$ и $\mathbf{v}_k$ — столбцы матриц U и V. Тогда разложение $A = U\Lambda V^T$ может быть записана в виде: $A = \sum_{k=1}^r A_k$, где $A_k = \mathbf{u}_k\lambda_k\mathbf{v}_k^T$. Если сингулярное значение $\lambda_k = 0$, то $A\mathbf{v}_k = \mathbf{0}$ и $\mathbf{v}_k$ находится в нуль-пространстве матрицы A, а если сингулярное значение $\lambda_k \neq 0$, то вектор $\mathbf{u}_k$ находятся в собственном пространстве матрицы A. Следовательно, можно сконструировать базисы для различных векторных подпространств, определенных матрицей A. Набор векторов $\mathbf{v}_1, \ldots, \mathbf{v}_k$ в векторном пространстве V формирует *базис* для V, если любой вектор $\mathbf{x}$ из V можно представить в виде линейной комбинации векторов $\mathbf{v}_1, \ldots, \mathbf{v}_k$ единственным способом. Пусть V_0 будет набором тех столбцов $\mathbf{v}_k$, для которых $\lambda_k \neq 0$, а V_1 — все остальные столбцы $\mathbf{v}_k$. Также, пусть U_0 будет набором столбцов $\mathbf{u}_k$, для которых $\lambda_k \neq 0$, а U_1 — все остальные столбцы $\mathbf{u}_k$, включая и те, для которых $k > n$. Тогда, если r — количество ненулевых сингулярных значений, то имеется r столбцов в наборе V_0 и $n-r$ столбцов в наборе V_1 и U_1, а также $m-n+r$ столбцов в наборе U_0. Каждый из этих наборов формирует базис векторного пространства матрицы A:

V_0 — ортонормальный базис для ортогонального комплементарного нуль-пространства A,

V_1 — ортонормальный базис для нуль-пространства A,

U_0 — ортонормальный базис для собственного пространства A,

U_1 — ортонормальный базис для ортогонального комплементарного нуль-пространства A.

Сингулярное разложение и нормирование матриц

Рассмотрим изменение длины вектора $\mathbf{x}$ до и после его умножения справа матрицу A. Евклидова норма вектора определена как

$$\|\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{x}.$$

При умножении вектора $\mathbf{x}$ на матрицу A справа, длина результирующего вектора $A\mathbf{x}$ изменяется. Если матрица A ортогональна, длина вектора $A\mathbf{x}$ остается неизменной. В противном случае, с помощью выражения $\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$, можно вычислить, насколько матрица A растянула вектор $\mathbf{x}$. Таким образом, евклидова норма матрицы, есть максимальный коэффициент растяжения вектора.

Евклидова норма матрицы определяется так. Пусть $\mathbf{x}$ является n-мерным вектором, и A — матрица размерности $(m \times n)$. Тогда Евклидова норма матрицы

$$\|A\|_E = \max_{\|\mathbf{x}\|=1} \left(\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \right).$$

Альтернативной Евклидовой норме является норма Фробениуса:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}.$$

Если известно сингулярное разложение, то обе эти нормы легко узнать. Пусть $U\Lambda V^T$ — сингулярное разложение $(m \times n)$-матрицы A. Тогда

$$\|A\|_E = \lambda_1,$$

и

$$\|A\|_F = \sqrt{\sum_{k=1}^{r} \lambda_k^2}.$$

Сингулярные значения матрицы A — это длины осей эллипсоида, заданного множеством $\{A\mathbf{x}|1 = \|\mathbf{x}\|_2\}$. (картинка)

Нахождение псевдообратной матрицы с помощью SVD

Классическая задача наименьших квадратов ставиться следующим образом. Даны действительная $(m{\times}n)$-матрица A и действительный (m)-вектор Y. Требуется найти действительный (n)-вектор $\mathbf{w}$, минимизирующий Евклидову длину вектора невязки,

$$\|Y - A\mathbf{w}\|_2 \longrightarrow \min.$$

Решение задачи наименьших квадратов —

$$\mathbf{w} = (A^T A)^{-1}(A^T Y).$$

Для отыскания решения $\mathbf{w}$ требуется обратить матрицу $A^T A$. Рассмотрим более общий случай обращения матриц.

Если $(m \times n)$-матрица A является вырожденной или прямоугольной, то обратной матрицы A^{-1} для нее не существует. Однако, для A может быть найдена псевдообратная матрица A^+ — такая матрица, для которой выполняются условия

$$\begin{aligned} &A^+A = I_n, \\ &AA^+ = I_m, \\ &A^+AA^+ = A^+, \\ &AA^+A = A. \end{aligned} \tag{7}$$

Пусть найдено разложение матрицы A вида

$$A = U\Lambda V^T,$$

где $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_r)$, $r = \min(m, n)$ и $U^TU = I_m, VV^T = I_n$. Тогда справедлива

Теорема 2. *Матрица $A^+ = V^T\Lambda^{-1}U$ является для матрицы A псевдообратной.*

Доказательство. Согласно теореме 1 и определению (7), $A^+A = V\Lambda^{-1}U^TU\Lambda V^T = I_n$, $AA^+ = U\Lambda V^TV\Lambda^{-1}U^T = I_m$. □

Усеченное обращение матриц

Алгоритм обращения матрицы посредством усеченного сингулярного разложения состоит в следующем. Пусть матрица исходных данных A представлена в виде $A = U\Lambda V^T$. Тогда при нахождении обратной матрицы $A^+ = V\Lambda^{-1}U^T$ в силу ортогональности матриц U и V: $U^TU = VV^T = I$, и в силу условия убывания диагональных элементов матрицы $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$, псевдообратная матрица A^+ будет более зависеть от тех элементов матрицы Λ, которые имеют меньшие значения, чем от первых сингулярных чисел. Действительно, если, по условию теоремы о сингулярном разложении матрица A имеет сингулярные числа $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_n$, то сингулярные числа матрицы A^+ равны $\Lambda^{-1} = \mathrm{diag}(\frac{1}{\lambda_1}, ..., \frac{1}{\lambda_n})$ и $\frac{1}{\lambda_1} \leqslant \frac{1}{\lambda_2} ... \leqslant \frac{1}{\lambda_n}$. Считая первые r сингулярных чисел определяющими собственное пространство матрицы A, используем при обращении матрицы A первые r сингулярных чисел. Тогда обратная матрица A^+ будет найдена как $A^+ = V\Lambda_r^{-1}U^T$.

Определим усеченную псевдообратную матрицу A_r^+ как

$$A_r^+ = V\Lambda_r^{-1}U^T, \tag{8}$$

где $\Lambda_r^{-1} = \mathrm{diag}(\lambda_1^{-1}, ..., \lambda_r^{-1}, 0, ..., 0)$ — диагональная матрица размерности $n \times n$.